Pattern Recognition 46 (2013) 3157–3173

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

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

A literature survey on robust and efficient eye localization in real-life scenarios

Fengyi Song [a], Xiaoyang Tan [a,*], Songcan Chen [a], Zhi-Hua Zhou [b]

[a] Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Yudao Street 29, Nanjing 210016, China
[b] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
17 April 2013
Accepted 11 May 2013
Available online 23 May 2013

Keywords:
Eye localization
Survey
Computer vision

ABSTRACT

Eye localization has gained a wide range of applications in face recognition, gaze estimation, pose estimation, expression analysis, etc. However, due to the high degree of appearance variability of eyes in size, shape, color, texture and various ambient environment changes, this task is challenging. During the last three decades, numerous techniques have been developed to meet these challenges. The goal of this paper is to categorize and evaluate these algorithms in a comprehensive way. We focus on the overall difficulties and challenges in real-life scenarios, and present a detailed review of prominent algorithms from the perspective of learning generalizable, flexible and efficient statistical eye models from a small number of training images. In addition, we organize the discussion of the global aspects of eye localization in uncontrolled environments, towards the development of a robust eye localization system. This paper concludes with several promising directions for future research.

1. Introduction

As one of the most salient facial features, eyes, which reflect the individual's affective states and focus attention, are one of the most important information sources for face analysis. Efficiently and accurately locating the eyes positions in a given face image is therefore essential to a wide range of face-related research efforts, including face alignment, face recognition, gaze estimation, pose estimation, expression analysis, etc., and has gained increasing attention from both the academic and industrial communities in the last three decades.

However, the task of accurate eye localization is challenging due to the high degree of eye's appearance variability. This variability may be caused either by intrinsic dynamic features of the eyes or by ambient environment changes. In particular, the following factors have significant influence on the states of the eyes:

- *Facial expression variations*: both the shape and appearance of the eyes are sensitive to the change of various expressions. For example, laughing may cause the eyes to close completely, and screaming may largely deform the shape of the eyes as well.
- *Occlusion*: in real application scenarios, the eyes are frequently occluded by hair, sunglasses, and myopia glasses with black frames.
- *Pose*: the appearance of the eyes differs between different camera-object poses (e.g., frontal, profile, upside down). Furthermore, it is possible that one eye is completely occluded in a profile face.
- *Imaging condition and quality*: ambient environment factors, such as lighting (varying in spectra, source distribution, and intensity), may change the appearance of the eyes in different ways. Moreover, the commonly seen factors in the real world, such as low resolution, blurring or detailed texture missing, may also lead to poor image quality. These cause great challenges to any eye localization algorithms.

These challenges are illustrated by the Labeled Face in the Wild (LFW) [42] database. As Fig. 1 shows, poor image quality and great appearance variations are the typical characteristics of images under the uncontrolled application scenarios. These factors pose great challenges to the existing eye localization techniques. Riopka and Boult [85] give a comprehensive eye perturbation sensitivity analysis, and their empirical evidence shows that the accuracy of eye localization has a significant effect on face recognition accuracy. Similar observations have also been made by many other authors [72,16,45,89], which urge the need for developing robust and accurate eye localization techniques in real-life scenarios.

Eye localization is closely related to but different from several tasks, such as eye detection, eye tracking, gaze estimation and blink detection. The purpose of eye detection is to determine the existence of eyes in an input image and, if any, find their positions. Eye localization, however, requires a much more accurate

* Corresponding author. Tel.: +86 25 8489 2956; fax: +86 25 8489 2452.
E-mail addresses: f.song@nuaa.edu.cn (F. Song), x.tan@nuaa.edu.cn (X. Tan), s.chen@nuaa.edu.cn (S. Chen), zhouzh@lamda.nju.edu.cn (Z.-H. Zhou).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.05.009

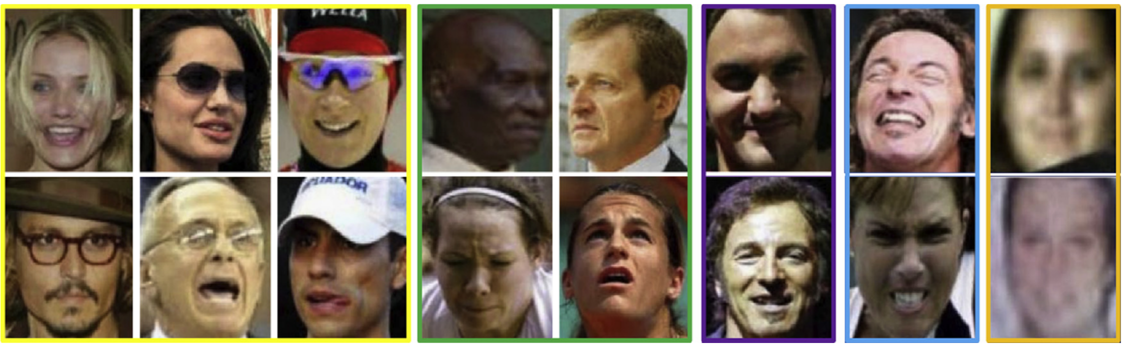

Fig. 1. An illustration of the great challenges of eye localization under the uncontrolled conditions (LFW) [42], from left to right: variations in occlusion, pose, lighting, expression, and blur.

prediction of the eye positions (usually with an error margin of a few pixels). Eye localization is generally treated as a subsequent fine tune step after eye detection. In eye tracking, another coordinate, i.e., time, is taken into account, and the redundant information between neighboring frames is usually explored to facilitate eye localization heuristically. Gaze estimation aims to infer individual focus attention by analyzing the pupil position in the eye socket. In blink detection, instances where the eye opens and closes are analyzed across the image sequence so as to estimate the individual's physical states (e.g., fatigue, active). The performance of these applications can be benefited from robust eye localization.

Numerous methods have been proposed for eye localization during the last three decades. Recently Campadelli et al. [14] surveyed several typical methods for eye localization under controlled conditions and proposed an objective performance evaluation criterion. Hansen and Ji [37] reviewed current progress in video-based eye detection and tracking techniques. However, it remains to be seen whether the state-of-the-art eye localization techniques perform well under uncontrolled conditions. As illustrated in Fig. 1, the problem of eye localization in real-life scenarios is much more challenging than that in controlled conditions, and is far from being resolved.

The major contribution of this paper is to give a comprehensive and critical survey of the ad hoc methods addressing these challenges, which we believe would be a useful complement to [14] and [37]. To be self-contained, some traditional methods for eye localizations covered in [14,37] are included in this work as well but due to the inherent complexity of eye localization in the wild, contrary to the previous works, we pay special attention on the problem of learning generalizable, flexible and efficient statistical eye models from a small number of training images, and the related topics such as the feature extraction and representation are discussed under this point of view. In addition, we organize the discussion of the global aspects of eye localization in uncontrolled environments, towards the development of a robust eye localization system (cf. Fig. 11), which in our opinion is a very important topic in practice but mostly ignored in previous studies.

In the following sections we first review state-of-the-art methods for eye localization and focus on the machine learning and computer vision techniques which are successfully applied to this problem. In Section 3, we investigate how the reviewed methods may be integrated in the development of a robust eye localization system, and several practical issues which have a critical influence on the system performance are also discussed in this global perspective. In Section 4, we discuss a few issues concerning performance evaluation. Finally, we conclude this paper with a discussion of several promising directions for eye localization in Section 5.

2. Localizing eyes in a single face

In this section, we review the existing techniques for eye localization. We broadly classify them into three categories based on the information or patterns that are used for model building. Note that some methods are at the overlapping category boundaries, such case will be discussed further at the end of this section.

- *Measuring eye characteristics*: this type of method exploits the inherent features of eyes as facial components, such as their distinct shapes and strong intensity contrast. Some context-related features such as the characteristics of the facial region between two eyes and the eye corner may also be useful. Due to the peculiarity of the eye features, eye localization can be performed by simply measuring such characteristics. However, reliable measuring is possible only under good imaging conditions.
- *Learning statistical appearance model*: this type of method tries to extract useful visual features from photometric appearance, based on which eye model is then learned from a large set of training images. The collected training data should cover representative variability of eye appearance.
- *Exploiting structural information*: this approach explores the spatial structure of interior components of eyes or the geometrical regularity between eyes and other facial features in the face context. The structural information cannot be used alone and is usually integrated in a statistical eye model to improve its stableness against complicated uncontrolled conditions.

Table 1 summarizes algorithms and representative works for eye localization within these three categories. Below, we discuss the motivations and general approaches of each category, and then give a review of specific methods followed with a discussion of their pros and cons.

2.1. Measuring eye characteristics

The idea of this line of research is to explore the distinct inherent features of eyes by treating eyes as a special facial feature by itself. Many eye-specific characteristics are ready to use in practice, such as the shape of the eyes and the intensity contrast between the eye white and the pupil. Once founded, these characteristics could be very reliable indicators of eyes. However, under the uncontrolled conditions, the measured characteristics tend to be less reliable, which may result in great performance loss. In addition, some characteristics, such as bright spot for infrared eye images, depend heavily on extra hardware devices and usually require active human cooperation.

2.1.1. *Shape characteristics*

Fig. 2 gives a typical shape model of eyes consisting of four major components: eyelids, the eye white, the iris and the pupil. Each of these components has a special geometric shape, e.g., elliptical shape for eyelids, circular shape for the iris and the pupil. Generally, there are two ways to represent these shapes, i.e., continuous or discrete. The representative work of the continuous shape model is Yuille et al.'s deformable model [111], while Active Shape Model (ASM) [19] represents the eyes in a discrete way.

In particular, Yuille et al. [111] built a parameterized deformable model which gives a continuous mathematical formulation of the shapes involved. More than that, their model takes into account the relevant features such as peaks and valleys of the accumulated intensity in the regions bounded by shapes. They have to search through a large continuous parameter space which covers almost all reasonable shape variations to fit the model to a testing image.

For a good fitting in practice, two factors are of importance [111,103,58] : (1) the distinctness and flexibility of the shape model, i.e., the model should expressive enough to explain large reasonable variability of shape while suppresses impractical deformations; (2) a properly initialized template. In [17,21,18], model parameters are redesigned to improve the discriminant and flexibility of deformable model, while others [103,58] propose to improve the quality of the initialization by localizing eye corners first. Despite these improvements, deformable models still suffer from their large search space and their dependence on good image quality and good initialization.

Table 1
Categorization of the popular approaches for eye localization.

Approach	Representative works
Measuring eye characteristics	
Shape and intensity contrast	Deformable model [110], circular shape of the pupil [78,46] Dark eye center [27,114], pupil centered outward gradient field [53,99,97]
Context information of eyes	Facial region between-eyes [50], eye corners [103,58,94]
Active eye localization	Localization and tracking for infrared eye images [117,116]
Learning statistical appearance model	
Bayesian model	Bayesian model [22] Multi-scale LBP feature based [55]
AdaBoost model	Discriminant feature based AdaBoost [101] 2D Cascaded AdaBoost [77] Bayesian criteria based AdaBoost [67] Probabilistic cascade [109]
Filtering model	Average of Synthetic Exact Filters (ASEF) [10]
Exploiting structural information	Active shape model [19], implicity shape model [62], enhanced pictorial model [92]

Another way to model the eye shape is to construct a statistical model for all the allowable shapes based on a series of discrete landmarks on the eyes. The Active Shape Model (ASM) [19] is the representative, which will be discussed in more detail in Section 2.3.

Besides building a global shape model for eyes, it is possible to learn different local shape models for different eye components, e.g., for the pupil and the iris. In [78], Hough transform was utilized to detect the circle shape of the eyes, and [46] proposed a neural network-based method with a compact circular perception fields for eye localization.

Since eye corners are less vulnerable to the changes of eye states, they are sometimes used to improve the initialization accuracy of the deformable model [103,58,47] (cf. Fig. 4). For example, [47] presented a local Hough voting based method for face alignment, using the spatial constraints imposed by the stable facial components to guide the search of other facial points.

2.1.2. *Intensity contrast characteristics*

The intensity distribution pattern of eyes is another useful cue for eye localization. For open eyes, the intensity contrast between eyes components such as the pupil, the iris, the eye white, and the eyelids is strong, while the gray intensity at the pupil region is usually much lower than that of iris and eye-white (cf. Fig. 3). Intensity patterns like these are commonly used as the heuristic evidences for the existence of eyes [114].

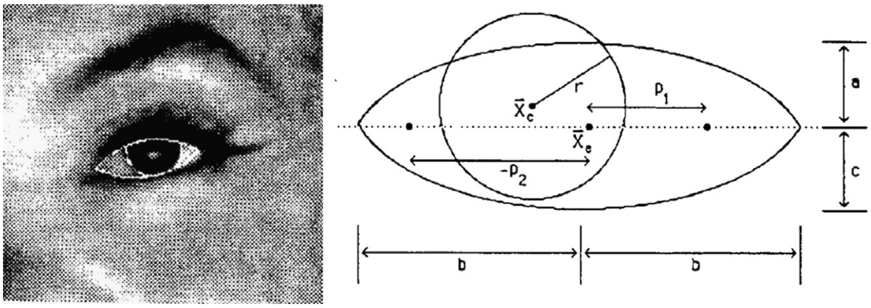

Fig. 2. Yuille et al.'s shape model of eyes [110].

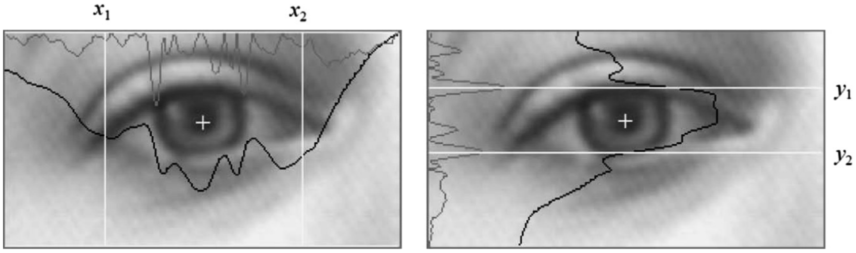

Fig. 3. Using projection function to locate the *x*-coordinate of the eye corners and the *y*-coordinate of the eyelids [114].

Typical methods to measure such patterns include Variance Projection Function (VPF) and Integral Projection Function (IPF) [27], both of which are adopted in General Projection Function (GPF) [114] under a unified architecture for accurate eye localization. These projection functions actually estimate the global intensity distribution around the coarse eye region. However, in some real-world applications, the global intensity distribution might be deteriorated by noisy light spots of iris. To address this, [65] proposed to accumulate locally smoothed version of pixel intensity, which tends to be more stable compared to the global one.

Alternatively, Wang et al. [99] developed a facial landscape navigation technique for eye localization, in which the interested intensity pattern (a pit at the eye center surrounded by hillside) is searched in a 3D terrain surface manifold of the face. In a similar method, gradient patterns [53] instead of intensity patterns are calculated around the eye region and served as the template for eye matching.

2.1.3. Context characteristics

When the shape or intensity characteristics of the eyes cannot be reliably measured, the context characteristics are very useful for eye localization. This is because eyes in the face context usually have stable relationship with other facial features in terms of both appearance and structure distribution. Therefore, one may exploit this prior knowledge to locate the positions of the eyes in the Bayesian framework. For example, Kawato and Ohya [50] proposed an eye tracking system through quickly localizing the 'between-eyes' region.

2.1.4. Active infrared lighting characteristics

One of the most effective ways to deal with the lighting changes is the active near-infrared (NIR) imaging techniques, due to the fact that under the active IR lighting, the pupil and iris will show different illumination properties. In particular, the pupil usually has a larger reflection rate than the iris, resulting in a bright spot at pupil position. This bright spot is a good indicator of the pupil and can be used for eye localization [117,116] (cf. Fig. 5). In practice, a near-infrared light source with a wavelength from 780 to 880 nm will meet the requirements of most in-door application scenarios. Due to its robustness against visible lighting changes, this method has been widely used in driver fatigue detection and face recognition [64]. It should be mentioned, however, that there exist several conditions (restrictions) that must be satisfied to ensure good performance, such as opened eye states and the on-axis light, together with NIR imaging hardware. Li et al. [64] observed that although active NIR makes the appearance of the eyes robust to different lighting conditions in general, the glasses and eye states may cause trouble for precise eye localization (cf. Fig. 5), and they proposed a tree-structured detector to carefully address this issue [64].

2.1.5. Discussion

In this section, we summarized several major eye localization methods which measure different characteristics of eyes, including the shape, strong intensity contrast, context information, and active NIR lighting characteristics. It is worth noting that most of these, except the active NIR technique, are developed at the early stage of eye localization research and their own limitations become more pronounced under the complicated uncontrolled conditions where the characteristics may not be reliably measured any more. To deal with this problem, most recent eye localization methods resort to more advanced statistic methods. This is the main topic of the next section.

2.2. Learning statistical appearance model

In contrast to aforementioned methods where eye-characteristics with intuitive visual meanings are measured, methods reviewed in

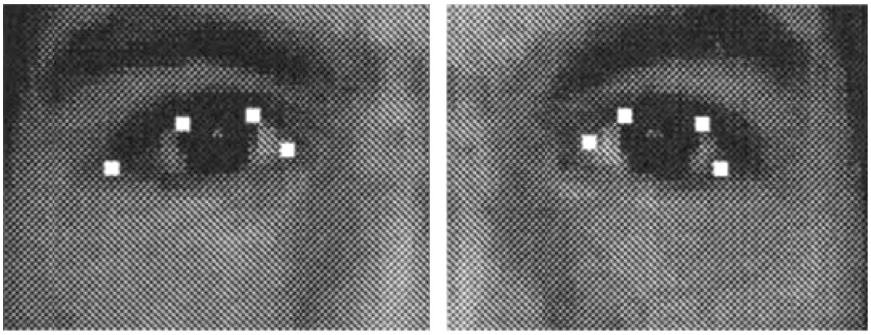

Fig. 4. Corners located for eyes [58].

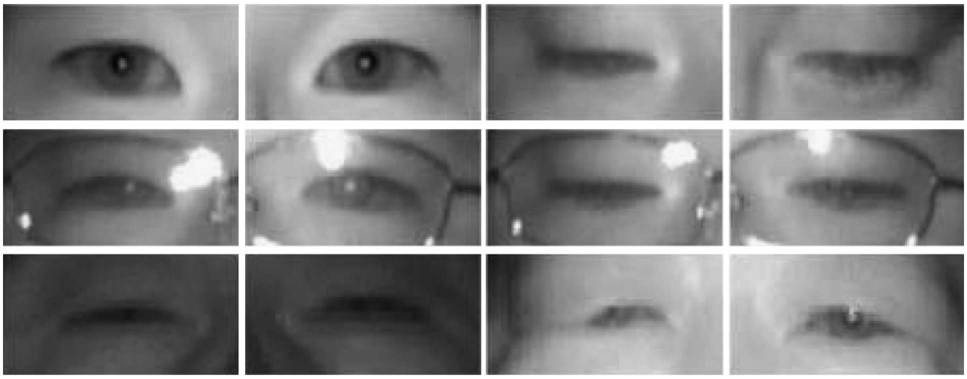

Fig. 5. Eyes examples under active near-infrared lights [64].

this section focus on statistical models using photometric appearance features extracted from cropped eye patches. The appearance based methods potentially use more information than those based on eye-characteristics, since the image content of eye patches contains both information about eye-characteristics (e.g., eye shape) and other relevant information that may be ignored or not easy to be measured. These lead to a more robust technique.

In what follows, we will first review the various appearance feature sets and then go on to review the statistical models built upon them.

2.2.1. Appearance features extraction and representation

To build a reliable statistic appearance model, one has to decide at first where and how to obtain a proper appearance representation. Most methods extract appearance features from a small patch image of eyes. Some subtle but important considerations are needed here. For example, how large should the patch be? Should it cover the pupil, the iris, and the eyelids only, or should it cover much context like the eyebrow? Unfortunately, most current discussions on these topics are empirical in nature. Intuitively, a relatively larger eye patch contains more discriminative information and thus tends to reduce the risk of false positive, but this will be at the cost of losing generalization capability as they will be less likely to be good representatives of eyes. In [108], it is shown that incorporating the eyebrow into the eye patch leads to better localization performance on FERET [82] and YALE [7] face databases. But in general, a cross validation procedure is recommended to search the best suitable setting [22].

Given an eye patch, several visual feature sets, from middle to high level, can be extracted from it. Each feature set is simply a transformation of a set of neighboring raw pixel values, designed to be invariant to certain changes. Since no single feature descriptor will satisfy all the needs, selecting the ones to use in practice is mostly application-driven and the factors that need to be taken into account mainly include: (1) the invariance properties it provides, e.g., to lighting changes or to variations in scale, orientation, and other affine transformations; (2) the information encoded and the discriminability preserved; (3) the computational efficiency. The first criterion helps to clarify which kind of change one wishes to compensate for after preprocessing the patches. The second one is essential to the subsequent statistical modeling/processing, while the third one concerns the computational aspects of the feature set.

Popular feature set descriptors include those in frequency domain, e.g., Harr wavelets features, Gabor features, and those in the spatial domain, especially various gradient-based features, such as Local Binary Patterns (LBP, [2]), Scale Invariant Feature Transform (SIFT [66]) and Gradient Location-Orientation Histogram (GLOH [73]). Some of them are illustrated in Fig. 6. We are not aware of any comparative work concerning the effect of different feature sets on the task of eye localization, but in the more general context of computer vision, Mikolajczyk and Schmid [73] reviewed some recently developed view-invariant local image descriptors and experimentally compared their performance. They found that GLOH [73] and SIFT [66] are the two best performed descriptors among others in their settings.

In [55], the authors presented a method based on the multi-scale LBP feature sets for eye localization in low and standard definition content (see [51] with multi-scale Gabor feature sets). In general LBP feature [2] is good at coding the details of appearance and texture, whereas Gabor features encode global shape and appearance over a range of coarse scales. Both representations are rich in information and computationally efficient and hence are widely used in the field of facial analysis. It is worth mentioning that eye localization is a complex task for which it is useful to include a lighting normalization stage before feature extraction and combine multiple types of feature sets [91].

The multi-scale method used in [55] is helpful to alleviate the problem of choosing the right patch for eye representation by properly fusing multi-scale context information. A similar idea is adopted in [107], where a pyramid of dictionaries is offline built at multiple scales (see Fig. 7). For online localization, the dictionaries are sequentially applied from the largest scale to the smallest one. This fitting procedure, however, involves solving an l_1 problem and is generally time consuming.

Besides usual feature descriptors which directly apply certain linear or nonlinear transformations on the given eye patches, recently there has been a trend to construct feature sets statistically [101,76,13,106,16,88]. Usually these feature sets are obtained by embedding the aforementioned middle-level feature sets into another feature space with desired properties (e.g., compact,

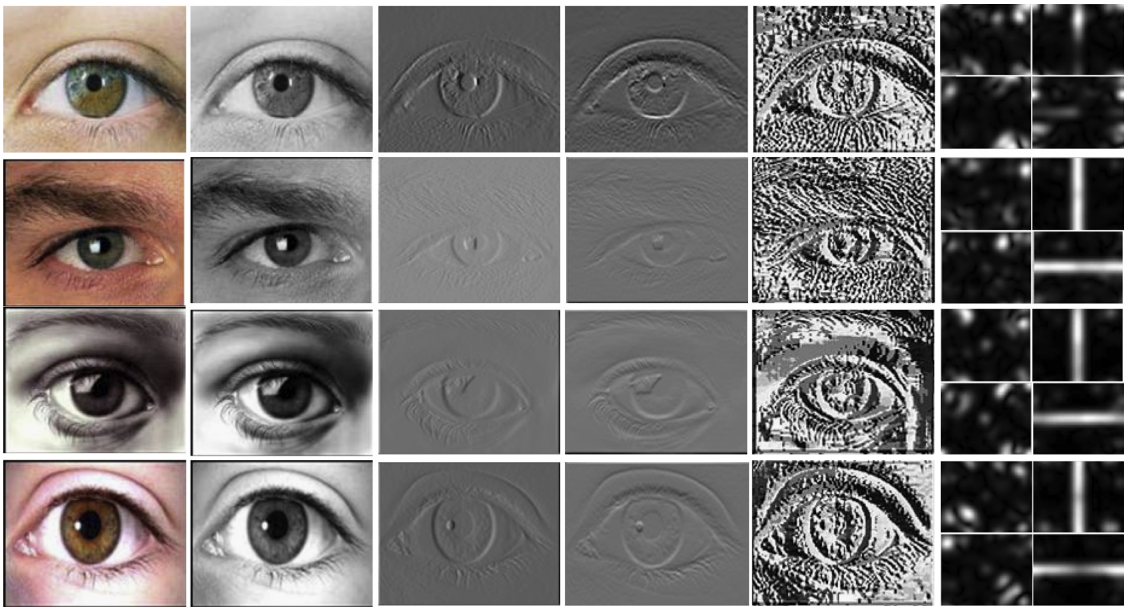

Fig. 6. Illustration of feature sets for eye patterns. From left to right: color image (best viewed in the electronic version), gray intensity, gradient images (in horizontal and vertical directions), Local Binary Patterns, and Gabor features (four directions and one scale).

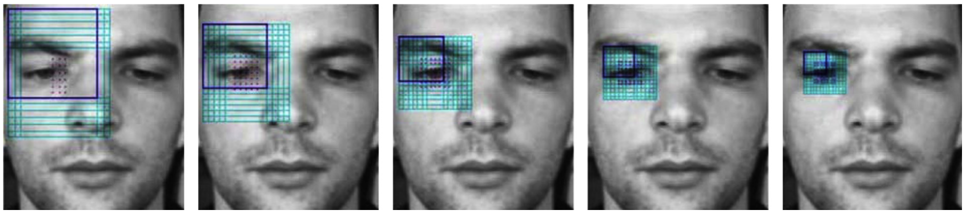

Fig. 7. Examples of multi-scale dictionaries [107].

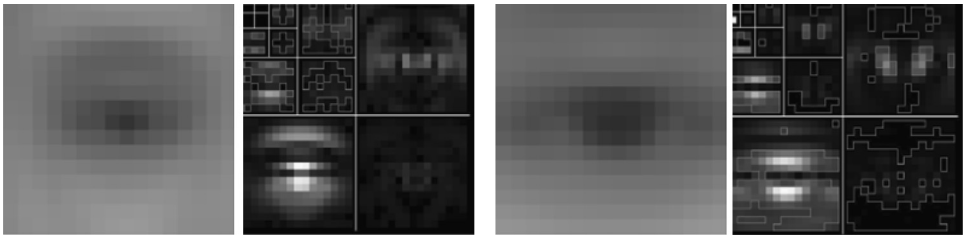

Fig. 8. Wavelet decomposition and feature selection for hierarchical SVM eye model [13].

discriminative, etc.). These welcome properties are introduced by learning from training examples either discriminatively or generatively. For example, in [101] the authors presented a method which learned an optimal discriminant feature space from training data by minimizing the empirical Bayes error using nonparametric discriminant analysis. Since the learned feature space is tuned to highlight the contrast between the positive eye samples and the false ones, it is expected to be able to characterize eye patterns better. In [76], the authors proposed an energy-based framework to jointly perform discriminative feature sets learning and eye localization, in which the parameters of statistical feature sets are optimized to maximize the generalization performance of an SVM classifier. Similarly, [13] proposed to group different sets of features hierarchically and built a cascade of SVM classifiers on them (see Fig. 8).

Compared to common feature descriptors, the major advantage of statistic-based feature sets lies in their stableness in handling uncertainty of image data, but at the cost of more computational efforts and of the needs for plenty of representative training data to ensure good performance.

2.2.2. *Statistical appearance models*

Even armed with a proper feature representation which accounts for certain variability associated with target occurrences, one still needs to build a final classification stage that can handle residual variability and learn effective models from relatively few training samples. This section describes some popular methods for this purpose.

Popular classifiers are constructed in either a generative or a discriminative way [8]. The generative methods try to recover the class conditional probability distribution of eyes and use it to see how likely a testing patch is generated from it. Due to the high dimensionality of eye patches, the conditional density is best estimated on a lower dimensional manifold, which can be found using manifold learning methods like Locally Linear Embedding (LLE) or Principal Component Analysis (PCA) [81].

Generative methods: a successful generative method is the probabilistic PCA model for object representation proposed by Moghaddam and Pentaland [74], in which the training samples are first projected into their column space using PCA, then a Gaussian density is estimated there to model the class conditional distribution of positive samples. Evergingham and Zisserman [22] extended this idea to include non-eye model $p(x|\overline{e})$. Then the prediction of a new patch x can be performed by looking at the margin (or log-likelihood ratio) between the outputs of the two models given $x: llr(x) = \log p(x|e) - \log p(x|\overline{e})$, where both eye model $p(x|e)$ and non-eye model $p(x|\overline{e})$ are assumed to be Gaussian. Even though more complicated generative model, e.g., Gaussian Mixture Model, could be applied, this simple method yields very good localization performance on several databases consistently, which clearly illustrates the power of generative methods in this challenging task. To ensure accurate distribution estimation, however, a relatively large number of labeled training samples are desired.

Discriminative methods: on the other hand, discriminative methods aim to find the discriminant function between eye and non-eye classes directly, in the form of decision surface, separating hyperplane, or threshold function. In this way, the problem of eye localization boils down to a binary classification problem. Typical classifiers for this include Support Vector Machine (SVM) [22,93,92], AdaBoost [98,101,112,67], neural networks [86], and so on. Among others, SVM is widely used in practice due to its ease of use and its good generalization capability [93,92]. However, in many cases a good generalization performance is only guaranteed with enough number of support vectors coupled with nonlinear kernel function, which may significantly increase the computational costs and memory when testing. However for certain specific kernels, the computation could be made very efficient [71] as well as the training method [70]. Alternatively, it is possible to adopt an efficient branch-and-bound optimization strategy to reduce the number of matching without sacrificing the accuracy of classifiers [59].

Another way to achieve a good tradeoff between accuracy and efficiency is the boosted cascade of features [98,67,101,77, 112,106], which adopts the ‘coarse-to-fine’ strategy to select a bunch of discriminative features and use them to construct more powerful classifiers increasingly. The testing step is very efficient since each level of the cascaded classifiers consist of a linear combination of a few simple weak classifiers and only those patches with high likelihood will be passed on for further examination. To enable reliable feature selection, sometimes the candidate patches are set to include both eyes [112,67]. To deal with the problem of eye localization in low quality face images, [109] introduced a quality adaptive cascade that works in a probabilistic framework (P-Cascade), which allows all image patches to contribute to the final result with some probability.

Optimizing the localization accuracy: compared to generative methods, discriminative methods are more efficient in exploiting different types of visual features from samples without assuming any dependence between them, but they may not behave so robustly against the problem of probability drifting as their generative counterparts. The reason for this is that for the discriminative methods only the area of decision boundary is important while the interior distribution of each class is completely ignored by the model. Evergingham and Zisserman [22] compared several classifiers including both generative and discriminative ones and found that the simple Bayesian classifier outperforms several discriminative methods including SVM and AdaBoost method. One possible explanation for this is that the discriminative methods are sensitive to small localization errors in the ground truth [22]. Moreover, since the objective of the discriminative method is set to optimize the classification accuracy rather than localization accuracy, the trained classifier may not give the maximal response at the right object location.

There are several approaches to address this problem. The simplest one is to explicitly tell the classifier which training examples are confusing, i.e., they look like but are actually not the target of interest. This can be done by the careful construction of the negative set. Those examples can either be sampled around random points in the iris neighborhood specified beforehand [76], or be generated in a boosting fashion by identifying them using the current classifier [92].

Another natural way to explicitly incorporate the positions of the eyes is to formulate the task of localization as a regression rather than a binary classification problem. In this setting, the training data are given by a set of input images and their corresponding eye positions, and the goal is to learn a regressor which maps from the input image to the predicted eye position (in some invariant coordinate system, cf. [61]).

Different regression methods can be adopted according to how to specify the eye localization. In [10], each training image is assigned a correlation image which is synthetically generated with a bright peak at the center of the target (e.g., the left eye) and small values everywhere else. This correlation image gives not only the locations of the eyes but also the corresponding desired response to be regressed. The regression is implemented by producing a correlation filter that exactly transforms each input image to its correlation image (see Fig. 9). The localization is performed by correlating a test image with the learned filter and selecting the global maximum in the correlation outputs. This testing procedure is very efficient without resorting to the time-consuming sliding window search. A similar method is adopted by Hefin et al. [38]. To deal with the ambiguity from multiple peak responses, they warp the face image using the candidate response pairs and select the one that leads to a face image with best quality [38].

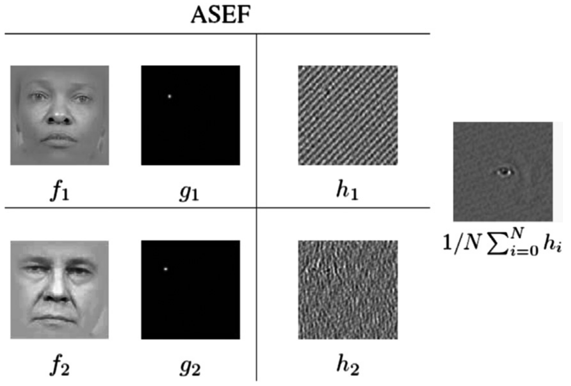

Fig. 9. Illustration of the ASEF method: The image f_i is an image in the training set and g_i is the corresponding desired filter output. A correlation filter h_i is produced by in the Fourier domain that exactly transforms f_i to g_i. The final correlation filter is produced by taking the average of many Exact Filters [10].

The above filter methods implicitly learn a regression model for eye localization. Alternatively, as proposed in [22], one can directly construct a linear regression model in the Hilbert space from an input image to its eye localization, i.e., a 2D coordinates in the image. To ignore the irrelevant variation, a square region centered at the mean over the training images of the ground truth eye positions is extracted for regressing. This method is shown to work better than SVM and AdaBoost-based methods, especially on the uncontrolled WWW images [22].

Moving forward with this line of research, Blaschko and Lampert [9] proposed a general object localization method which specifically optimizes localization accuracy within a large margin regression framework. In particular, instead of representing the positions of the eyes as 2D coordinates, a bounding box around the eye region, which is regarded as a more flexible and useful structure to be regressed in the output space, is adopted. With the help of joint kernel map, the training procedure is formulated under the framework of structured predicting [4]. This method is shown to boost the accuracy of object localization and has the potential to handle partial detections by properly evaluating the usefulness of image regions that contain portions of the object. Besides structured regression, Gaussian process can also be used for the purpose of object localization in general and eye localization in particular [44].

2.2.3. *Discussion*

In this section, we have summarized the motivations and popular techniques of appearance-based methods. Incorporating the feature extraction methods and advanced statistical techniques, appearance-based model can effectively explore a wide range of appearance variations of the eyes, and achieves good generalization ability on unseen eye images. However, it is still useful to consider the eye characteristics-measuring-based methods and the appearance-based methods under the same framework, and such a hybrid method is a promising direction for localizing eyes' positions under uncontrolled conditions. We will explore such a possibility in the next section.

2.3. *Exploiting structural information*

Unlike photometric appearance features which capture the visual aspects of the eyes, the spatial topological features characterize the pattern of the eyes in a different way. In particular, the eye consists of components such as eyelids, iris and pupil, and these components have regular structural relationships. Such structural features are less affected by environmental conditions than appearance features. And clearly, the two types of features (i.e., appearance and structural features) are complementary to each other in depicting eye patterns. Since the eye is relatively smaller than the face, it is also convenient to model eyes structural features in the face context.

Pictorial Structure Model (PS, [28]) is a typical method in the line. The basic idea is to decompose an object as a set of parts and then use a graph structure to model the topological relationship between them. After a model like this is constructed, it can be used to localize the object of interest by measuring the appearance fitness and structural deformation simultaneously,

$$L^* = \arg\min_{L}\left(\sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i,v_j)\in E} d_{ij}(l_i,l_j)\right) \tag{1}$$

where $m_i(l_i)$ is the similarity for part i at position l_i according to appearance model, and $d_{ij}(l_i,l_j)$ measures the spatial structure confidence for part i and part j located at position l_i and l_j correspondingly. The challenge is how to solve this energy minimization problem efficiently so as to find the optimal

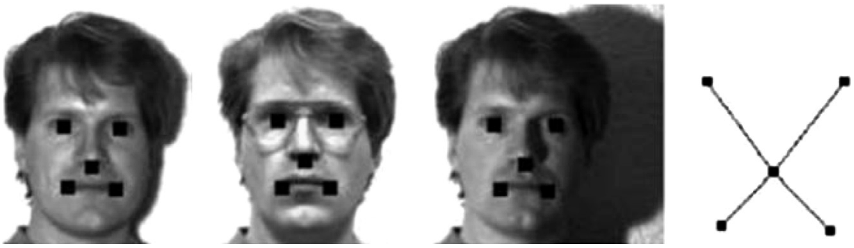

Fig. 10. Three examples from the training set showing the locations of the labeled features and the structure of the learned pictorial model [25].

configuration of parts positions L^*. For this purpose, Felzenszwalb and Huttenlocher [25] presented an efficient method to automatically estimate the topological structure by simplifying the graph structure as a tree and restricting the form of connections between parts as a linear one rather than quadratic in the number of possible locations for each part (Fig. 10). In a recent development [26], they treat the locations of the parts as latent variables and learn them in a discriminative framework. This essentially reformulates the energy function of Eq. (1) by canceling the second term and merging it within the appearance model in the first term to make it more computational tractable. This method has been successfully applied to the task of object detection although in principle it can also be used for the problem of eye localization.

To this end, Tan et al. [92] proposed a method which enhances the pictorial model in two aspects for the purpose of eye localization. First, to avoid estimating the multi-modal distribution of eye appearance under the complex real-life scenarios, a discriminative method is used to substitute the simple generative model used in [25]. This is in spirit similar to the latent SVM method in [26], but the difference lies in that it is the label rather than the location of each part that is treated as a latent variable in [92]. This helps the model capture the meaning of each part (i.e., identifying them respectively as eyes, nose, etc.). Second, an improved structural description method is introduced to make it more robust against affine transformations as rotation, scale and translation. This method is tested on the challenging LFW data set [42] with promising results. Campadelli et al. [15] also proposed to use the information of other facial features (mouth in their case) as further constraints to reduce the number of false positives. Their methods are tested on several public databases (cf. Table 3).

Active Shape Model (ASM) [19] is another representative method to model the structure information of an object. In ASM, the shape of an object is described in terms of shape vector, which is a set of coordinates of these landmark points arranged in a predefined fixed order. The landmark points need to be manually annotated, usually on the contour and the key components of an object. ASM then uses these annotated training data to build a statistical model, so as to guide an iterative search on a test image for its shape vector which conforms to all the shape variations possible for this particular type of object (e.g., face). This model is suitable for locating multiple facial features simultaneously under high image resolution.

One of the major advantages of structural methods lies in that it provides a nice mechanism to infer the location of an object by estimating the locations of its parts. Hough transform based methods have been successfully adapted to this purpose recently [61–63,5]. Implicit Shape Model (ISM, [62,63]) is a typical voting-based structure model, in which the shape model is not explicitly constructed but represented loosely in terms of a bag of patches. Different from the usual bag of patches model, the key idea of ISM is to maintain a spatial occurrence distribution for each visual codeword, such that it can be used not only for the representation of local appearance but also for casting votes for possible positions of the object center as well. This Hough voting mechanism effectively bypasses the difficulty of optimization problem encountered in many structured methods (cf. Eq. (1)). Furthermore, it is robust against partial occlusion and large appearance variations. Recently, Barinova et al. [5] provided a new probabilistic formulation for Hough voting and Lehmann et al. [61] interpreted Hough voting as a dual implementation of linear sliding window detection, which leads to a fast implementation of ISM. Kozakaya et al. [54] proposed to improve the stableness of voting space using the dense sampling strategy.

2.4. Discussion

In this section, we have reviewed major automatic eye localization techniques which involve feature extraction, texture representation, model training and optimization algorithm. In principle the methods reviewed can be also used for related tasks such as eye detection, eye tracking, gaze estimation and blink detection aforementioned in Section 1. Nevertheless, it is worth mentioning that extracting and representing task-specific feature sets is extremely important for good performance. For example, extracting motion-based features, which is not reviewed here (but cf. [37]), is crucial to the task of eye tracking. On the other hand, for a complex task like eye localization, it is difficult to address all the variations using a single type of feature set or localization method. Therefore, taking careful consideration about the system architecture from a global view is extremely useful, and we will discuss this further in the following section.

3. Towards the development of a robust eye localization system

As mentioned in Section 1, the eye appearance and shape signals are subject to many kinds of undesirable variations, and any mismatch between the training and testing conditions may dramatically decrease the performance of eye localization. Tracking this mismatch and related variations is the main focus of eye localization research, and due to its complexity, it is often the case that no single modality is enough. In this section, we first present a global system architecture for eye localization (Fig. 11) and then have a closer look at possible strategies to improve the robustness of eye localization under this architecture.

3.1. The global system architecture for eye localization

Due to the complexity of the eye localization, it is better to use a divide and conquer strategy to handle different variations at different stages. Inspired by [24], we give a global architecture for eye localization (see Fig. 11). Note that this is only to illustrate the general pipeline for an eye localization task and it is possible to instantiate different systems according to this architecture but not all components are mandatory in a practical system. For example, the pictorial structure model [25] and ASM model [19] do not rely on a careful face normalization to build their shape model.

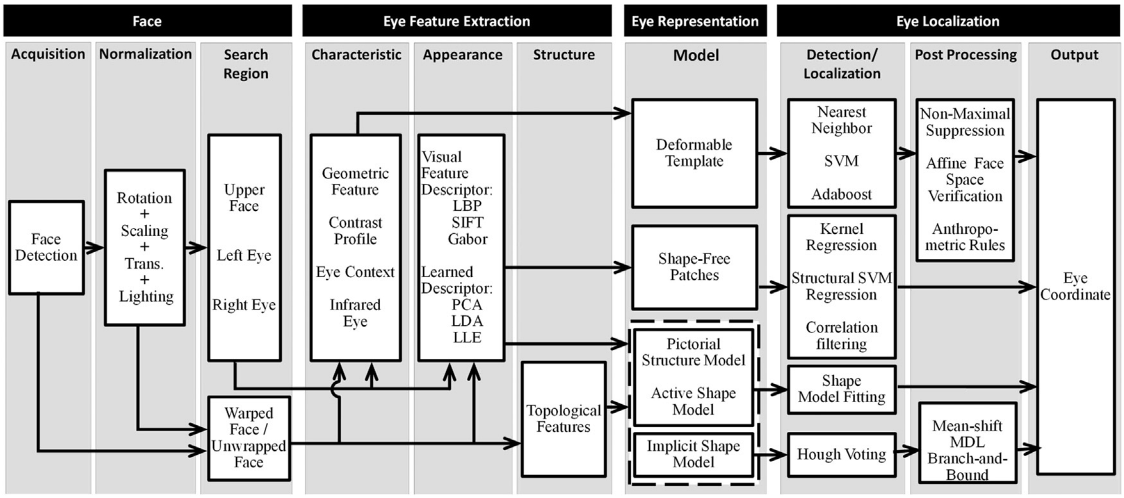

Fig. 11. A global system architecture for eye localization.

As shown in Fig. 11, the architecture is roughly divided into three components, i.e., face processing, feature extraction and representation for eye model, and eye localization. The face processing component can be thought of as a preprocessing stage for eye feature extraction. Meanwhile, the problem of the accuracy and efficiency tradeoff is essential for any practical eye localization system. Finally, before the locations of the eyes are actually outputted, some postprocessing which combines multiple searching results of eye detectors needs to be done properly. Next we will have a closer look at these issues.

3.2. *Preprocessing methods*

The goal of preprocessing stage is to facilitate the task of feature extraction by removing noise information and narrowing down the search region. The first step for this is to locate and segment out faces from complex scenes with cluttered backgrounds. Although this can be done by hand, in most cases a modern face detector [98] will do it quickly and accurately. The output of a face detector is a rough face region possibly with different scales, rotations and lightings. Therefore it might be useful to remove these variations from the acquired faces prior to further analysis.

Handling pose variations: pose variations due to scale changes and in-plane rotations are commonly handled by geometric transformation. In many face recognition applications this is performed based on the coordinates of some predefined landmarks like eyes or the nose, which are unfortunately assumed to be unknown in eye localization. Therefore warping methods which do not rely on such information are preferred for eye localization. In the ASM method [19], a set of face images are simultaneously aligned using an iterative Procrustes algorithm which essentially aims to find the best affine transformation under a least square type criterion. Alternative normalization methods are based on the information theory. In [41], the authors proposed an unsupervised method for face alignment with no need of landmarks by looking for a transformation for each face such that the empirical joint entropy is maximized after transforming.

It is worth mentioning that the function of geometric normalization method should not be overstated and there exist alternative ways for this purpose. For example, to handle scale changes, one can scan the face images of different sizes by subsampling instead of doing geometric normalization. In practice, out-of-plane rotations with limited angle can be handled through geometric normalization, but the problem of compensating missing information for distorted face due to large in-depth rotation is quite difficult if no further prior knowledge is provided.

Image enhancement: pixel values are raw intensity features and could be easily affected by illumination variations. Traditional approaches for dealing with this issue can be broadly classified into three categories: appearance-based methods, normalization-based, and feature-based methods. In direct appearance-based approaches, training examples are collected under different lighting conditions used to learn a global model of the possible illumination variations, for example a linear subspace or manifold model, which then generalizes to the variations seen in new images [6,60]. Direct learning of this kind of model makes few assumptions but it requires a large number of training images and an expressive feature set, therefore it is seldom used in the application of eye localization.

Normalization based approaches seek to reduce the image to a more ‘canonical’ form in which the illumination variations are suppressed. Histogram equalization is one simple example, but purpose-designed methods often exploit the fact that (on the scale of a face) naturally-occurring incoming illumination distributions typically have predominantly low spatial frequency and soft edges so that high frequency information in the image is the predominant signal (i.e., intrinsic facial appearance). For example, the Multi-scale Retinex method of Jobson et al. [49] cancels much of the low frequency information by dividing the image by a smoothed version of itself. More recently, Gross and Brajovic (GB) [33] developed an anisotropic smoothing method that relies on the iterative estimation of a blurred version of the original image. In [91], a signal processing approach motivated more by bottom-up human perception than by Retinex theory is presented. Overall, these methods are quite effective and efficient but the problem of handling spatially non-uniform variations remains an open problem.

The third approach extracts illumination-insensitive feature sets directly from the given image. These feature sets range from geometrical features [11] to image derivative features such as edge maps [1], Local Binary Patterns (LBP) [90], Local Ternary Patterns (LTP) [91], Local Phase Quantization (LPQ) [80], Distance Vector Field (DVF, [3]), Gabor wavelets [102], and local autocorrelation filters [35]. For instance, Nanni and Lumini [75] built a well performed eye classification module based on multiresolution LTP [91] and LPQ [80] descriptors. Yang et al. [108] designed a ‘Gabor-Eye’ representation for eyes which exploits the frequency and direction sensitivity properties of Gabor feature sets. Qian and Xu [84] selected some Gabor-transformed images to reconstruct an intermediate representation suitable for eye localization, in which the signal of the eye region is enhanced and the influence of lighting variations is reduced. To stabilize the Gabor filter direction selection and increase the robustness to the accessories appearing on the eye-and-brow region, Xiong et al. [104] present a method based on the face tilt angle detection and adjustment techniques. Alternatively, Crowley et al. [48] used the first and second order Gaussian derivatives as filters to extract feature vectors which are robust to chrominance variations of light.

Shrinking search areas: narrowing down the search region is helpful to improve both the efficiency and the effectiveness of eye localization in general. The search region plays the role of hypothesis space, reflecting the prior knowledge about the place where the eyes are assumed to lie. The size is crucial in setting the hypothesis space—a small one may simply not contain eyes and a large one may lead to unnecessarily computational cost. Usually a face detection operation is first performed to locate the face region then a search window is put at the eye region based on knowledge from the training set. This strategy works well in most cases, but care must be taken when the location of the eye is unpredictable (e.g., being occluded due to pose variations). In the later cases, expanding the search space to include the entire image might be a better choice [10]. Alternatively, one may progressively use different templates with different scales to narrow down the search space. For example, in [29], a multi-view face detector is first used to estimate the rotation state of the face, based on which a coarse-to-fine strategy is further applied to locate brow-eye regions and the eyes respectively.

In practice, some heuristics may be used to locate the search window. The simplest one is based on the fact that eyes are almost always lying in the upper part of a face and the intensity of eye center always has lower contrast compared to its surroundings. Such cues are very informative to give a coarse prediction of eye positions [67]. Actually one can even try to learn the prior distribution of coordinates of the eyes given a large number of samples [55]. Another useful heuristic is that the eye regions have low intensity, low red chrominance, and high blue chrominance, when compared to the forehead region of the face. This heuristic is commonly used to segment the eye region from the face image, followed by some morphological operations (e.g., dilation and erosion) to enhance certain interesting pixels of the eye region [40,32,100]. Other eye features described in Section 2.1, such as intensity contrast and edge map, can also be used to roughly locate the eye region efficiently. In [104], candidate eye region is extracted through the horizontal intensity gradient integral projection and Gabor filtering. Campadelli and Lanzarotti [12] uses a man-made eye template to roughly locate 10 candidate eye regions and selects the best one based on the symmetry condition and majority voting methods.

3.3. Accuracy and efficiency tradeoffs

Localizing eyes from faces in real time is crucial to many practical applications. The efficiency depends not only on the eye prediction model but also on the search and matching strategy applied. Sliding window technique is commonly used for object detection, while the strategy of ‘coarse-to-fine’ is usually adopted to achieve better tradeoff between accuracy and efficiency. To solve the high accuracy and low complexity dilemma, and considering that most search regions are backgrounds, it is wise to use discriminant model but with less complexity to quickly filter out noise regions and use more complex model on most probable eye regions. Campadelli et al. [13] build two cascading SVM classifiers with different complexities and accuracies to achieve accurate eye localization in coarse-to-fine way.

On the other hand, more efficient eye model design is helpful. In the above introduced cascading classifiers, the last level classifiers usually have high accuracy but also have high model complexity; the time cost is still expensive even evaluating in the reduced search space. Many factors contribute to the complexity of models and there are many ways to measure the complexity in literatures, such as VC dimension and Minimum Description Length (MDL), but it is generally agreed that the degree of complexity increases with the number and magnitude of the effective parameters involved. One way to achieve high accuracy but with controllable complexity is to linearly integrate a few of simple classifiers, as does in the AdaBoost framework for object detection [98]. Similarly, [30] linearly combined a set of linear SVMs, which are shown to have equivalent accuracy to nonlinear SVM, but with less model complexity. Saragih et al. [87] extended this idea to implement an efficient ASM model for facial feature localization.

Recently, compress sensing techniques originated from the field of signal processing are popularly applied to select the most informative feature set and hence reduce the complexity of the final model in object detection [105]. Others propose to use machine learning methods like Fisher discriminant analysis and SVM to extract/select discriminative feature sets so as to improve the efficiency and effectiveness of eye localization [101,15].

The drawback of sliding windows technique is the huge size of the search space. To overcome this, researchers try some interesting ways that go beyond sliding window technique. For instance, Lampert et al. [59] propose the *Efficient Subwindow Search* method, which adopts a branch-and-bound scheme to find the global optimum of the quality function over all sub-images, and runs in linear time. This method has recently been extended in the context of Hough voting for object localization [61]. On the other hand, some methods are motivated by the efficient visual attention mechanism of human visual system. For instance, inspired by the concept of visual routines [95], Huang and Wechsler [43] proposed an eye localization method using navigational routines. They automated the derivation of such routines using evolution and learning, which effectively filters out unlikely eye locations and limits the use of the more expensive classifier. Alternatively, [34] implemented the attention mechanisms using the phase spectrum information of Fourier transform. The Correlation filter methods [68,69,10] effectively avoid exhaustively searching in the image domain by taking advantage of the fact that the correlation computation in the image domain is equivalent to element-wise multiplication in Fourier domain.

3.4. Postprocessing methods

For those methods localizing eyes localization through detection [22,93,101,112,67], a postprocessing step is usually needed to further evaluate the most probable eye positions or return feedbacks to final decision. One reason for this is that they commonly use a classifier like SVM or AdaBoost to find the best decision boundary which discriminates the positive samples from negative ones in terms of classification accuracy rather than optimizing the objective to give highest score to the true object of interest at the right position. Besides, the ground truth with small localization errors may seriously mislead the learning procedure of classifier [22]. Therefore, we should not over-interpret the output of the classifier—it is just an indicator about whether or not eyes are probably present at some positions, but not necessarily their exact positions. Actually, due to the spatial dependence between candidate patches, it is very likely that the eye detector yields several positive responses with similar scores in the vicinity of some locations. Consequently, postprocessing like non-maximal suppression is needed to remove such ambiguity. However, it is worth mentioning that unlike postprocessing in the general context of object localization where multiple instances may exist simultaneously in one image, for eyes we know a prior that only one eye exists in one search window.

One of the best known non-maximal suppression methods is developed by Viola and Jones in their face detector [98], in which highly overlapping detections are merged into one representative face using a clustering algorithm. To this end, various heuristics can be used to fine localization and false alarm dismissal. Garcia and Delakis [31] interpreted the fine localization as a local search procedure and proposed to accumulate evidences for positives in the scale space, based on the observation that true objects usually give a significant number of high positive responses in consecutive scales, which is not often the case for negatives. However, in some Hough voting methods (e.g., ISM [63], PRISM [61]) the searching space is modeled as continuous rather than discrete, which makes the searching become much more complicated. Many efforts have been made to address this issue [63,61] as reviewed in the previous sections.

The anthropometric measures are also frequently used for postprocessing, due to the fact that the spatial structures between two eyes and other facial features could pose rather strong (and useful) constraints on the locations of the eyes. Actually, most of the methods reviewed in Section 2.3 can be thought in this way—they simply combine the task of eye detection and fine localization into one single objective function. But many methods do perform their postprocessing explicitly. In [112], the authors used an AdaBoost detector to segment the eye region and then a fast radial symmetry operator to finely locate the centers of the eyes. However, this geometric constraint also implies that it can only be used for open eyes. To leverage on the effectiveness of postprocessing, [93] relaxed the constraint criteria of the AdaBoost eye model so as to generate more candidates for further verification. Instead of using a single-eye patch, patches containing both eyes could be utilized to train the SVM classifier as well [67], which implicitly exploits the spatial dependence between two eyes. To eliminate small spurious detected regions, Crowley et al. [48] used a connected components analysis algorithm and computed the bounding box around. Regions with a small bounding box are eliminated.

Hamouz et al. [36] described a different method for postprocessing by checking the consistency of candidates in a lower dimensional space. In particular, they first transform all the data into a canonical space and then find the best candidate there using an SVM. The canonical space is trained with ground truth and hence serves to encode the constraints about what the 'good' samples should look like. This idea is similar to that in ASM [19], which recommends a better shape vector for face alignment in a pre-learned shape space given the current fitness.

4. Performance evaluation

In order to evaluate eye localization model objectively and make a fair comparison among different methods, ideally performance should be reported on the representative benchmark database and follow a standard and reasonable experiment protocol. However, in reality many algorithms are evaluated in different ways with variations in databases, measure metrics, training samples, testing samples, etc., which makes it almost impossible to directly compare eye localization results found in the literature. Nevertheless, we still listed some of reported performance in Table 3 for reference. Next we will review the performance measure metric and the major face databases used for evaluation.

4.1. Measure metric

The design of the ground truth: the localization accuracy is basically measured by the distance between the predicted position and ground truth location of the eye. The ground truth is defined as a representation of the agreed correct result of the ideal eye localization method and forms the basis for all performance comparisons among the methods to be evaluated. Therefore, the design of the ground truth is crucial.

We note, however, that this is task-specific. Take the task of gaze estimation for example. It would be reasonable to define the ground truth as the center of the pupil. While in the task of automatic face recognition, the procedure of eye localization mainly serves to provide the anchor information for the subsequent geometric normalization. But under the side-looking gaze direction while keeping frontal face pose, pupil positions will deviate away from eye center. In this case face will be misaligned if we still take pupil positions as reference points (see Fig. 12). What is more, labeling pupils is error-prone under uncontrolled conditions featured with closure eye states, poor image quality, glasses reflection, and eye partial occlusion. In these cases it would be better to take the actual center of the eye as its position.

Recently, Kostinger et al. [52] propose a method which defines facial landmark positions based on a rigid 3D face model (see Fig. 13). One advantage of this method lies in its capability to deal with the variability due to face pose changes. It has been successfully used to label the centers of the eyes and other facial landmarks on a large scale face database [52].

Localization error measurement: after defining the ground truth, the next step is to measure the localization error. Among others, most commonly used measurement is the normalized eye localization error proposed in [46], which is defined in terms of the eye center positions according to

$$d_{eye} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|}, \quad (2)$$

where C_l and C_r are the ground-truth positions and d_l and d_r are the Euclidean distances between the detected eye centers and the

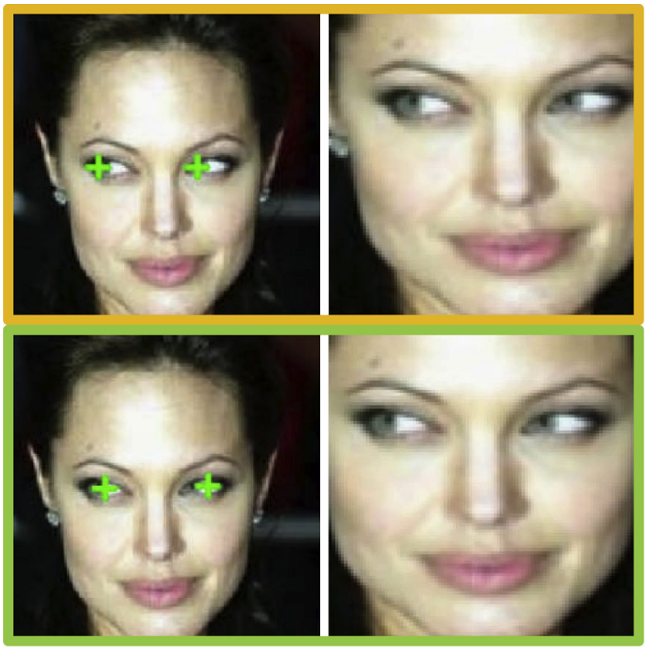

Fig. 12. Illustration of the influence of eye labels on face alignment. The ground truth is defined as the center of the pupil (above row) and the center of the eye (below) respectively.

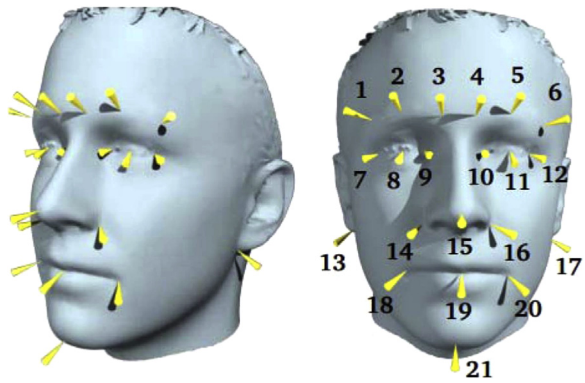

Fig. 13. The 3D model used to define the centers of the eyes and other facial landmarks [52].

ground-truths. The definition shows that this measure is the worst location results analysis standard, and the normalization term in this formulation can eliminate unreasonable measurement variations caused by variations of face scales and image resolutions. In practice, for eye detection, usually $d_{eye} < 0.25$ is required, but for eye localization, $d_{eye} < 0.05$ or $d_{eye} < 0.1$ is more meaningful.

For the face under out-of-plane rotation, the two eyes distance cannot reflect the actual face scale, and the normalized localization error measurement may be biased. In such case more general evaluation measures such as mean and variance could be used [22]. Alternatively, depending on where the localization system is used, the performance can also be evaluated by checking how well the interested task is solved (e.g., in terms of the improvement on the accuracy of face recognition [101]).

4.2. Databases and performance evaluation

Although there are many face databases [113] developed, they are not originally aiming for the problem of eye localization and are collected under well controlled laboratory conditions with normal lighting, neutral expression and high image quality. Despite these drawbacks, they contain many kinds of interesting eye pattern variations and hence are widely used for the eye localization research nowadays. Among others, popular face databases include FERET [82], FRGC (Face Recognition Grand Challenge, [83]), JAFFE [20], BioID [23], LFW (Labeled Face in the Wild, [42]) , FaceTracer [56] and so on. Fortunately, most of these databases are shipped with the ground truth of eye positions and thus can be used for model training. Alternatively, plenty of sources for reliable labels of eyes are publicly available in the area of face alignment. For details, see Kostinger et al. [52]. More information about published miscellaneous face databases can be found in the Internet [39], and some useful sources are listed in Table 2.

Table 3 summarizes the performance evaluated on some of these databases and Fig. 14 gives a visual illustration of the localization results on them. In the table we highlight the major challenges contained in each database, the size of test images, the localization performance in terms of the percentage of images that have been successfully handled corresponding to a certain localization error (cf. Eq. (2)). Note also that the time costs given in this table may be recorded under different settings. Among them, FERET [82] is one of the most popular databases for both face recognition and eye localization. Face images in this database are taken under controlled conditions, thus they are suitable for general evaluation of eye localization algorithms. The table shows that most methods perform well on this database. The face images of JAFFE [20] have rich expression variations and most with eyes wide open, which makes them particularly suitable for evaluating those methods based on measuring the geometrical or intensity characteristics of eyes (e.g., [114]).

Table 2
A list of data sources and tools related to eye localization.

Sources	Links
Data and annotations	
Low quality eye [109]	http://www.cbsr.ia.ac.cn/users/dyi/eyelocalization.htm
RPI ISL Eye Database	http://www.ecse.rpi.edu/~cvrl/database/database.html
AR, BANCA, FRGC1.0, FRGC2.0	http://lipori.di.unimi.it/download/gt.html
AR, BioID, XM2VTS, Talking Face, PETS ICVS 2003	http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html
FERET	http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
Cohn Kanade Database	http://lipori.di.unimi.it/download/gt2.html
Annotated Facial Landmarks in the Wild [52]	http://lrs.icg.tugraz.at/research/aflw/
IMM Face [79]	http://www2.imm.dtu.dk/aam/datasets/datasets.html
Code and toolBox	
Hough transform [5]	http://graphics.cs.msu.ru/en/science/research/machinelearning/hough
Structured output regression [9]	https://sites.google.com/site/christophlampert/software
Pictorial structures [25,26]	http://people.cs.uchicago.edu/~rbg/latent/
Enhanced Pictorial Model [92]	http://parnec.nuaa.edu.cn/xtan/data/eyedetector.html
Max-Margin Additive [70]	http://www.cs.berkeley.edu/~smaji/projects/add-models/
Fast Intersection Kernel [71]	http://www.cs.berkeley.edu/~smaji/projects/fiksvm/
Lighting processing [91]	http://lear.inrialpes.fr/people/triggs/src/amfg07-demo-v1.tar.gz
Landmark Localization [115]	http://www.ics.uci.edu/~xzhu/face/
Eye+Mouth [15]	http://homes.di.unimi.it/~lipori/download.html
Annotation tool [96]	http://cmp.felk.cvut.cz/~uricamic/flandmark/
OCOFTools [10]	http://www.cs.colostate.edu/~ross/ocof_toolset_2012/

The BioID is a much more challenging database for eye localization, featured with a large variety of changes in illumination, background and face size. Actually, performance reported on this database drops about 7% compared to that on the relatively simple databases [67,93]. This indicates that complicated uncontrolled conditions are still big challenges for current eye localization methods. In this aspect, methods exploiting rich feature sets and structural information may have advantages (cf. Fig. 14). Finally, the FRGC [83] face data are featured with large lighting changes but with high image resolution, and most methods show satisfying performance on this database.

5. Conclusion and prospect

Eyes are arguably one of the most salient features of the human face, and locating eyes precisely and efficiently is of great importance for a wide range of real-world applications. It is not uncommon to mistakenly think that eye localization is a simple task since eyes are just a simple structure in the face. However, eyes have its unique geometric, photometric and motion characteristics, and changes due to these three characteristics would lead to a very complicated nonlinear manifold of eyes. Developing effective eye models to describe this manifold and performing efficient searching within this manifold is therefore of both academic and practical interest. In the recent thirty years, intensive studies on this problem have resulted in a great amount of achievements.

Table 3
Lists of eye localization performance evaluated on various face databases.

Databases	**Challenges**	**Methods**	**#Test**	**Accuracy (%)**		**Time (s)**
				$d_{eye} < 0.1$	$d_{eye} < 0.25$	
FERET [82]	Lighting, expression, pose	General-to-specific [13]	375	89.5	96.4	12
		Eyes + mouth [15]	1175	97.3	99.7	3
		Multi-scale LBP [55]	3368	97.6	99.6	
		Avgerage of synthetic exact filters [10]	1699	98.5		0.03
		Bayesian method [22]	1000	99.0		
		Multi-scale Gabor [51]	488 (fa)	91.8 (0.07)		
		Combining face detector [75]	1175	99.7 (0.05)		
JAFFE [20]	Expression	General projection function [114]	213		97.2	
		Locally smoothed IPF [65]	213		99.1	
		AdaBoost [67]	213	98.6	100 (0.12)	0.06
		AdaBoost + SVM [93]	213	99.5	100	0.04
		2D Cascaded AdaBoost [77]	213	100		
		Probabilistic cascade [109]	213	100		
		Multi-scale Gabor [51]	213	100 (0.07)		
		Multi-view eyes localization [29]	213	60.9	98.6	
BioID [23]	Lighting,	General projection function [114]	1521		94.8	
		Locally smoothed IPF [65]	489		95.2	
	Background,	Isophote curvature [97]	1521	90.9	98.5	0.01
	scale	Eyes + mouth [15]	1521	93.2	99.3	3
		Multi-scale sparse dictionary [107]	1521	95.5	99.1	
		Multi-scale LBP [55]	1512	97.9	99.9	0.95
		Probabilistic Cascade [109]	1521	99.0	100	
		Multi-scale Gabor [51]	1521	96.4	98.8	
		Combining face detector [75]	1521	99.3 (0.06)		
		AdaBoost + SVM [93]	1521	91.8	98.1	0.06
		2D Cascaded AdaBoost [77]	1521	93.0	97.3	0.03
LFW [42]	Uncontrolled	Enhanced pictorial model [92]	1000	98.4		0.1
		Probabilistic cascade [109]		88.1	99.8	
		Intensity filtering and clustering [84]	1192	90.6	96.1	
		Novel correlation filter [38]	1540		86.0[a]	
		Multi-view eyes localization [29]	400	75.4	97.2	
		Locally smoothed IPF [65]	1000		95.6	
FRGC [83]	Controlled	Eyes + mouth [15]	748 (v1)	97.7	100	3
			1430(v2)	95.0	99.9	3
	Uncontrolled	Eyes + mouth [15]	409 (v1)	91.2	95.1	3
			1430(v2)	89.1	94.5	3
	Controlled/Uncontrolled	General-to-specific [13]	862	92.8	97.1	12
		Multi-scale LBP [55]	39094	98.6	99.6	
		Discriminant features [101]	4715 (v1)	99.0		

[a] The error measured in pixels is converted into normalized localization error (cf. Eq. (2)) by assuming that the eyes distance is 40 pixels for faces in LFW [42].

In this paper, we attempt to provide a comprehensive survey of current researches in this area. In particular, we have focused on the overall difficulties and challenges in real-life scenarios and described the current state of the art in dealing with these challenges. The ad hoc methods are mostly reviewed from the perspective of pattern recognition and computer vision with the bias on how to learn generalizable, flexible and efficient statistical eye models from a small number of training images, which, we believe, could help us to develop a clear understanding of this particular problem. We also pay much attention to the practical issues concerning the development of a robust eye localization system, ranging from various preprocessing methods to postprocessing ones. Furthermore, we reviewed the relative localization performance of current algorithms on popular face databases and discussed various factors (e.g., the design of ground truth) that should be taken into account when evaluating a system. Finally, it is worth mentioning that some closely related problems are deliberately ignored in this paper, such as gaze estimation and eye tracking, which are also very important in practice. For these topics, we refer to [37] for a detailed discussion.

Despite of many efforts devoted to eye localization during the last several decades, we have to admit that this problem is far from being resolved and several promising research directions could be suggested.

Firstly, from the application's point of view, locating eyes under uneven lighting and occlusion, variable face pose, and low image resolution remains largely problematic. Each of these difficulties is non-trivial and is of great academic interest as well. While the final solution of these problems depends heavily on the advance of related areas, such as computer vision, pattern recognition and machine learning, promising progress can be made in developing new problem-specific but flexible feature extraction techniques and constructing effective mathematical model for lighting, pose, occlusion and image enhancement. In addition, considering that human can perform the task of object detection and recognition task effectively and efficiently, insights into the mechanism of human brain and neural network will definitely promote development in eye localization, for example, through deriving novel visual attention model.

Secondly, accuracy and efficiency are conflicting in the modeling of eyes. It is still worth discussing on how to find feature sets and to develop optimization techniques compromising the requirements for both the accuracy and efficiency. 'Coarse-to-fine' and 'branch-and-bound' are currently two major fast searching techniques for real-time eye localization. However, constructing a

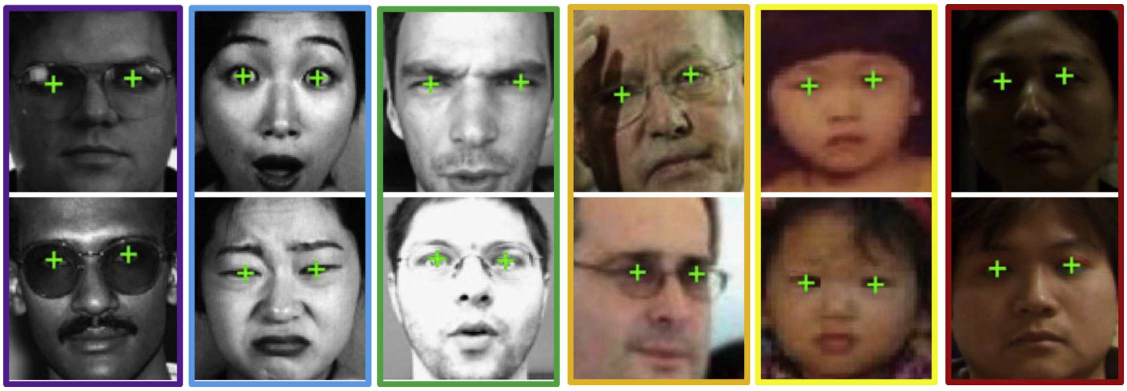

Fig. 14. Example results of eye localization of the enhanced pictorial structure model [92] on several typical databases, from left to right: FERET [82], JAFFE [20], BIOID [23], LFW [42], FaceTracer [56], and FRGC [83].

cascade of classifiers with increasing complexity may miss the global solution, while the branch and bound searching avoids this problem but its usage depends on the heuristic exploring of specific model structure. We are therefore expecting to see more effective optimization models in the next few years which accurately model the eyes while being fitted efficiently.

Thirdly, there are needs to develop benchmark eye databases with carefully designed ground truth. Although current face databases contain useful eye patterns, they are not originally designed for eye localization and the ground truth are mainly provided for face recognition, which cannot be used for other eye-related tasks such as gaze estimation. Furthermore, many face images collected in the laboratory have eyes looking in the frontal direction, which is rarely the case in reality. While the LFW database [42] is closed to the real-life scenarios, it lacks ground truth for eyes partly due to the difficulties involved in labeling those realistic face images. Fortunately, we see that at least two large-scale annotated face databases, i.e., FaceTracer [56] and AFLW [52] with images harvested from the web, have emerged to meet the need for ground truth.

Finally, the question of whether the application areas for eye localization actually need precise eye locations bears mentioning. Specifically, it is usually assumed that better face alignment leads to better performance. However, it has been shown that accurate recognition is possible (e.g., using high-level attribute features [57]) without requiring face alignment and, hence, do not rely on eye localization. Will the existence of such robust methods remove the need for eye localization algorithms? Probably not, but this is a topic that needs to be addressed. After all, it is well-known that the warping operation has the side effect of image distortion.

Nevertheless, the aforementioned difficulties do not mean that addressing this problem cannot be achieved under the current technique framework. As shown in Table 3, considerable efforts in this field are very encouraging. Due to the inherent complexity of this problem and its wide practical applications, we believe that this area will draw increasing attention from a variety of fields beyond computer vision, pattern recognition and machine learning.

Conflict of Interest

None declared.

Acknowledgments

The authors want to thank the editors and reviewers for helpful comments and suggestions. The authors also want to thank Dr. Xin Geng for proof reading. This work was supported by the National Science Foundation of China (61073112, 61035003, 61273301), Jiangsu Science Foundation (BK2012793), Qing Lan Project, Specialized Research Fund for the Doctoral Program of Higher Education (20123218110033) and the Fundamental Research Funds for the Central Universities (CXLX11_0204).

References

[1] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 721–732.

[2] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.

[3] S. Asteriadis, N. Nikolaidis, I. Pitas, Facial feature detection using distance vector fields, Pattern Recognition 42 (7) (2009) 1388–1398.

[4] G.H. Bakir, T. Hofmann, B. Schölkopf, A.J. Smola, B. Taskar, S.V. N. Vishwanathan, Predicting Structured Data (Neural Information Processing), The MIT Press, 2007, ISBN 0262026171.

[5] O. Barinova, V. Lempitsky, P. Kohli, On the detection of multiple object instances using hough transforms, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2233–2240.

[6] P.N. Belhumeur, D.J. Kriegman, What is the set of images of an object under all possible illumination conditions, International Journal of Computer Vision 28 (3) (1998) 245–260.

[7] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[8] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006, ISBN 0387310738.

[9] M. Blaschko, C. Lampert, Learning to localize objects with structured output regression, in: Proceedings of European Conference on Computer Vision, 2008, pp. 2–15.

[10] D.S. Bolme, B.A. Draper, J.R. Beveridg, Average of synthetic exact filters, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2105–2112.

[11] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (10) (1993) 1042–1052.

[12] P. Campadelli, R. Lanzarotti, Fiducial point localization in color images of face foregrounds, Image and Vision Computing 22 (11) (2004) 863–872.

[13] P. Campadelli, R. Lanzarotti, G. Lipori, Precise eye localization through a general-to-specific model definition, in: Proceedings of the British Machine Vision Conference, vol. 1, 2006, pp. 187–196.

[14] P. Campadelli, R. Lanzarotti, G. Lipori, Eye localization: a survey, in: The Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue NATO Science Series, vol. 18, 2007, pp. 234-245.

[15] P. Campadelli, R. Lanzarotti, G. Lipori, Precise eye and mouth localization, International Journal of Pattern Recognition and Artificial Intelligence 23 (3) (2009) 359–377.

[16] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2707–2714.

[17] G. Chow, X. Li, Towards a system for automatic facial feature detection, Pattern Recognition 26 (1993) 1739–1755.

[18] C. Colombo, A. Del Bimbo, Real-time head tracking from the deformation of eye contours using a piecewise affine camera, Pattern Recognition Letters 20 (7) (1999) 721–730.

[19] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.

[20] The JAFFE database. 〈http://www.kasrl.org/jaffe.html〉.

[21] J. Deng, F. Lai, Region-based template deformation and masking for eye-feature extraction and description, Pattern Recognition 30 (1997) 403–419.

[22] M.R. Everingham, A. Zisserman, Regression and classification approaches to eye localization in face images, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2006, pp. 441–448.

[23] The BioID face database. 〈http://www.bioid.com/index.php?q=downloads/software/bioid-face-database.html〉.

[24] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognition 36 (1) (2003) 259–275.

[25] P.F. Felzenszwalb, D.P. Huttenlocher, Pictorial structures for object recognition, International Journal of Computer Vision 61 (1) (2005) 55–79.

[26] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1627–1645.

[27] G.C. Feng, P.C. Yuen, Variance projection function and its application to eye detection for human face recognition, Pattern Recognition Letters 19 (9) (1998) 899–906.

[28] M.A. Fischler, R.A. Elschlager, The representation and matching of pictorial structures, IEEE Transactions on Computers 22 (1) (1973) 67–92.

[29] Y. Fu, H. Yan, J. Li, R. Xiang, Robust facial features localization on rotation arbitrary multi-view face in complex background, Journal of Computers 6 (2) (2011) 337–342.

[30] Z. Fu, A. Robles-Kelly, On mixtures of linear SVMs for nonlinear classification, in: Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition, 2008, pp. 489–499.

[31] C. Garcia, M. Delakis, Convolutional face finder: a neural architecture for fast and robust face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (11) (2004) 1408–1423.

[32] A. Geetha, V. Ramalingam, S. Palanivel, B. Palaniappan, Facial expression recognition—a real time approach, Expert Systems with Applications 36 (1) (2009) 303–308.

[33] R. Gross, V. Brajovic, An image preprocessing algorithm for illumination invariant face recognition, in: International Conference on Audio- and Video-Based Biometric Person Authentication, 2003, pp. 1055–1055.

[34] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[35] F. Guodail, E. Lange, T. Iwamoto, Face recognition system using local autocorrelations and multiscale integration, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (10) (1996) 1024–1028.

[36] M. Hamouz, J. Kittler, J.K. Kamarainen, P. Paalanen, H. Kalviainen, J. Matas, Feature-based affine invariant localization of faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (9) (2005) 1490–1495.

[37] D.W. Hansen, Q. Ji, In the eye of the beholder: a survey of models for eyes and gaze, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (3) (2010) 478–500.

[38] B. Heflin, W. Scheirer, T.E. Boult, For your eyes only, in: Proceedings of IEEE Workshop on the Applications of Computer Vision, 2012, pp. 193–200.

[39] Face Recognition Homepage. 〈http://www.face-rec.org/databases/〉.

[40] R.L. Hsu, M. Abdel Mottaleb, A.K. Jain, Face detection in color images, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 696–706.

[41] G.B. Huang, V. Jain, E. Learned-Miller, Unsupervised joint alignment of complex images, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 1–8.

[42] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[43] J. Huang, H. Wechsler, Visual routines for eye location using learning and evolution, IEEE Transactions on Evolutionary Computation 4 (1) (2000) 73–82.

[44] V. Jain, E. Learned-Miller, Online domain adaptation of a pre-trained cascade of classifiers, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 577–584.

[45] R. Javier, V. Rodrigo, C. Mauricio, Recognition of faces in unconstrained environments: a comparative study, European Association for Signal Processing Journal on Advances in Signal Processing 1–19 (2009) 2009.

[46] O. Jesorsky, K. Kirchberg, R. Frischholz, Robust face detection using the Hausdorff distance, in: Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication, 2001, pp. 90–95.

[47] X. Jin, X. Tan, L. Zhou, Face alignment using local hough voting, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2013.

[48] J.L. Crowley, N. Gourier, D. Hall, Facial features detection robust to pose, illumination and identity, in: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 617–622.

[49] D.J. Jobson, Z.U. Rahman, G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Transactions on Image Processing 6 (7) (1997) 965–976.

[50] S. Kawato, J. Ohya, Real-time detection of nodding and head-shaking by directly detecting and tracking the between-eyes, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2000, pp. 40–45.

[51] S. Kim, S.T. Chung, S. Jung, D. Oh, J. Kim, S. Cho, Multi-scale Gabor feature based eye localization, in: Proceedings of World Academy of Science, vol. 21, 2007, pp. 483–487.

[52] M. Kostinger, P. Wohlhart, P.M. Roth, H. Bischof, Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization, in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 2144–2151.

[53] R. Kothari, J.L. Mitchell, Detection of eye locations in unconstrained visual images, in: Proceedings of International Conference on Image Processing, vol. 3, 1996, pp. 519–522.

[54] T. Kozakaya, T. Shibata, M. Yuasa, O. Yamaguchi, Facial feature localization using weighted vector concentration approach, Image and Vision Computing 28 (5) (2010) 772–780.

[55] B. Kroon, S. Maas, S. Boughorbel, A. Hanjalic, Eye localization in low and standard definition content with application to face matching, Computer Vision and Image Understanding 113 (8) (2009) 921–933.

[56] N. Kumar, P.N. Belhumeur, S.K. Nayar, FaceTracer: a search engine for large collections of images with faces, in: Proceedings of European Conference on Computer Vision, October 2008, 2008, pp. 340–353.

[57] N. Kumar, A.C. Berg, P. N. Belhumeur, S. K. Nayar, Attribute and simile classifiers for face verification, in: Proceedings of IEEE International Conference on Computer Vision, 2009, pp. 365–372.

[58] K. Lam, H. Yan, Locating and extracting the eye in human face images, Pattern Recognition 29 (1996) 771–779.

[59] C.H. Lampert, M.B. Blaschko, T. Hofmann, Beyond sliding windows: object localization by efficient subwindow search, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[60] K.C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (5) (2005) 684–698.

[61] A. Lehmann, B. Leibe, L. Van Gool, Fast PRISM: branch and bound hough transform for object class detection, International Journal of Computer Vision 94 (2) (2011) 175–197.

[62] B. Leibe, A. Leonardis, B. Schiele, An implicit shape model for combined object categorization and segmentation, in: Towards Category-Level Object Recognition, vol. 4170, 2006, pp. 508–524.

[63] B. Leibe, A. Leonardis, B. Schiele, Robust object detection with interleaved categorization and segmentation, International Journal of Computer Vision 77 (1) (2008) 259–289.

[64] S.Z. Li, R.F. Chu, S.C. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 627–639.

[65] W. Li, Y. Wang, Y. Wang, Eye location via a novel integral projection function and radial symmetry transform, International Journal of Digital Content Technology and its Applications 5 (8) (2011) 70–80.

[66] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[67] Y. Ma, X. Ding, Z. Wang, N. Wang, Robust precise eye location under probabilistic framework, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2004, pp. 339–344.

[68] A. Mahalanobis, B.V.K. Kumar, D. Casasent, Minimum average correlation energy filters, Applied Optics 26 (17) (1987) 3633–3640.

[69] A. Mahalanobis, B.V.K. Vijaya Kumar, S. Song, S.R.F. Sims, J.F. Epperson, Unconstrained correlation filters, Applied Optics 33 (17) (1994) 3751–3759.

[70] S. Maji, A.C. Berg, Max-margin additive classifiers for detection, in: Proceedings of IEEE International Conference on Computer Vision, 2009, pp. 40–47.

[71] S. Maji, A.C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[72] J. Marques, N.M. Orlans, A.T. Piszcz, Effects of eye position on eigenface-based face recognition scoring, Technical Paper of the MITRE Corporation, October 2003.

[73] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10) (2005) 1615–1630.

[74] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 696–710.

[75] L. Nanni, A. Lumini, Combining face and eye detectors in a high-performance face-detection system, IEEE MultiMedia 19 (2012) 20–27.

[76] M.H. Nguyen, J. Perez, F. De La TORRE, Facial feature detection with optimal pixel reduction SVMs, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2008, pp. 1–6.

[77] Z. Niu, S. Shan, S. Yan, X. Chen, W. Gao, 2D cascaded AdaBoost for eye localization, in: International Conference on Pattern Recognition, 2006, pp. 1216–1219.

[78] M. Nixon, Eye spacing measurement for facial recognition, in: Proceedings of Society of Photo-Optical Instrument Engineers, vol. 575, 1985, pp. 279–285.

[79] M.M. Nordstrøm, M. Larsen, J. Sierakowski, M.B. Stegmann, The IMM face database—an annotated dataset of 240 face images, Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, May 2004.

[80] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: Proceedings of International Conference on Image and Signal Processing, Lecture Notes in Computer Science, 2008, pp. 236–243.

[81] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 84–91.

[82] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

[83] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 947–954.

[84] Z. Qian, D. Xu, Automatic eye detection using intensity filtering and K-means clustering, Pattern Recognition Letters 31 (12) (2010) 1633–1640.

[85] T. Riopka, T. Boult, The eyes have it, in: Proceedings of ACM SIGMM Multimedia Biometrics Methods and Applications Workshop, 2003, pp. 9–16.

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

[87] J.M. Saragih, S. Lucey, J.F. Cohn, Deformable model fitting with a mixture of local experts, in: Proceedings of IEEE International Conference on Computer Vision, 2009, pp. 2248–2255.

[88] H.J. Seo, P. Milanfar, Training-free, generic object detection using locally adaptive regression kernels, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2010) 1688–1704.

[89] S. Shan, Y. Chang, W. Gao, B. Cao, P. Yang, Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2004, pp. 314–320.

[90] M. Pietikainen, T. Ojala, D. Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition 29 (1996) 51–59.

[91] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Transactions on Image Processing 19 (2010) 1635–1650.

[92] X. Tan, F. Song, Z.H. Zhou, S. Chen, Enhanced pictorial structures for precise eye localization under uncontrolled conditions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 20–25 June 2009, 2009, pp. 1621–1628.

[93] X. Tang, Z. Ou, T. Su, H. Sun, P. Zhao, Robust precise eye location by AdaBoost and SVM techniques, in: Proceedings of International Conference on Advances in Neural Networks, 2005, pp. 93–98.

[94] Y. Tian, T. Kanade, J.F. Cohn, Dual-state parametric eye tracking, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2000, pp. 110–115.

[95] S. Ullman, Visual routines, Cognition 18 (1984) 97–159.

[96] M. Uricár, V. Franc, V. Hlavác, Detector of facial landmarks learned by the structured output SVM, in: Proceedings of International Conference on Computer Vision Theory and Applications, vol. 1, 2012, pp. 547–556.

[97] R. Valenti, T. Gevers, Accurate eye center location and tracking using isophote curvature, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[98] P. Viola, M.J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.

[99] J. Wang, L. Yin, J. Moore, Using geometric properties of topographic manifold to detect and track eyes for human-computer interaction, ACM Transactions on Multimedia Computing, Communications, and Applications 3 (4) (2007).

[100] J.G. Wang, E. Sung, Frontal-view face detection and facial feature extraction using color and morphological operations, Pattern Recognition Letters 20 (10) (1999) 1053–1068.

[101] P. Wang, M. Green, Q. Ji, J. Wayman, Automatic eye detection and its validation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 3, 2005, pp. 164–172.

[102] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 775–779.

[103] X. Xie, R. Sudhakar, H. Zhuang, On improving eye feature extraction using deformable templates, Pattern Recognition 27 (1994) 791–799.

[104] F. Xiong, Y. Zhang, G. Zhang, A pupil localization algorithm based on adaptive Gabor filtering and negative radial symmetry, in: Proceedings of International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, vol. 4679, 2007, pp. 87–96.

[105] R. Xu, B. Zhang, Q. Ye, J. Jiao, Human detection in images via L1-norm minimization learning, in: IEEE International Conference on Acoustics Speech and Signal Processing, 2010, pp. 3566–3569.

[106] S. Yan, S. Shan, X. Chen, W. Gao, Locally Assembled Binary (LAB) feature with feature-centric cascade for fast and accurate face detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–7.

[107] F. Yang, J. Huang, P. Yang, D. Metaxas, Eye localization through multiscale sparse dictionaries, in: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, 2011, pp. 514–518.

[108] P. Yang, B. Du, S. Shan, W. Gao, A novel pupil localization method based on GaborEye model and radial symmetry operator, in: Proceedings of International Conference on Image Processing, vol. 1, 2004, pp. 67–70.

[109] D. Yi, Z. Lei, S.Z. Li, A robust eye localization method for low quality face images, in: Proceedings of International Joint Conference on Biometrics, 2011, pp. 1–6.

[110] A.L. Yuille, D.S. Cohen, P.W. Hallinan, Feature extraction from faces using deformable templates, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1989, pp. 104–109.

[111] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable template, International Journal of Computer Vision 8 (2) (1992) 99–111.

[112] W. Zhang, H. Chen, P. Yao, B. Li, Z. Zhuang, Precise eye localization with AdaBoost and fast radial symmetry, in: International Conference on Computational Intelligence and Security, vol. 1, 2006, pp. 725–730.

[113] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, Face recognition: a literature survey, ACM Computing Surveys (2003) 399–458.

[114] Z.H. Zhou, X. Geng, Projection functions for eye detection, Pattern Recognition 37 (5) (2004) 1049–1056.

[115] X. Zhu and D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2879–2886.

[116] Z. Zhu, Q. Ji, Robust real-time eye detection and tracking under variable lighting conditions and various face orientations, Computer Vision and Image Understanding 98 (1) (2005) 124–154.

[117] Z. Zhu, K. Fujimura, Q. Ji, Real-time eye detection and tracking under various light conditions, in: Proceedings of the 2002 Symposium on Eye Tracking Research and Applications, vol. 25, 2002, pp. 139–144.

Fengyi Song received the B.Sc. degree in computer science from Henan University, China, in 2006. In 2009, he received his M.Sc. degree in computer applications from Nanjing University of Aeronautics and Astronautics (NUAA), China. Now he is currently a Ph.D. student at the Department of Computer Science and Engineering, NUAA. His research interests include face recognition and computer vision.

Xiaoyang Tan received his B.Sc. and M.Sc. degree in computer applications from Nanjing University of Aeronautics and Astronautics (NUAA) in 1993 and 1996, respectively. Then he worked at NUAA in June 1996 as an assistant lecturer. He received a Ph.D. degree from Department of Computer Science and Technology of Nanjing University, China, in 2005. From September 2006 to October 2007, he worked as a postdoctoral researcher in the LEAR (Learning and Recognition in Vision) team at INRIA Rhone-Alpes in Grenoble, France. His research interests are in face recognition, machine learning, pattern recognition, and computer vision. In these fields, he has authored or coauthored over 20 scientific papers.

Songcan Chen received his B.Sc. degree in mathematics from Hangzhou University (now merged into Zhejiang University) in 1983. In December 1985, he completed his M.Sc. degree in computer applications at Shanghai Jiaotong University and then worked at the Nanjing University of Aeronautics and Astronautics (NUAA) in January 1986 as an assistant lecturer. There he received a Ph.D. degree, in 1997, in communication and information systems. Since 1998, as a full-time professor, he has been with the computer science and engineering department at NUAA. His research interests include pattern recognition, machine learning and neural computing. In these fields, he has authored or coauthored over 130 scientific journal papers.

Zhi-Hua Zhou (S'00-M'01-SM'06-F'13) received his B.Sc., M.Sc. and Ph.D. degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honor. He joined the Department of Computer Science & Technology of Nanjing University as an Assistant Professor in 2001, and at present he is a Professor and Deputy Director of the National Key Lab for Novel Software Technology, and Founding Director of LAMDA (the Institute of Machine Learning and Data Mining) at Nanjing University.

He has wide research interests, mainly including machine learning, data mining, pattern recognition and artificial intelligence. In these areas he has published more than 100 papers in leading international journals or conferences, and holds 12 patents. He has been awarded with various honors such as the IEEE Computational Intelligence Society

Outstanding Early Career Award (2013), the Fok Ying Tung Young Professorship First-Grade Award (2010), the National Science & Technology Award for Young Scholars of China (2006), the Microsoft Young Professorship Award (2006), and a number of journal/conference paper awards. He is a Fellow of the IEEE, Fellow of the IAPR, and Fellow of the IET/IEE.

He serves/ed as an Associate Editor-in-Chief of the Chinese Science Bulletin, and Associate Editor or editorial boards member of various journals including the IEEE Transactions on Knowledge and Data Engineering and ACM Transactions on Intelligent Systems and Technology. He also served as Guest Editor for Machine Learning, Pattern Recognition, IEEE Intelligent Systems, etc. He founded ACML and serves as its Steering Committee Chair. He also serves/ed as Steering Committee member of PAKDD and PRICAI, General Chair/Co-Chair of ACML'12 and ADMA'12, Program Committee Chair/Co-Chair of PAKDD'07, PRICAI'08, ACML'09, SDM'13, etc., and Vice Chair or Area Chair of various conferences. He also chaired many domestic conferences in China.

He is the Chair of the Artificial Intelligence and Pattern Recognition Technical Committee of the CCF (China Computer Federation), Chair of the Machine Learning Technical Committee of the CAAI (Chinese Association of Artificial Intelligence), Vice Chair of the IEEE Computational Intelligence Society Data Mining Technical Committee, Vice Chair of the IEEE Nanjing Section, and the Chair of the IEEE Computer Society Nanjing Chapter.