

# Face Verification Across Age Progression Using Discriminative Methods

Haibin Ling, *Member, IEEE*, Stefano Soatto, *Senior Member, IEEE*, Narayanan Ramanathan, *Student Member, IEEE*, and David W. Jacobs, *Member, IEEE*

**Abstract**—Face verification in the presence of age progression is an important problem that has not been widely addressed. In this paper, we study the problem by designing and evaluating discriminative approaches. These directly tackle verification tasks without explicit age modeling, which is a hard problem by itself. First, we find that the gradient orientation, after discarding magnitude information, provides a simple but effective representation for this problem. This representation is further improved when hierarchical information is used, which results in the use of the gradient orientation pyramid (GOP). When combined with a support vector machine GOP demonstrates excellent performance in all our experiments, in comparison with seven different approaches including two commercial systems. Our experiments are conducted on the FGnet dataset and two large passport datasets, one of them being the largest ever reported for recognition tasks. Second, taking advantage of these datasets, we empirically study how age gaps and related issues (including image quality, spectacles, and facial hair) affect recognition algorithms. We found surprisingly that the added difficulty of verification produced by age gaps becomes saturated after the gap is larger than four years, for gaps of up to ten years. In addition, we find that image quality and eyewear present more of a challenge than facial hair.

**Index Terms**—Age progression, face verification, gradient orientation pyramid (GOP), support vector machine (SVM).

## I. INTRODUCTION

### A. Background

**F**ACE verification is an important problem in computer vision and has a very wide range of applications, such as surveillance, human–computer interaction, image retrieval, etc. A thorough survey can be found in [42]. A large amount

of research effort has been focused on pursuing robustness to different imaging conditions, including illumination change, pose variation, expression, etc. Despite decades of study on face image analysis, age-related facial image analysis has not been extensively studied until recently. Most of these works focus on age estimation [8]–[11], [14], [15], [24], [30], [40], [41], [43] and age simulation [18], [35], [36], [38]. In addition, some researchers study the effect of age progression on face profiles and appearances [16], [31], [32], [37].

Face verification across age has been subject to relatively little attention. Some previous work applies age progression for face verification tasks. When comparing two photos, these methods either transform one photo to have the same age as the other, or transform both to reduce the aging effects. One of the earliest works appears in Lanitis *et al.* [18], where a statistical model is used to capture the variation of facial shapes over age progression. The model is then used for age estimation and face verification. Ramanathan and Chellappa [31] use a face growing model for face verification tasks for people under the age of 18. This assumption limits the application of these methods, since ages are often not available. A recent work by Biswas *et al.* [4] studies feature drifting on face images at different ages and applies it to face verification tasks. Other studies using age transformation for verification include [9], [25], [26], [34], and [40].

The above methods can be roughly categorized as generative methods since aging needs to be modeled. In fact, most of them use verification to evaluate the age modeling algorithm. While these methods explicitly address the aging issue, they usually require additional information about the images being compared, such as actual age. In addition, many landmark points are often used for modeling age progression or building statistical models. All the methods mentioned above use the 68 landmarks that are pre-labeled for each photo in the FGnet dataset [1]. Furthermore, both age estimation and age simulation are still open problems and may bring instabilities to the generative methods. To avoid these problems, we study discriminative methods that directly tackle the face verification problem.

Discriminative approaches have been used for face verification across age progression. The study most related to our work is [30], where the probabilistic eigenspace framework [22] is adapted for face identification across age progression. Instead of using a whole face, only a half face (called a PointFive face) is used to alleviate the nonuniform illumination problem. Then, eigenspace techniques and a Bayesian model are combined to capture the intrapersonal and extrapersonal image differences. An eigenspace is also used in [17] in combination with a statistical model on the FGnet dataset [1] and in [33] on the MORPH dataset. We study the same task as that studied in [30]. As will

Manuscript received April 12, 2009; accepted November 05, 2009. First published December 15, 2009; current version published February 12, 2010. The work of H. Ling and S. Soatto was supported in part by ONR 67F-1080868 and NSF ECS-0622245. The work of D. W. Jacobs was supported by a Honda Research Initiation Grant and from the Office of Naval Research Under the MURI Grant N00014-08-10638. The work of H. Ling, N. Ramanathan, and D. W. Jacobs was supported in part by a fellowship from Apttis Inc. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick J. Flynn.

H. Ling is with the Department of Computer and Information Science, Temple University, Philadelphia, PA 19122 USA (e-mail: hbling@temple.edu).

S. Soatto is with the Computer Science Department, University of California, Los Angeles, CA 90092 USA (e-mail: soatto@cs.ucla.edu).

N. Ramanathan is with Cernium Corporation, Reston, VA 20191 USA (e-mail: nramanathan@cernium.com).

D. W. Jacobs is with the Computer Science Department, University of Maryland, College Park, MD 20742 USA (e-mail: djacobs@cs.umd.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2009.2038751

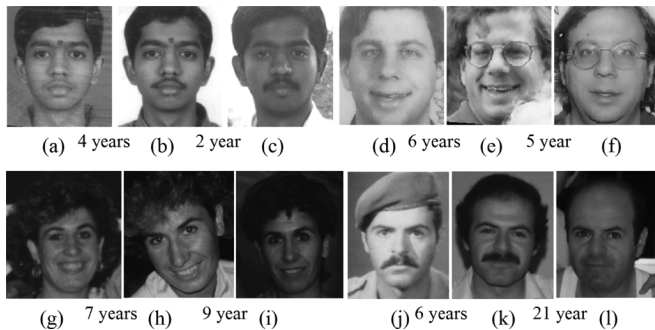


Fig. 1. Typical images with age differences. Top row: scanned passport or visa photos. Bottom row: photos from the FG-NET Aging Database [1].

be clarified in the following sections, our work differs from previous studies in both the representation [we use gradient orientation pyramids (GOPs)] and the classification frameworks [we use support vector machine (SVM)]. Part of this work was published in a preliminary conference version [19].

### B. Tasks and Challenges

The goal of our study is two-fold. The first is to investigate representations and algorithms for verification. The second is to study the effect of age gaps and related issues (including image quality, spectacles, and facial hair) on verification algorithms. We use three datasets in our study. Two of them are passport datasets involving more than 1800 subjects, which to the best of our knowledge are the largest datasets ever studied for the task. We also use the FG-NET Aging Database [1] that is widely used for image-based face aging analysis.

The challenges of face verification across age progression are due to several sources. The first source is the biometric change over years, including facial texture [e.g., wrinkles as on the forehead in Fig. 1(i)], shape [e.g., weight gain, Fig. 1(d)–(f)], facial hair [mustache and beard, e.g., Fig. 1(a)–(c), (k)–(l)], presence of glasses [e.g., Fig. 1(d)–(e)], scars, etc. The second source is the change in the image acquisition conditions and environment, including the illumination conditions, the image quality change caused by using different cameras, etc. In addition, for images converted from nondigital photos, additional artifacts [e.g., saturation in Fig. 1(e)] sometimes appear due to scanning processes and sometimes the original photos are smudged. Some examples of these challenges are shown in Fig. 1.

### C. Contribution

We make several contributions in this study. First, we propose using the GOP for the task. We show that, when combined with the SVM [39], GOP demonstrates excellent performance for face verification with age gaps. This is mainly motivated by the illumination insensitivity of gradient orientation (GO) as shown in [6]. We conjecture in our preliminary work [19] that GO is robust to aging processes under some flexible conditions that are usually true in the context of face verification. The pyramid technique is used to capture hierarchical information that further improves the representation. Then, given a face image pair, we use the cosines between GOPs at all scales to build the feature vector. The feature vector is then combined with an SVM for face verification in a way similar to [27].

Our second contribution is thorough empirical experiments. We evaluated nine different approaches, including two baseline methods ( $l_2$  norm and GO), four different representations with the same SVM-based framework (intensity difference [27], gradient with magnitude, GO, and GOP), the Bayesian face [30], and two commercial face verification systems. The evaluations are conducted on the three datasets mentioned above. To the best of our knowledge, this is the largest reported evaluation in both the size of dataset and the number of tested methods.

Our third contribution is the empirical study of how verification performance varies with increasing age gaps and related issues. We found surprisingly that the added difficulty of verification produced by age gaps becomes saturated after the gap is larger than four years, for gaps of up to ten years. This is observed with different image representations that have been tested. In addition, on the FGnet dataset, we observed that the image quality and presence of eye glasses bring more challenges than facial hair.

The rest of the paper is organized as follow. In Section II, we formulate the task of face verification using an SVM framework. Then, we introduce the GO pyramid in Section II-B. After that, Section III describes our experiments on two passport image datasets and the FG-NET dataset, which have large age separations. Section IV presents our empirical study of how age gaps affect verification algorithms. Section V reports the verification experiments on face images from children. Finally, Section VI concludes the paper.

## II. PROBLEM FORMULATION

### A. Face Verification Framework

In this paper, we study face verification tasks as in [30]. In verification, one must determine whether two images come from the same person, as opposed to recognition, in which an individual is identified from a large gallery of individuals. An advantage of this problem is that it does not require many images for each subject, which is often difficult for collections across aging. Furthermore, this problem directly relates to the passport renewal task that is important for the passport datasets in our experiments. In the task, a newly submitted photo needs to be compared with an old one, to ensure that the request is valid. Face verification as a two-class classification problem has been studied for general face analysis tasks. For example, Moghaddam *et al.* [23] used a Bayesian framework for the intrapersonal and extrapersonal face classification. Phillips [27] used SVM for face recognition problems and observed good results on the FERET database [28] compared to component-based approaches. Jonsson *et al.* [13] used SVM for face authentication problems. All of the above methods use intensity (sometimes normalized intensity) as their representation. In comparison, we use the GOP and apply the framework for problems involving large age differences.

As in [13], [23], and [27], we model face verification as a two-class classification problem. Given an input image pair  $I_1$  and  $I_2$ , the task is to assign the pair as either *intrapersonal* (i.e.,  $I_1$  and  $I_2$  from the same people) or *extrapersonal* (i.e.,  $I_1$  and  $I_2$  from different individuals). We use an SVM [39]. Specifically,

given an image pair  $(I_1, I_2)$ , it is first mapped onto the feature space as

$$\mathbf{x} = \mathcal{F}(I_1, I_2) \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^d$  is the feature vector extracted from the image pair  $(I_1, I_2)$  through the feature extraction function  $\mathcal{F} : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}^d$  ( $\mathcal{F}$  will be described in the following subsections),  $\mathcal{I}$  is the set of all images, and  $\mathbb{R}^d$  forms the  $d$ -dimensional feature space.

Then SVM is used to divide the feature space into two classes, one for intrapersonal pairs and the other for extrapersonal pairs. Using the same terminology as in [27], we denote the separating boundary with the following equation:

$$\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b = \Delta \quad (2)$$

where  $N_s$  is the number of support vectors and  $\mathbf{s}_i$  is the  $i$ th support vector.  $\Delta$  is used to trade off the *correct reject rate* (CRR) and *correct acceptance rate* (CAR) as described in (3) and (4).  $K(\cdot, \cdot)$  is the kernel function that provides SVM with nonlinear abilities. In our experiments, we use the LibSVM library [5].

For verification tasks, the CRR and the CAR are two critical criteria

$$\text{CRR} = \frac{\#\text{correctly rejected extra-personal pairs}}{\#\text{total extra-personal pairs}} \quad (3)$$

$$\text{CAR} = \frac{\#\text{correctly accepted intra-personal pairs}}{\#\text{total intra-personal pairs}} \quad (4)$$

where “accept” indicates that the input image pair are from the same subject and “reject” indicates the opposite. In addition, the *equal error rate* (EER), defined as the error rate when a solution has the same CAR and CRR, is frequently used to measure verification performance.

### B. Gradient Orientation and Gradient Orientation Pyramid

Now we need to decide the representation for feature extraction, i.e.,  $\mathcal{F}(\cdot, \cdot)$ . A natural choice is to use the intensity difference between  $I_1$  and  $I_2$ , which is called *difference space* in [23] and also has been used in [27] and [30]. The difference space can be made robust to affine lighting changes by an appropriate intensity normalization. However, the affine lighting model is not always sufficient for face images, especially for images taken at times separated by years.

Motivated by previous study of the robustness of GO [2], [3], [6], [12], we propose to use GO for face verification across age progression. Specifically, in [6], GO is shown to be robust to illumination change and successfully applied for face recognition tasks. Furthermore, it has been shown in [37] and [38] that the change of face color across age progression can be factored to two components, hemoglobin and melanin, according to skin anatomy. This observation inspired our preliminary study [19], which shows that the GO of each color channel of human faces is robust under age progression. In addition, we collect GO in a hierarchical way, which has been shown to retain most visual information as in [2] and [12].

Note that gradient-based representations are recently widely used in computer vision and pattern recognition tasks, such as

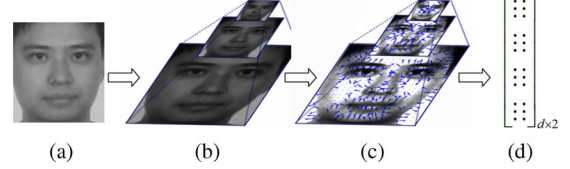


Fig. 2. Computation of a GOP from an input image  $I$ . Note: In (c), the figure is made brighter for better illustration. (a) Image  $I$ . (b) Pyramid  $\mathcal{P}(I)$ . (c) GOP. (d)  $\mathcal{G}(I)$ .

the scale invariant feature transfer (SIFT) [20] for object and category classification and the histogram of orientation (HOG) [7]. In these works, the gradient directions were weighted by gradient magnitudes. In contrast, we discard magnitude information and use only orientations, which demonstrates significant improvement in our experiments (Section III). Furthermore, the gradient directions at different scales are combined to make a hierarchical representation.

Given an image  $I(\mathbf{p})$ , where  $\mathbf{p} = (x, y)$  indicates pixel locations, we first define the pyramid of  $I$  as  $\mathcal{P}(I) = \{I(\mathbf{p}; \sigma)\}_{\sigma=0}^s$  with

$$\begin{aligned} I(\mathbf{p}; 0) &= I(\mathbf{p}) \\ I(\mathbf{p}; \sigma) &= [I(\mathbf{p}; \sigma - 1) * \Phi(\mathbf{p})] \downarrow_2, \quad \sigma = 1, \dots, s \end{aligned} \quad (5)$$

where  $\Phi(\mathbf{p})$  is the Gaussian kernel (0.5 is used as the standard deviation in our experiments),  $\otimes$  denotes the convolution operator,  $\downarrow_2$  denotes half size downsampling, and  $s$  is the number of pyramid layers. Note that in (5) the notation  $I$  is used both for the original image and the images at different scales for convenience.

Then, the GO at each scale  $\sigma$  is defined by its normalized gradient vectors at each pixel

$$g(I(\mathbf{p}; \sigma)) = \begin{cases} \frac{\nabla(I(\mathbf{p}; \sigma))}{|\nabla(I(\mathbf{p}; \sigma))|}, & \text{if } |\nabla(I(\mathbf{p}; \sigma))| > \tau \\ (0, 0)^T, & \text{otherwise} \end{cases} \quad (6)$$

where  $\tau$  is a threshold for dealing with “flat” pixels. The GOP of  $I$  is naturally defined as  $\mathcal{G}(I) = \text{stack}(\{g(I(\mathbf{p}; \sigma))\}_{\sigma=0}^s) \in \mathbb{R}^{d \times 2}$  that maps  $I$  to a  $d \times 2$  representation, where  $\text{stack}(\cdot)$  is used for stacking GOs of all pixels across all scales and  $d$  is the total number of pixels. Fig. 2 illustrates the computation of a GOP from an input image.

### C. Kernels Between GOPs

Given an image pair  $(I_1, I_2)$  and corresponding GOPs ( $G_1 = \mathcal{G}(I_1), G_2 = \mathcal{G}(I_2)$ ), the feature vector  $\mathbf{x} = \mathcal{F}(I_1, I_2)$  is computed as the cosines of the difference between GOPs at all pixels over scales

$$\mathbf{x} = \mathcal{F}(I_1, I_2) = (G_1 \odot G_2) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (7)$$

where  $\odot$  is the element-wise product. Next, we apply the Gaussian kernel to the extracted feature  $\mathbf{x}$  to be used with the SVM framework. Specifically, our kernel is defined as

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\gamma \|\mathbf{x}_1 - \mathbf{x}_2\|^2) \quad (8)$$

where  $\gamma$  is a parameter determining the size of RBF kernels ( $\gamma = 1/d$  is used in our experiments). In the rest of the paper, we use SVM+GOP to indicate the proposed approach.

The proposed SVM+GOP approach demonstrates excellent performance in our experiments (Section III). In the following, we summarize its advantages:

- 1) Being a discriminative method, SVM+GOP tackles the face identification problem directly. This way, it not only avoids the potential instability brought by age estimation and simulation, but also requires less prior information about photos under comparison. Consequently, the proposed approach is more applicable than previously proposed generative methods (see Section I).
- 2) GOP is insensitive to illumination changes [6]. As a result, no normalization is needed on the input images.
- 3) As shown in the preliminary study [19] using anatomic studies of skin color over age, GO is fairly robust across age progression for face verification tasks where high resolution images are avoided.
- 4) The pyramid technique provides a natural way to perform face comparison at different scales.
- 5) As demonstrated in our experiments (Section III), the proposed GO+SVM and GOP+SVM significantly outperform most of its competitors. The performances of two commercial systems are similar to our proposed methods. However, our methods are much simpler than these commercial systems and have potential to be combined with other approaches to further boost the performance.

### III. FACE VERIFICATION EXPERIMENTS

#### A. Experimental Setup

**Datasets:** We conduct face verification experiments on three databases: two passport databases, named Passport I and Passport II, and the FGnet database [1]. All datasets are dominated by Caucasian descendants. Details of these databases are given in the following subsections.

In our experiments, the images are preprocessed using the same scheme as in [30]. This includes manual eye location labeling, alignment by eyes and cropping with an elliptic region. For computational reasons, image sizes are reduced to  $96 \times 84$  for Passport I,  $72 \times 63$  for Passport II, and  $96 \times 84$  for the FGnet database. To alleviate the alignment problem, we tried different alignments with small shifts (up to two pixels), using the shift that led to greatest image similarity. In our experiments, this improved performance by around 0.5% (EER). A similar technique is used by [21].

**Approaches:** We compared the following approaches.

- 1) SVM+GOP: the approach proposed in this paper.
- 2) SVM+GO: this is similar to SVM+GOP, except that only the GO at the finest scale is used without a hierarchical representation.
- 3) SVM+G: this one is similar to SVM+GO, except that the gradient (G) itself is used instead of GO. It can also be viewed as weighting GOs with gradient magnitudes.
- 4) SVM+diff [27]. As in [27], we use the differences of normalized images as input features combined with SVM.
- 5) GO: this is the method using GO proposed in [6].
- 6)  $l_2$ : this is a baseline approach that uses the  $l_2$  norm to compare two normalized

TABLE I  
PASSPORT DATASETS FOR FACE VERIFICATION TASKS. "STD." IS SHORT FOR STANDARD DEVIATION

Dataset	# intra pair	mean age	std. age	mean age diff.	std. age diff.
Pass. I	452	39	10	4.27	2.9
Pass. II	1824	48	14.7	7.45	3.2

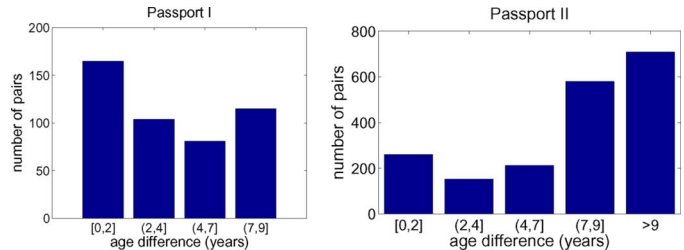


Fig. 3. Distribution of age differences in the passport image databases. Left: Passport I. Right: Passport II.

images. 7) Bayesian+PFF [30]. This is the approach combining Bayesian framework [22] and PointFive Face (PFF) [30]. In addition, two commercial systems are tested on the datasets, which we will name Vendor A and Vendor B.<sup>1</sup>

The first four approaches use exactly the same configurations and the same SVM framework, but different representations. The purpose is to study the value of the proposed GOP representation. The other five approaches are different from our method in both representations and classification frameworks. For intensity-based representations (i.e.,  $l_2$ , SVM+G, SVM+diff), the image intensities are first normalized (by subtracting mean intensities and dividing by the standard deviation of intensities) to achieve affine invariance.

**Experimental Evaluation:** The performance of algorithms is evaluated using the CRR-CAR curves that are usually created by varying some classifier parameters. We used three-fold cross validation in our experiments. For each experiment, the CRR-CAR curve is created by adjusting parameter  $\Delta$  in (2). The total performance is evaluated as the average of the output CRR-CAR curves of three folds. For Vendors A and B, all original color images are input to their systems. To compare with Bayesian+PFF, we also test SVM+GOP in the experimental setup according to [30], i.e., we use 200 positive and 200 negative pairs as a training set. We also use EERs for evaluation.

#### B. Experiments With Passport Datasets

We tested the proposed approach on two real passport image datasets, which we will refer to as Passport I and Passport II, respectively. Passport I is the dataset used in [30]. It contains 452 intrapersonal image pairs (several duplicate pairs were removed) and 2251 randomly generated extrapersonal image pairs. Passport II contains 1824 intrapersonal image pairs and 9492 randomly generated extrapersonal image pairs. The extrapersonal pairs are generated in such a way that there is no overlapping of subjects between training and testing sets (during cross validation), as in [30]. Images in both datasets are scanned passport images. They are in general frontal images with small pose variations. The lighting condition varies, and

<sup>1</sup>Anonymous due to agreements with the companies.



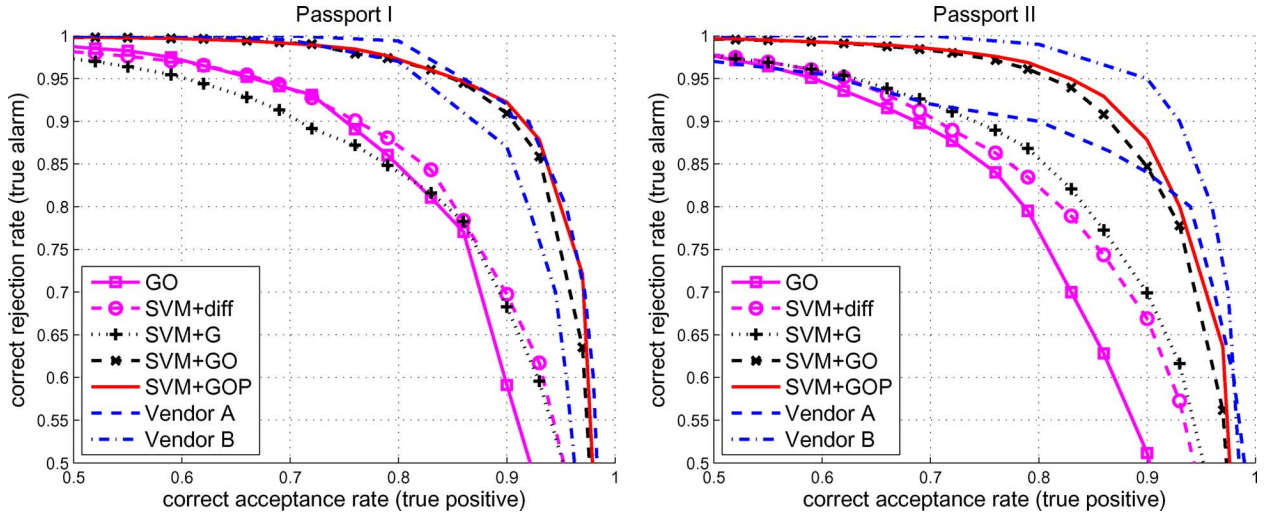


Fig. 4. CRR-CAR curves for three-fold cross validation experiments. Left: on Passport I. Right: on Passport II. This figure is better viewed in color.

TABLE II  
EQUAL ERROR RATES. LEFT: EXPERIMENTS OF THREE-FOLD CROSS VALIDATION. RIGHT: EXPERIMENTS USING 200 INTRA AND 200 EXTRA PAIRS AS TRAINING, AS IN [30]

	GO [6]	SVM+diff [27]	SVM+G	SVM+GO	SVM+GOP	Vendor A	Vendor B	SVM+GOP	Bayesian [30]
Pass. I	17.6%	16.5%	17.8%	9.5%	8.9%	9.5%	11.5%	5.1%	8.5%
Pass. II	20.7%	18.8%	17.4%	12.0%	11.2%	13.5%	8.0%	10.8%	12.5%

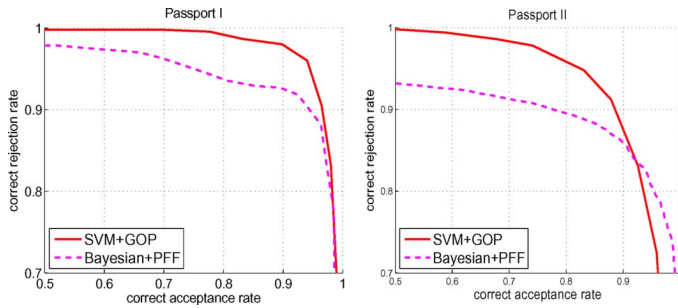


Fig. 5. CRR-CAR curves for experiments with 200 intra and 200 extra pairs for training.

can be nonuniform and saturated. The age differences between image pairs are summarized in Table I. It shows that both datasets have significant age gaps for intrapersonal images. Fig. 3 further shows the distribution of age differences of intrapersonal pairs in the datasets. Intuitively, Passport II is more challenging than Passport I for verification tasks because of the relatively larger age differences. Furthermore, we observed that the image resolution change in Passport II is also larger than that in Passport I.

Figs. 4 and 5 show the CRR-CAR curves for the experiments. In addition, Table II lists the EERs (i.e., when  $CRR=CAR$ ). There are several observations from the experimental results.

First, among the SVM-based approaches, GOP works the best. The gradient direction obviously plays a main role in GOP's excellent performance, since both SVM+GOP and SVM+GO largely outperform SVM+G, which includes the gradient magnitude information. In comparison, the use of a hierarchical structure in GOP further improves upon GO.

Second, SVM+GO greatly outperforms GO. Note that, for face verification, SVM+diff was previously used in [27] and GO was previously used in [6]. This shows that our method, as a combination of these two, greatly improves both of them.

Third, SVM+GOP outperforms the Bayesian approach [30] on both datasets. In addition, from Fig. 5, it is obvious that SVM+GOP is more suitable for passport verification tasks because it performs much better at a high CRR, which is desired as mentioned in Section II-A. Furthermore, given an image pair, our approach does not require the information of which one is older, which is used in the Bayesian approach as a prior.

Fourth, on Passport I, SVM+GOP performs similarly to Vendor A while much better than Vendor B, while on Passport II, SVM+GOP outperforms Vendor A but performs worse than Vendor B (interestingly, the ranks of Vendor A and Vendor B alternate). This observation shows that, though very simple, our approach performs close to commercial systems, which combine many additional heuristic techniques and are well tuned. Furthermore, only low resolution gray images are used in our approach, while the original color images are used in both commercial systems.

### C. Experiments on the FGnet Database

The FGnet Aging Database [1] is widely used for research of age-related facial image analysis. The database contains 1002 images from 82 subjects, over large age ranges. Consequently, there is an average of 12 images per subject in the FGnet database, which is much more than that in the passport databases (only two images per subject). This property makes the FGnet very useful for age progression study such as estimation and simulation. All images in the database are annotated with landmark points, age information, and pose information.

TABLE III  
FGNET DATABASE USED IN FACE VERIFICATION TASKS. “STD.” IS SHORT FOR STANDARD DEVIATION

# subject	# intra pair	mean age	std. age	mean age diff.	std. age diff.
62	665	29.5	11.3	12.3	9.7

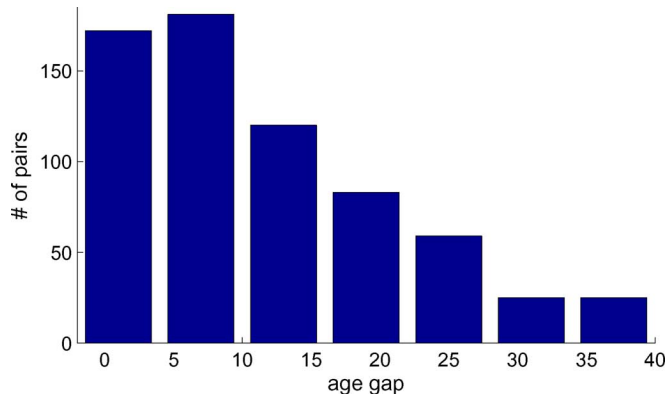


Fig. 6. Distribution of age differences in the FGnet dataset.

TABLE IV  
EERs FOR EXPERIMENTS ON THE FGNET DATABASE [1]

	$l_2$	GO	SVM+diff	SVM+G	SVM+GO	SVM+GOP
EER	40.6%	32.3%	31.2%	28.5%	25.2%	24.1%

We use a subset of the FGnet database that contains only images that are taken above age 18 (including 18) and roughly frontal, which is consistent with the study on the passport databases and in [30]. The effects of aging in children are quite different, and we discuss them in Section V. For notational convenience, we still call this subset FGnet in the following. The subset contains 272 images from 62 subjects. Age statistics of FGnet are shown in Table III and Fig. 6.

We emphasize the importance of experiments on FGnet due to the following reasons:

- 1) FGnet is very challenging for our task in two ways. First, it contains much larger age gaps. The largest gap is 45 years in FGnet, compared to 12 years in the passport databases. Second, the number of subjects is very limited, which makes learning very difficult.
- 2) Since FGnet is a publicly available dataset, experiments on FGnet will serve as a benchmark/baseline for future studies on the topic.

For verification tasks, we generate 665 intrapersonal pairs by collecting all image pairs from the same subjects. Extrapersonal pairs are randomly selected from images from different subjects. Three-fold cross validation is used, such that in each fold images from the same subject never appear in both training and testing pairs. Each fold contains about 220 intrapersonal pairs and 2000 extrapersonal pairs.

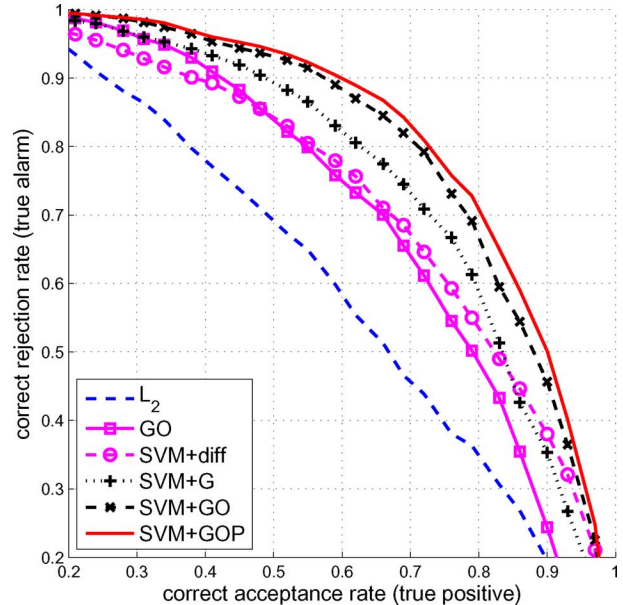


Fig. 7. CRR-CAR curves for three-fold cross validation experiments on FGnet dataset. This figure is better viewed in color.

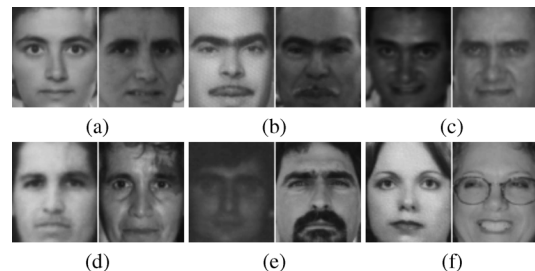


Fig. 8. Example results of SVM+GOP on the FGnet datasets at the EER. (a)–(c) Three correctly accepted intrapersonal pairs. (d)–(f) Three incorrectly rejected intrapersonal pairs. The listed years indicate age gaps in the corresponding pairs. (a) 18 years. (b) 31 years. (c) 7 years. (d) 35 years. (e) 23 years. (f) 32 years.

The experimental results are shown in Fig. 7 and Table IV.<sup>2</sup> Examples of correct as well as incorrect classification for intrapersonal pairs are shown in Fig. 8. The results indicate that, again, the proposed approach outperforms all others. In addition, we also tried combining SVM+GOP with the PointFive Face approach [30] but observed no improvement. This confirms to some degree that our method is insensitive to illumination change, because PointFive Face is designed to be robust to illumination variations.

#### IV. EFFECTS OF AGE PROGRESSION ON VERIFICATION PERFORMANCE

In this section, we empirically study how verification performance is affected by age gaps and related issues, including image quality, presence of eye glasses, and facial hair.

<sup>2</sup>The commercial systems were not available for testing in this experiment.

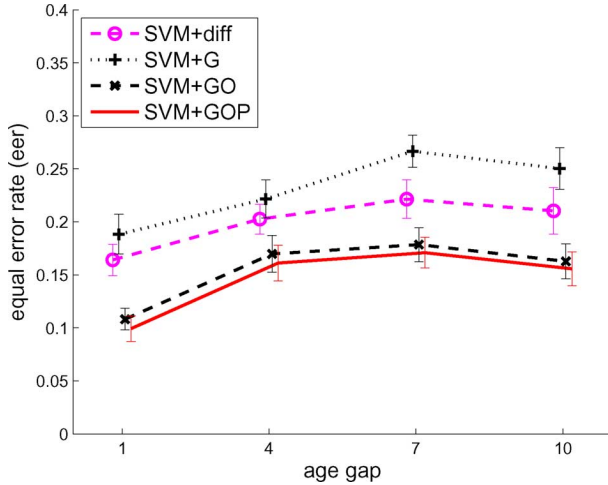


Fig. 9. Effect of aging on verification performance. The curves are shifted a bit along the x axis for better illustration.

### A. Effects of Age Gaps

We are interested in how age differences affect the performance of machine verification algorithms. Taking advantage of the large number of image pairs in Passport II, an empirical study of this problem is conducted.

First, intrapersonal image pairs are grouped into four classes according to their age gaps. Specifically, these are groups with age gaps from 0 to 2 years, 3 to 5 years, 6 to 8 years, and 9 to 11 years. The goal is to test verification performance for different groups. Specifically, we use the average EERs as a criterion. For each group, 80 intra pairs and 80 extra pairs are randomly selected as the training set. Testing sets are created similarly but with 15 intra pairs and 15 extra pairs. There is no overlap between training and testing sets. After that, four SVM-based approaches are tested on the data sets and EERs are recorded. To reduce the variance caused by the lack of training samples, 20 different training/testing sets are generated and the average EERs are recorded. The above experiments have been run 50 times with randomly chosen training/testing sets (i.e.,  $50 \times 20$  training/testing sets). Finally, the mean and standard deviation of EERs are summarized to evaluate the performance.

Fig. 9 shows the performance of the experiments on all four groups. From the plots, we found that faces separated by more than a year are more difficult than those within one year. What surprised us is that the difficulty becomes saturated after the age gap is larger than four years. This phenomenon is observed on all four different representations tested in the experiments.

### B. Effects of Age-Related Issues

When comparing two images of the same person taken at different years, several nonanatomic issues often happen in practice. The FGnet dataset has detailed descriptions associated with each image. Using these descriptions, we analyze the verification results on the FGnet dataset to study the effects of the following three issues: 1) *Quality*, photos taken a long time ago sometimes have poor quality due either to the photographic environment or scanning artifacts. An intrapersonal pair is treated as *high* quality if both photos have good image quality and *low* otherwise. 2) *Glasses*: an intrapersonal pair is treated as *different*

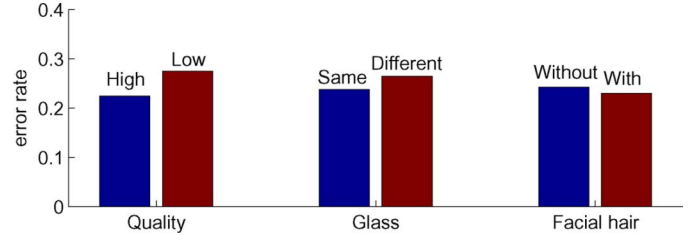


Fig. 10. Error analysis of face verification experiments on the FGnet dataset.

if one photo has spectacles and the other does not. Otherwise, the pair is treated as *same*. 3) *Facial hair*: an intrapersonal pair is treated as *without* facial hair if none of photos has facial hair (including mustache and beard). Otherwise, the pair is treated as *with* facial hair.

Once we have assigned each intrapair with the above labels, we can compare the error verification rate for each label and then compare how related issues affect verification algorithms. For example, the error rate of *high* (quality) inner-pairs is calculated as

$$1 - \frac{\#\text{correctly classified high-quality intra-pairs}}{\#\text{high-quality intra-pairs}}.$$

Fig. 10 shows the error rates of different labels. These error rates are computed using SVM+GOP on the FGnet dataset and taken at the EERs (see Section III). From the figure, we see that low quality and spectacles do increase the difficulties for face verification. However, the proposed SVM+GOP seems to be robust to the presence of facial hair. One reason to this observation is, though facial hair sometimes adds difficulties to verification tasks, they often provide discriminative cues as well. For example, some people have similar beard styles over the years.

## V. FACE VERIFICATION ACROSS AGING IN CHILDREN

The appearance changes of human faces are very different in children than in adults [29]. In this paper, we mainly focus on face images taken above age 18, after which face profiles remain stable [29]. However, it is helpful to understand the performance of the above tested methods on faces from children as well. In this section, we report our experiments on the children face images from the FGnet dataset.

We first extract two face datasets from FGnet, in the same way as in Section III-C. One dataset, named *FGnet-18*, contains 311 face images from 79 subjects, taken at ages in the range [8 18]. The other dataset, named *FGnet-8*, contains 290 face images from 74 subjects, taken at ages in the range [0 8].

For verification tasks, we follow the same scheme as in Section III-C; we generate 577 intrapersonal pairs and 6000 extrapersonal pairs for FGnet-18, and 580 intrapersonal pairs and 6000 extrapersonal pairs for FGnet-8. Three-fold cross validations are conducted for each dataset. Then, the average EERs and CRR-CAR curves are reported in Table V and Fig. 11.

From these experiments, we have the following observations. First, the verification tasks for childrens' faces are much harder than for adult faces. This is clear when we compare results in Tables V and IV. Second, GO-based methods still work well



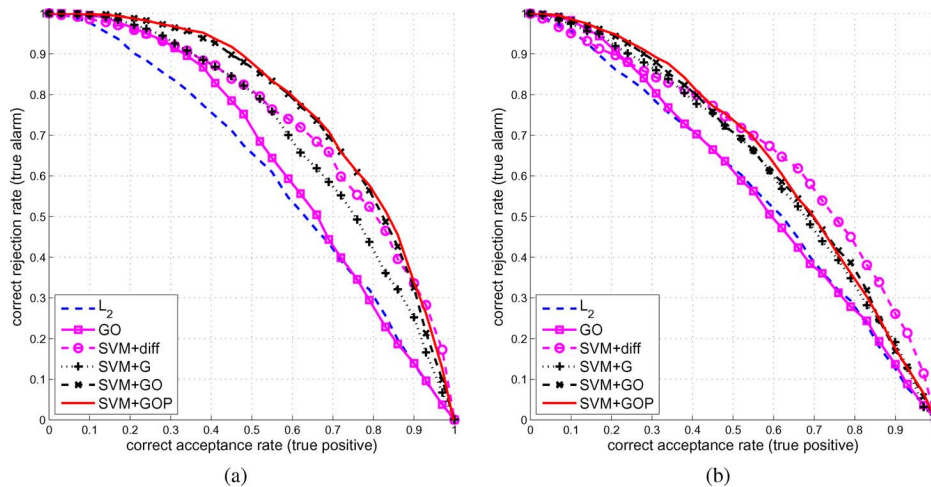


Fig. 11. CRR-CAR curves for three-fold cross validation experiments on the children images of the FGnet dataset. These figures are better viewed in color. (a) FGnet-18 (age range [8 18]). (b) FGnet-8 (age range [0 8]).

TABLE V  
EERs FOR EXPERIMENTS ON THE CHILDREN IMAGES OF FGNET DATABASE [1]

	$l_2$	GO	SVM+diff	SVM+G	SVM+GO	SVM+GOP
FGnet-18	42.9%	40.9%	32.3%	36.1%	30.7%	30.5%
FGnet-8	44.0%	44.6%	36.2%	40.0%	39.8%	38.6%

for age changes of teenagers, though the hierarchical information does not help much any more. Third, the task becomes extremely difficult for small children with ages from 0 to 8, where all methods work poorly.

The major challenge of verifying children faces across aging comes from the alignment problem, because face profiles undergo large variations before age 18. This explains why the intensity- (after normalization) based method, SVM+diff, works relatively better. Generative approaches can provide helpful guidance here, though age information is often requested. It is an interesting future direction to combine generative and discriminative approaches for this task.

## VI. CONCLUSION AND DISCUSSION

In this paper, we studied the problem of face verification with age variation using discriminative methods. First, we proposed a robust face descriptor, the GOP, for face verification tasks across ages. Compared to previously used descriptors such as image intensity, the new descriptor is more robust and performs well on face images with large age differences. In our experiments, with comparison to several techniques, the new approach demonstrated very promising results on two challenging passport databases and the FGnet dataset. In addition, being a discriminative approach, the proposed method requires no prior age knowledge and does not rely on age estimation and simulation algorithms. Second, the effect of the aging process on verification algorithms are studied empirically. In the experiments, we observed that the difficulty of face verification algorithms saturated after the age gap is larger than four years (up to ten years). We also studied the effects of age-related issues including image quality, presence of spectacles, and facial hair.

We plan to investigate several directions in our future work. First of all, testing on a large public dataset will be conducted for

deeper understanding of the proposed approaches. We plan to work on the MORPH dataset [33] for this purpose. Second, we plan to apply other discriminative approaches (e.g., boosting) for simultaneous feature analysis and classification.

## ACKNOWLEDGMENT

The authors would like to thank R. Chellappa for helpful comments and S. Shirdhonkar for help on experiments. The authors would like to thank the editors and all anonymous reviewers for their constructive comments on the early versions of this paper.

## REFERENCES

- [1] Face and Gesture Recognition Working Group 2000 [Online]. Available: <http://www-prima.inrialpes.fr/FGnet/>
- [2] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, "Axioms and fundamental equations of image processing," *Arch. Ration. Mech. Anal.*, vol. 123, no. 3, pp. 199–257, 1993 [Online]. Available: <http://www.springerlink.com/content/t6h8523k61074131>
- [3] M. Bichsel, "Strategies of Robust Object Recognition for the Automatic Identification of Human Faces," Ph.D. thesis, ETH Zurich, Switzerland, 1991.
- [4] S. Biswas, G. N. R. Aggarwal, and R. Chellappa, "A non-generative approach for face recognition across aging," in *2nd IEEE Int. Conf. Biometrics: Theory, Applications and Systems, 2008 (BTAS 2008)*, Oct. 29–1, 2008, pp. 1–6.
- [5] C.-C. Chang and C.-J. Lin, Libsvm : A Library for Support Vector Machines 2001 [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [6] H. Chen, P. Belhumeur, and D. Jacobs, "In search of illumination invariants," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2000, vol. 1, pp. 254–261.
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. 886–893.
- [8] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. 2008.
- [9] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.
- [10] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.
- [11] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 112–119.



- [12] D. K. Hammond and E. P. Simoncelli, *Nonlinear Image Representation via Local Multiscale Orientation* New York University, Courant Institute, Tech. Rep. TR2005-875, 2005.
- [13] K. Jonsson, J. Kittler, Y. Li, and J. Matas, "Support vector machines for face authentication," *Image Vision Comput.*, vol. 20, no. 5-6, pp. 369-375, 2002.
- [14] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Comput. Vis. Image Understanding*, vol. 74, no. 1, pp. 1-21, 1999.
- [15] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621-628, Feb. 2004.
- [16] A. Lanitis, "Comparative evaluation of automatic age-progression methodologies," *EURASIP J. Adv. Signal Process.*, vol. 8, no. 2, pp. 1-10, 2008.
- [17] A. Lanitis, "Evaluating the performance of face-aging algorithms," in *Int. Conf. Face & Gesture Recognition (FG)*, Amsterdam, The Netherlands, Sep. 17-19, 2008.
- [18] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442-455, Apr. 2002.
- [19] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "A study of face recognition as people age," in *Proc. Int. Conf. Computer Vision (ICCV'07)*, 2007, pp. 1-8.
- [20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [21] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748-763, Jun. 2002.
- [22] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696-710, Jul. 1997.
- [23] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," in *Int. Conf. Face & Gesture Recognition (FG)*, 1998, pp. 30-35.
- [24] A. Montillo and H. Ling, "Age regression from faces using random forests," in *IEEE Int. Conf. Image Processing*, Cairo, Egypt, Nov. 7-10, 2009.
- [25] U. Park, Y. Tong, and A. K. Jain, "Face recognition with temporal invariance: A 3d aging model," in *Int. Conf. Face & Gesture Recognition (FG)*, Amsterdam, The Netherlands, Sep. 17-19, 2008.
- [26] E. Patterson, A. Sethuram, M. Albert, K. Ricanek Jr., and M. King, "Aspects of age variation in facial morphology affecting biometrics," in *2nd IEEE Int. Conf. Biometrics: Theory, Applications and Systems, 2007 (BTAS 2007)*, 2007, pp. 1-6.
- [27] P. J. Phillips, "Support vector machines applied to face recognition," in *Advances in Neural Information Processing Systems 16 (NIPS)*, 1999, vol. 2, pp. 803-809.
- [28] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090-1104, Oct. 2000.
- [29] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception," *J. Exp. Psychology: Human Perception and Performance*, vol. 1, no. 4, pp. 374-382, 1975.
- [30] N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3349-3361, Nov. 2006.
- [31] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 1, pp. 387-394.
- [32] N. Ramanathan and R. Chellappa, "Modeling shape and textural variations in aging faces," in *Int. Conf. Face & Gesture Recognition (FG)*, Amsterdam, The Netherlands, Sep. 17-19, 2008.
- [33] K. Ricanek Jr. and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *7th IEEE Int. Conf. Automatic Face and Gesture Recognition (FG 2006)*, Southampton, U.K., Apr. 10-12, 2006, pp. 341-345.
- [34] R. Singh, M. Vatsa, A. Noore, and S. K. Singh, "Age transformation for improving face recognition performance," in *2nd Int. Conf. Pattern Recognition and Machine Intelligence (PREMI)*, 2007, pp. 576-583.
- [35] J. Suo, S. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.* DOI: 10.1109/TPAMI.2009.39.
- [36] J. Suo, X. Chen, S. Shan, and W. Gao, "Learning long term face aging patterns from partially dense aging databases," in *Proc. Int. Conf. Computer Vision (ICCV'09)*, 2009, pp. 622-629.
- [37] N. Tsumura, H. Haneishi, and Y. Miyake, "Independent component analysis of skin color image," *J. Opt. Soc. Amer. A*, vol. 16, pp. 2169-2176, 1999.
- [38] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Horii, and Y. Miyake, "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 770-779, 2003.
- [39] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- [40] J. Wang, Y. Shang, G. Su, and X. Lin, "Age simulation for face recognition," in *18th Int. Conf. Pattern Recognition (ICPR'06)*, 2006, pp. 913-916.
- [41] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proc. Int. Conf. Computer Vision (ICCV'07)*, 2007, pp. 1-8.
- [42] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399-458, 2003.
- [43] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *IEEE Int. Conf. Computer Vision (ICCV)*, 2005, vol. 1, pp. 541-548.



**Haibin Ling** (S'04-M'08) received the B.S. degree in mathematics and the M.S. degree in computer science from Peking University, China, in 1997 and 2000, respectively, and the Ph.D. degree in computer science from the University of Maryland, College Park, in 2006.

From 2000 to 2001, he was an assistant researcher in the Multi-Model User Interface Group at Microsoft Research Asia. From 2006 to 2007, he worked as a postdoctoral scientist at the University of California Los Angeles. After that, he joined Siemens Corporate Research as a research scientist. Since fall 2008, he has been an Assistant Professor at Temple University, Philadelphia, PA. His research interests include computer vision, medical image analysis, human-computer interaction, and machine learning.

Dr. Ling received the Best Student Paper Award at the ACM Symposium on User Interface Software and Technology (UIST) in 2003.



**Stefano Soatto** (S'96-M'03-SM'08) received the D.Ing. degree (highest honors) from the University of Padova, Italy, in 1992, and the Ph.D. degree in control and dynamical systems from the California Institute of Technology in 1996.

He joined UCLA in 2000 after being Assistant and then Associate Professor of Electrical and Biomedical Engineering at Washington University, and Research Associate in Applied Sciences at Harvard University. Between 1995 and 1998 he was also Ricercatore in the Department of Mathematics and Computer Science at the University of Udine, Italy. His general research interests are in computer vision and nonlinear estimation and control theory. In particular, he is interested in ways for computers to use sensory information (e.g., vision, sound, touch) to interact with humans and the environment.

Dr. Soatto is the recipient of the David Marr Prize (with Y. Ma, J. Kosecka and S. Sastry of U.C. Berkeley) for work on Euclidean reconstruction and reprojection up to subgroups. He also received the Siemens Prize with the Outstanding Paper Award from the IEEE Computer Society for his work on optimal structure from motion (with R. Brockett of Harvard). He received the National Science Foundation Career Award and the Okawa Foundation Grant. He is Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI) and a Member of the Editorial Board of the *International Journal of Computer Vision (IJCV)* and *Foundations and Trends in Computer Graphics and Vision*.



**Narayanan Ramanathan** (S'05) received the B.E. (Hons) degree in electrical and electronics engineering from the Birla Institute of Technology and Science, Pilani, in 2002. He received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 2004 and 2008, respectively.

Currently, he is a research scientist at Cernium Corporation, Reston, VA. His research interests span computer vision, pattern recognition, and image processing. He was awarded the university gold medal for the Batch of 2002 from the Birla Institute of Technology and Science.

Royal Institute of Technology (KTH) in Stockholm. In 2002, he joined the faculty of the Computer Science Department at the University of Maryland, College Park. In 2008, he spent a sabbatical at L'Ecole Normale Supérieure de Cachan. His research has focused on human and computer vision, especially in the areas of object recognition and perceptual organization. He has also published articles in the areas of motion understanding, memory and learning, computer graphics, human-computer interaction, and computational geometry.

Dr. Jacobs has served as an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and has assisted in the organization of many workshops and conferences, including serving as Program co-Chair for CVPR 2010. He and his coauthors received honorable mention for the best paper award at CVPR 2000. He also coauthored a paper that received the best student paper award at UIST 2003.



**David W. Jacobs** (S'91–M'92) received the B.A. degree from Yale University in 1982. From 1982 to 1985, he worked for Control Data Corporation on the development of database management systems, and attended graduate school in computer science at New York University. From 1985 to 1992 he attended M.I.T., where he received the M.S. and Ph.D. degrees in computer science.

From 1992 to 2002, he was a Research Scientist and then a Senior Research Scientist at the NEC Research Institute. In 1998, he spent a sabbatical at the