

A Study of Face Recognition as People Age

Haibin Ling^{1*}, Stefano Soatto², Narayanan Ramanathan³, and David W. Jacobs⁴

¹ Integrated Data Systems Department, Siemens Corporate Research, Princeton, NJ

² Computer Science Department, University of California, Los Angeles, CA

³ Electrical and Computer Engineering Department, University of Maryland, College Park, MD

⁴ Computer Science Department, University of Maryland, College Park, MD

{haibin.ling}@siemens.com , {soatto}@cs.ucla.edu, {ramanath,djacobs}@umiacs.umd.edu

Abstract

In this paper we study face recognition across ages within a real passport photo verification task. First, we propose using the gradient orientation pyramid for this task. Discarding the gradient magnitude and utilizing hierarchical techniques, we found that the new descriptor yields a robust and discriminative representation. With the proposed descriptor, we model face verification as a two-class problem and use a support vector machine as a classifier. The approach is applied to two passport data sets containing more than 1,800 image pairs from each person with large age differences. Although simple, our approach outperforms previously tested Bayesian technique and other descriptors, including the intensity difference and gradient with magnitude. In addition, it works as well as two commercial systems. Second, for the first time, we empirically study how age differences affect recognition performance. Our experiments show that, although the aging process adds difficulty to the recognition task, it does not surpass illumination or expression as a confounding factor.

1. Introduction

Face recognition across ages is an important problem and has many applications, such as passport photo verification, image retrieval, surveillance, etc. This is a challenging task because human faces can vary a lot over time in many aspects, including facial texture (e.g. wrinkles), shape (e.g. weight gain), facial hair, presence of glasses, etc. In addition, the image acquisition conditions and environment often undergo large changes, which can cause non-uniform illumination and scale changes. Fig. 1 shows several typical images with different age gaps. Despite its importance, recognition across ages has been relatively less

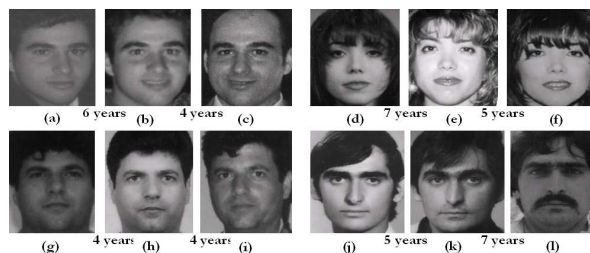


Figure 1. Typical images with age differences. Source images are from the FG-NET Aging Database, <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm>.

studied, which is mainly due to the lack of suitable data sets.

In this paper we study the problem of face recognition across ages and how age gaps affect face recognition tasks. The study is applied to passport photo verification tasks that involve more than 1,800 image pairs, where each pair is the same person taken at different years.

First, we are interested in finding robust face descriptions especially to lighting and aging. Gradient orientation has been shown to be insensitive to lighting change under a Lambertian assumption and has been applied to face recognition [5]. Based on previous study of craniofacial growth [18] and skin color models [8, 22], we conjecture that the gradient direction can be used to build descriptors that are insensitive to the aging process as well. Furthermore, a pyramid technique is used for hierarchical modeling. With this idea, we propose using the *gradient orientation pyramid* (GOP) for robust face representation. Then, given a face image pair, we use the cosines between gradient orientations at all scales to build the “difference” between the pair. Finally, the “difference” is combined with an SVM for face verification tasks in a way similar to [17]. We applied the proposed approach for passport verification tasks and tested it on two passport image datasets with large age differences. Promising results are observed in comparison with several other approaches, including three different rep-

*The work was done when the first author was with University of Maryland College Park and University of California Los Angeles.

representations with the same SVM-based framework (intensity difference [17], gradient with magnitude, gradient orientation), the Bayesian face [19], and two commercial face recognition products. It is also worth noting that our approach is rather simple in comparison to its competitors.

Second, we study how recognition performance varies with increasing time lapses between images. This is the first such study, to our knowledge, and it uses the largest dataset reported in the literature. Surprisingly, we found that the added difficulty of recognition 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 all different image representations that have been tested.

The rest of the paper is organized as follow. Sec. 2 discusses related work. After that, the framework for face verification using a support vector machine is described in Sec. 3. Then, we introduce the gradient orientation pyramid in Sec. 4. Sec. 5 describes our experiments on two passport image datasets involving large age separations. Sec. 6 presents our empirical study of how age gaps affect recognition algorithms. Sec. 7 gives a preliminary study on the relation between aging and gradient orientation. Finally, Sec. 8 concludes.

2. Related Work

Face recognition and detection has been widely studied for several decades. A lot of work has been done to handle the problem under different conditions, including lighting, pose, expression, etc. A thorough survey can be found in [26]. The aging process and its effect on face analysis, which we are interested in, has recently attracted research effort. Most work has focused on modelling the aging process [20], age estimation [10, 12, 19, 27], and simulation [11, 21]. In comparison, face verification across ages is far less studied [19].

Ramanathan and Chellappa [19] adapted the probabilistic eigenspace framework [16] for face identification across age progression, which is most closely related to our work. Instead of using a whole face, only a half face (called a PointFive face) is used to alleviate the non-uniform illumination problem. Then, eigenspace techniques and a Bayesian model are combined to capture the intra-personal and extra-personal image differences. Targeting the same task, our work differs from their work in both the representation (we use gradient orientation pyramids) and the classification frameworks (we use SVM).

Modelling face verification as a two-class classification problem is not new. Moghaddam et al. [15] used a Bayesian framework for the intra-personal and extra-personal face classification. Phillips [17] used SVM for face recognition problems and observed good results on the FERET dataset compared to component based approaches. Jonsson et al. [9] used SVM for face authentication problems. Our work

is different in that we use the gradient orientation pyramid instead of intensity differences [15, 17] or the intensity itself [9] as a face description. Furthermore, we are more interested in passport photos with age differences.

Image gradients are widely used for feature building and image representation, e.g. [13, 6, 25]. In most of these works, the gradient magnitude information is included (e.g. to build a weighted histogram). The direction of image gradient has been proposed for lighting insensitive recognition (e.g. [3]) and was shown to be insensitive to changes in lighting direction under a Lambertian assumption [5]. Recently, Hammond and Simoncelli [7] proposed a nonlinear image representation using only the orientation information. Inspired by these works, we propose using the gradient orientation for robust face representation with age variation. To the best of our knowledge, it is the first time the gradient direction is combined with SVM for face verification problems. Our experiments show that by discarding magnitude information, the gradient orientation achieves significantly better recognition performance. We also propose using hierarchical structure to further improve discriminability.

Our work also relates to work on skin appearance [8]. It is known that melanin and hemoglobin are the two most important factors that determine human skin color. Tsumura et al. [22] used independent component analysis (ICA) to separate the effect of melanin and hemoglobin as two independent components from skin color. They showed that the logarithm of skin color can be decomposed as linear combinations of two components corresponding to melanin and hemoglobin respectively. We show that this fact is closely related to the insensitivity of gradient orientation across the aging process.

3. Problem Formulation

3.1. Task Description

Passport photo verification is important in the process of passport renewal and related face authentication applications. For example, when a person submits a new photo for renewal, the ideal system can automatically tell whether he is an imposter by comparing the new photo to previous photos that were usually taken years before.

A common way to evaluate verification uses two criteria: the rate of correct rejection on imposter images (correct reject rate) and the rate of correct acceptance (correct accept rate) on true images. These two rates conflict although we want both rates to be as high as possible. In practice, for the passport verification task, the correct reject rate is most important because rejected images will be examined by a human. For example, we can fix such a rate at a high level (e.g. 99%) while making the correct acceptance rate as high as possible, which measures the human labor that is saved.

3.2. Classification Framework

We model face verification as a two-class classification problem [15, 17, 9]. Given an input image pair I_1 and I_2 , the task is to assign the pair as either *intra-personal* (i.e. I_1 and I_2 from the same people) or *extra-personal* (i.e. I_1 and I_2 from the different individuals). In this section we briefly describe the framework of using a support vector machine for this task. Details about support vector machines can be found in [24].

Given any image pair (I_1, I_2) , it is first mapped it onto the feature space. Formally, we have $\mathbf{x} = \mathcal{F}(I_1, I_2) \in \mathbb{R}^d$, where \mathbf{x} is the feature vector extracted from the image pair (I_1, I_2) through the feature extraction function (functional) $\mathcal{F}(\cdot, \cdot)$, and \mathbb{R}^d is the d -dimensional feature space.

Then the support vector machine is used to divide the feature space into two classes, one for intra-personal pairs and the other for extra-personal pairs. Using the same terminology as in [17], we denote the separating boundary with the following equation

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

where N_s is the number of support vectors and s_i is the i -th support vector. Δ is used to trade off the correct reject rate and correct accept rate as described in Sec. 5. $K(\cdot, \cdot)$ is the kernel function that provides SVM with non-linear abilities. The RBF kernel is chosen for our task due to its effectiveness and efficiency. The RBF kernel is defined as

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

where γ is a parameter determining the size of RBF kernels ($\gamma = \frac{1}{d}$ is used in our experiments).

4. Gradient Orientation Pyramid

There is one question left open in the previous section: what is $\mathcal{F}(\cdot, \cdot)$? A natural choice is to use the intensity difference between I_1 and I_2 , which is called *difference space* in [15] and also has been used in [19, 17]. With an appropriate normalization scheme, the intensity difference can be made robust to affine lighting changes. However, the affine lighting model is not always sufficient for face images, especially for images taken at times separated by years. Instead, to alleviate this problem, we propose using the cosine of gradient orientations as features of image pairs, because gradient orientation is known to be insensitive to lighting change [5]. In addition, we organize the gradient orientation in a hierarchical fashion, which further improves its discriminability.

4.1. Gradient Orientation Pyramid

Our proposed features are partly motivated by recent work using gradient information for object representations

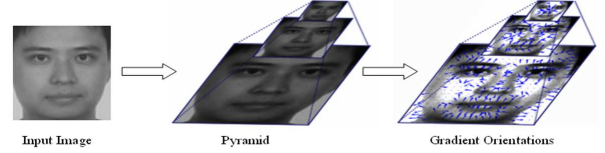


Figure 2. Computation of a GOP from an input image. Note: 1) In the right figure, the gradient orientations at “flat” regions are excluded. 2) The right figure is made brighter for better illustration.

([13, 6]). In these works, the gradient directions were weighted by gradient magnitudes. In contrast, inspired by [5], we discard magnitude information and use only orientations, which demonstrates significant improvement in our experiments (Sec. 5). 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 $\mathcal{P}(I) = \{I(\mathbf{p}; \sigma)\}_{\sigma=0}^s$ as

$$\begin{aligned} I(\mathbf{p}; 0) &= I(\mathbf{p}) \\ I(\mathbf{p}; \sigma) &= [I(\mathbf{p}; \sigma - 1) * \Phi(\mathbf{p})] \downarrow_2 \quad \sigma = 1, \dots, s \end{aligned} \quad (3)$$

where $\Phi(\mathbf{p})$ is the Gaussian kernel (0.5 is used as the standard deviation in our experiments), \downarrow_2 denotes half size downsampling, and s is the number of pyramid layers. Note that in (3) the notation I is used both for the original image and the images at different scales for convenience.

Then, the gradient orientation at each scale σ is defined by its normalized gradient vectors at each pixel.

$$g(I(\mathbf{p}; \sigma)) = \nabla(I(\mathbf{p}, \sigma)) / |\nabla(I(\mathbf{p}, \sigma))| \quad (4)$$

Naturally, the *gradient orientation pyramid* (GOP) of I , is defined as $\mathcal{G}(I) = \{g(I(\mathbf{p}, \sigma))\}_{\sigma=0}^s$. Fig. 2 illustrates the computation of a GOP from an input image.

4.2. Differences Between GOPs

Given an image pair (I_1, I_2) and corresponding GOPs $(\mathcal{G}(I_1), \mathcal{G}(I_2))$, the feature vector $\mathbf{x} = \mathcal{F}(I_1, I_2)$ is computed as the concatenation of the cosines of the difference between gradient orientations at each pixel and all scales. The computation can be efficiently achieved through the inner product of the corresponding entries of GOP, i.e., for pixel \mathbf{p} at scale σ , it is computed as

$$f(I_1(\mathbf{p}; \sigma), I_2(\mathbf{p}; \sigma)) = g(I_1(\mathbf{p}; \sigma)) \cdot g(I_2(\mathbf{p}; \sigma)) \quad (5)$$

The cosine values are organized into a feature vector \mathbf{x} as

$$\mathbf{x} = \mathcal{F}(I_1, I_2) = (\dots, f(I_1(\mathbf{p}; \sigma), I_2(\mathbf{p}; \sigma)), \dots)^T \quad (6)$$

where \mathbf{p} is organized in lexicographic order and σ in increasing order.

We summarize the advantages of using GOPs for face verification tasks as follow.

- GOP is insensitive to illumination changes [5]. As a result, no normalization is needed on the input images.
- The pyramid technique provides a natural way to perform face comparison at different scales.
- The inner product between normalized gradients lies in a finite range $([-1, 1])$ that automatically limits the effect of outliers.

In the following, we use SVM+GOP to indicate the proposed approach.

5. Verification Experiments

5.1. 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 [19]. It contains 452 intra-personal image pairs (several duplicate pairs were removed) and 2251 randomly generated extra-personal image pairs. Passport II contains 1824 intra-personal image pairs and 9492 randomly generated extra-personal image pairs. Images in both datasets are scanned passport images. They are in general frontal images with small pose variations. The lighting condition varies, and can be non-uniform and saturated. The age differences between image pairs are summarized in Table 1. It shows that both datasets have significant age gaps for intra-personal images. Fig. 3 further shows the distribution of age differences 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.

In our experiments (SVM-based approaches), the images are preprocessed using the same scheme as in [19]. This includes manual eye location labelling, alignment by eyes and cropping with an elliptic region. Image sizes are reduced to 96×84 for Passport I and 72×63 for Passport II. Example images at these resolution are shown in Fig. 8, where we can also see the “disappearance” of wrinkles. While relatively larger resolutions could be used to make use of more information that is still insensitive to wrinkles (e.g. 160×140 as in Fig. 8 (c)), we run out of memory with them in the current Matlab-based implementation.

To alleviate the alignment problem, when comparing two GOPs, we tried different alignments with small shiftings (two pixels). In our experiments it helped to improve the performance by around 0.5% (equal error rate). A similar technique is used by [14].

Table 1. Passport datasets for identification tasks. “Std.” is short for standard deviation.

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

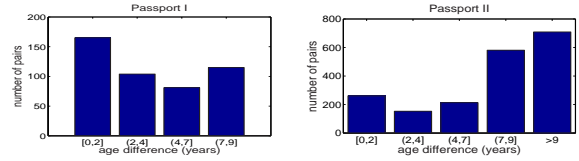


Figure 3. Distribution of age differences in the passport image databases. Left: Passport I. Right: Passport II.

5.2. Methods

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 gradient orientation (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 gradient orientation. It can also be viewed as weighting gradient orientations with gradient magnitudes. 4) SVM+diff [17]. As in [17], we use the differences of normalized images as input features combined with SVM. 5) GO: this is the method using gradient orientation proposed in [5]. 6) Bayesian+PFF [19]. This is the approach combining Bayesian framework [16] and PointFive Face (PFF) [19]. In addition, two commercial systems are tested on the datasets, which we will name Vendor A and Vendor B¹.

The first approaches use exactly the same configurations and the same SVM framework, but different representations. The purpose is to study the superiority of the proposed GOP representations. The other four approaches are different from our method in both representations and classification frameworks.

5.3. Experimental Evaluation

For verification tasks, the correct reject rate (CRR) and the correct acceptance rate (CAR) are two critical criteria, which are defined as

$$\begin{aligned}
 CRR &= \frac{\# \text{ correct rejected extra-personal pairs}}{\# \text{ total extra-personal pairs}} \\
 CAR &= \frac{\# \text{ correct accepted intra-personal pairs}}{\# \text{ total intra-personal pairs}}
 \end{aligned} \quad (7)$$

The performance of algorithms is evaluated using the CRR-CAR curves that are usually created by varying some classifier parameters. As mentioned in Sec. 3.1, although both rates are desired to be as high as possible, the more important is the performance when CRR is high.

¹Anonymous due to agreements with the companies.

We used three-fold cross validation in our experiments. For each experiment, the CRR-CAR curve is created by adjusting parameter Δ in (1). The total performance is evaluated as the average of the output CRR-CAR curves of three folds. For Vendor 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 [19], i.e., we use 200 positive and 200 negative pairs as a training set.

Fig. 4 and Fig. 5 show the CRR-CAR curves for the experiments. In addition, Table 2 lists the equal error rates (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. Later we will show that the improvement is more obvious when larger image resolution is used.

Second, SVM+GO largely outperforms GO. Note that, for face recognition, SVM+diff is previously used in [17] and GO is previously used in [5]. This shows that our method, as a combination of these two, greatly improves both of them.

Third, SVM+GOP outperforms the Bayesian approach [19] 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 correct reject rate, which is desired as mentioned in Sec. 3.1. 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.

In addition to the verification experiments mentioned above, another two experiments have been done to further understand the proposed approach. One of them studies the hierarchical information captured by GOP. The other tries to find irrelevant features on the face that are related to wrinkles to some degree.

Study on Hierarchical Information. As mentioned above, the improvement from adding hierarchical information is not significant due to the limited image sizes. To further

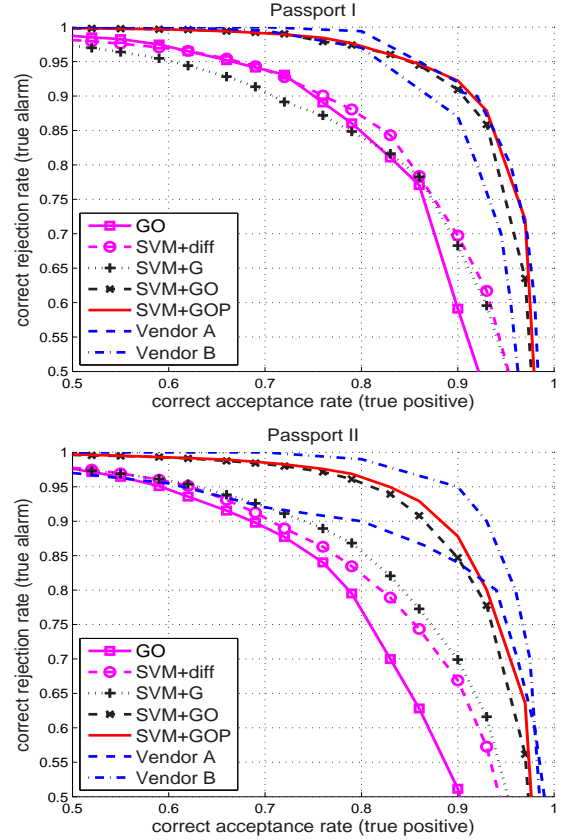


Figure 4. CRR-CAR curves for three-fold cross validation experiments. Top: on Passport I. Bottom: on Passport II. This figure is better viewed in color.

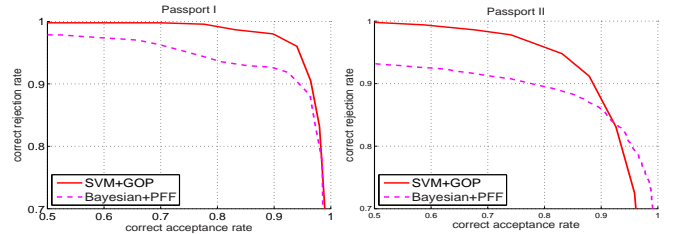


Figure 5. CRR-CAR curves for experiments with 200 intra- and 200 extra-pairs for training.

study the effect of the pyramid, we conducted experiments comparing SVM+GOP and SVM+GO on larger resolution images. We used randomly chosen subsets of Passport I and Passport II due to memory limitations. Specifically, for Passport I, one third of the image pairs are used with the size 207×180 . Note that the wrinkle effect at this resolution may still exist (e.g. Fig. 8 (b)), but we focus on the study of hierarchical information instead. For Passport II, one fourth of the image pairs are used with the size 160×140 (wrinkles almost disappear at this resolution, e.g. Fig. 8 (c)). The improvement of GOP from GO for different resolutions are summarized in Table 3. The table shows that the improve-

Table 2. Equal error rates. Left table: experiments of three-fold cross validation. Right table: experiments using 200 intra- and 200 extra-pairs as training, as in [19].

	GO [5]	SVM+diff [17]	SVM+G	SVM+GO	SVM+GOP	Vendor A	Vendor B	SVM+GOP	Bayesian [19]
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%

ment of using pyramid is larger when higher resolution images are used. This validates including hierarchical information in the GOP representation.

Table 3. The improvement of GOP from GO. The content inside each cell is the difference of equal error rates from GO to GOP.

Reso.	96×84	207×180	Reso.	160×140	72×63
Pass. I	0.6%	1.9%	Pass. II	0.8%	1.6%

Study of Irrelevant Wrinkle Related Features. This experiment provides a preliminary study of how wrinkle related regions affect the recognition system. For this task, we use two feature masks (Fig. 6). One is the usual ellipse mask. The second one removed the forehead and cheeks, which are known to carry much wrinkle information [4] but little shape information. We tested SVM+GO and SVM+diff with both masks. The performance curves shown in Fig. 6 show that there is no performance drop by removing the wrinkle regions. This suggests that SVM is already learning to ignore these regions. Or it might mean that the noise added by wrinkles balances whatever positive information is there.

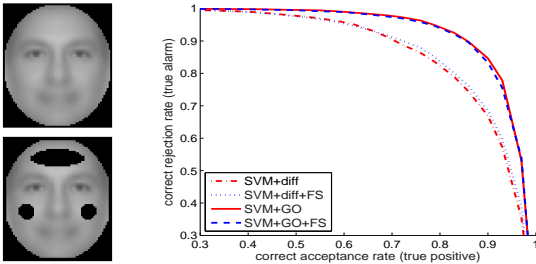


Figure 6. Experiment on irrelevant wrinkle region selection. Left: two masks with resolution 72×63 (enlarged for better illustration). Right: performance curves, where “+FS” means with feature selection masks (the bottom one on the left).

6. Effects of Age Gaps on Recognition Performance

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

First, intra-personal image pairs are grouped into four classes according to their age gaps. Specifically, these are groups with ages gaps from 0 to 2 years, 3 to 5 years, 6 to

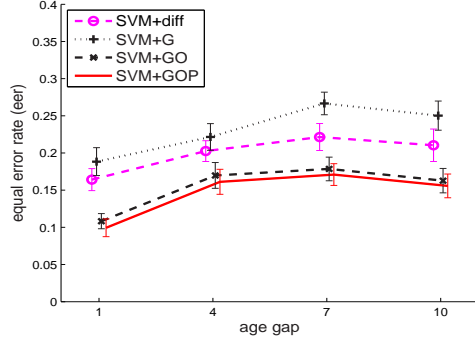


Figure 7. Effect of aging on recognition performance. The curves are shifted a bit along the x axis for better illustration.

8 years, and 9 to 11 years. The goal is to test the recognition performance for different groups. Specifically, we use the average equal error rates as a criterion. For each group, 80 intra pairs and 80 extra pairs are randomly selected as training set. Testing set 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 equal error rates are recorded. To reduce the variance caused by the lack of training samples, 20 different training/testing sets are generated and the average equal error rates are recorded. The above experiments have been run 50 times with randomly chosen training/testing sets (i.e., 50×20 training/testing sets). Finally, the mean and standard deviation of equal error rates are summarized to evaluate the performance.

Fig. 7 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. It shows that, when the age gap is larger than four years (up to ten years), the difficulty added by aging is not a dominant factor compared to other nuisance factors such as illumination, expression, etc.

7. Gradient Orientation Under Aging Progress

As shown in Sec. 5, gradient orientation yields a very robust representation for faces undergoing aging. Though we have not developed rigorous theoretic support for this, we give some intuitive explanations of why the gradient orien-

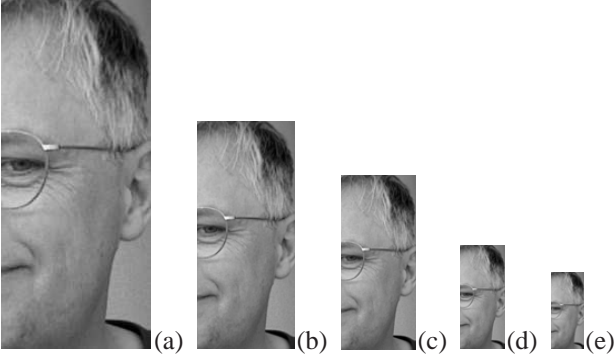


Figure 8. Perception of wrinkles at different resolution (in pixels) of a face image at age 50. The resolutions in (b-e) will be used in our experiments (see Sec. 5). (a) Original image, height=347. (b) Height=207. (c) Height=160. (d) Height=96. (e) Height=72.

tation is robust to aging for face recognition tasks.

The change of skin appearance over time is very complicated [8]. Discriminative invariance to aging is very difficult to achieve (if it exists). Focusing on aging problem and based on the observation from our datasets, we made following assumptions. 1) We assume that all faces are frontal and there is no alignment problem. With this assumption, we can ignore viewpoint issues in image formation. 2) Reflectance can be modelled as Lambertian, which has been widely and successfully done for face recognition tasks. 3) We are interested in faces with age ranging from 20 to 70. All these assumptions are roughly satisfied by the passport photos used in our experiments.

With the above assumptions, we model a human face as a surface $\xi(\mathbf{x}, t) = \{\mathbf{n}(\mathbf{x}, t), \rho(\mathbf{x}, t) : \mathbf{x} \in \Omega\}$ described by surface normals $\mathbf{n}(\mathbf{x}, t)$ and albedos $\rho(\mathbf{x}, t)$ at age or time $t \geq 0$. For recognition tasks, $\Omega \subset \mathbb{R}^2$ is the shared support for both the face surface and the corresponding image plane. With the Lambertian model, the image $I(\mathbf{x}, t)$ of a face $\xi(\mathbf{x}, t)$ is generated by the following formula

$$I(\mathbf{x}, t) = \rho(\mathbf{x}, t) \langle \mathbf{n}(\mathbf{x}, t), \mathbf{l}(t) \rangle \quad \forall \mathbf{x} \in \Omega, t \geq 0 \quad (8)$$

where $\mathbf{l}(t) \doteq \sum_i \mathbf{l}_i(t) \in \mathbb{R}^3$, and $\mathbf{l}_i(t)$ are different lighting sources. Next, we will discuss how \mathbf{n} and ρ change over time with respect to recognition tasks.

For the task of face recognition, we can roughly treat the surface normal $\mathbf{n}(\mathbf{x}, t)$ as consistent over age gap Δt (e.g. $\Delta t = 10$ in our experiments). This is based on the following arguments. First, previous study of craniofacial growth [18] has shown that face profiles (e.g. boundary shapes, eye locations) are relatively stable for people older than 18 year. This motivates assumption 3. Second, it has been shown that wrinkles are one of the most important factors for the visual perception of age [1]. In fact, the width and depth of wrinkles grow roughly linearly with age [1]. While wrinkles do affect the surface normals, the effects are largely reduced due to following facts.

- Wrinkles are hardly perceptible for images with low resolution [10], which are used for most recognition tasks. For example, the wrinkles on the face in Fig. 8 “disappear” when the resolutions decreases.
- The distribution of wrinkles on human faces is known to be nonhomogeneous. For example, Boissieux et al. [4] summarized eight generic wrinkle maps from the qualitative data from L’Oreal. From the studies in cosmetics, we know that wrinkles appear frequently on forehead and cheeks, which are not important for face recognition. This means that most regions with heavy wrinkles can be ignored by face recognition algorithms, either through manually adjusted masks or by automatic feature selection (e.g. through SVM). One support comes from the experiment in Sec. 5. It compares face recognition with forehead and cheeks to recognition without them, and observes no significant difference in recognition performance.

As a result, we can roughly write $\mathbf{n}(\mathbf{x}, t) \approx \mathbf{n}(\mathbf{x})$ and simplify (8) as²

$$I(\mathbf{x}, t) \approx \rho(\mathbf{x}, t) \langle \mathbf{n}(\mathbf{x}), \mathbf{l}(t) \rangle \quad \forall \mathbf{x} \in \Omega, t \geq 0. \quad (9)$$

The gradient orientation is known to be robust under the Lambertian lighting model [5, 2], which is true if we assume ρ is independent of t . Alternatively, if $\rho(\mathbf{x}, t)$ is a homogeneous scaling function across t , then the gradient of $I(\mathbf{x}, t)$ is invariant across t . Specifically, we want ρ to have the following decomposition

$$\rho(\mathbf{x}, t) \approx a(t) \rho_0(\mathbf{x}) \quad (10)$$

where $a(t)$ is the positive scaling function and $\rho_0(\mathbf{x})$ is a “reference” albedo (equivalently, (10) can also be written as $\rho(\mathbf{x}, t) \approx a(t - t_0) \rho(\mathbf{x}, t_0)$).

Next we will show the support for (10) from studies of skin optics. Studies in skin optics [8] show that skin color is mainly determined by two kind of chromophores, melanin and hemoglobin. Tsumura et al. [22] used ICA to separate the effect of melanin and hemoglobin as two independent components of the logarithm of skin color. By adjusting the amount of the two components, their approach is successfully applied to skin color synthesis tasks [23]. With their model, the skin color $\mathbf{c}(\mathbf{x}) \in \mathbb{R}^3$ can be derived as³

$$\mathbf{c}(\mathbf{x}, t) = \mathbf{c}_1^{p(\mathbf{x}, t)} \mathbf{c}_2^{q(\mathbf{x}, t)} \mathbf{c}_3 \quad (11)$$

where $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^3$ correspond to melanin and hemoglobin respectively. p and q correspond to the amount of associated pigments at time t and are mutually independent. \mathbf{c}_3 denotes a constant shift. As a result, the change of skin

²Strictly speaking, we should use $\Omega_0 \subset \Omega$ in (9) for feature selection.

³In (11) and (12) the product between vectors are element wise.

color over time is roughly determined by the change of p and q .

In tasks where age gaps are limited (e.g. $t \leq 10$ in our passport verification task, Sec. 5.1), the change of q is small. This is because hemoglobin reflects blood color that is relative stable. As a result, we approximate q with, $q(\mathbf{x}, t) \approx q(\mathbf{x})$. Under the same assumption, and considering the fact that the change of melanin is mainly due to exposure to sunlight (for healthy skin), we approximately treat the change of p as additive, $p(\mathbf{x}, t) \approx p_0(\mathbf{x}) + p_1(t)$. Therefore, (11) can be approximated as

$$c(\mathbf{x}, t) \approx c_1^{p_0(\mathbf{x})} c_1^{p_1(t)} c_2^{q(\mathbf{x})} c_3 \doteq a(t) c_0(\mathbf{x}) \quad (12)$$

where $a(t) \doteq c_1^{p_1(t)}$ and $c_0(\mathbf{x}) \doteq c_1^{p_0(\mathbf{x})} c_2^{q(\mathbf{x})} c_3$. Equation (12) is close to (10). We expect future works to reveal deeper connections between this decomposition and the aging insensitivity of gradient orientations.

8. Conclusion and Discussion

In this paper we studied the problem of face recognition with age variation. First, we proposed a robust face descriptor, the gradient orientation pyramid, 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 addition, the pyramid technique enables the descriptor to capture hierarchical facial information. In our experiments with comparison to several techniques, the new approach demonstrated very promising results on two challenging passport databases. Second, the effect of the aging process on recognition algorithms are studied empirically. In the experiments we observed that the difficulty of face recognition algorithms saturated after the age gap is larger than four years (up to ten years).

Acknowledgement

We thank Rama Chellappa for helpful comments and Sameer Shirdhonkar for help on experiments. Ling, Ramanathan, and Jacobs were supported in part by a fellowship from Apptis Inc.. Ling and Soatto were supported in part by ONR 67F-1080868 and NSF ECS-0622245.

References

- [1] S. Akazaki, H. Nakagawa, H. Kazama, O. Osanai, M. Kawai, Y. Takema, and G. Imokawa. "Age-related changes in skin wrinkles assessed by a novel three-dimensional morphometric analysis", *British J. of Dermatology*, 147(4):689-95, 2002. 7
- [2] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel. "Axioms and fundamental equations of image processing", Springer Berlin / Heidelberg, 1993. 7
- [3] M. Bichsel, "Strategies of Robust Object Recognition for the Automatic Identification of Human Faces", *PhD. thesis*, ETH Zurich, No. 9467, 1991. 2
- [4] L. Boissieux, G. Kiss, N. Magnenat-Thalmann, and P. Kalra. "Simulation of skin aging and wrinkles with cosmetics insight". *Eur. Wksp Animation & Simulation*, 15, 2000. 6, 7
- [5] H. Chen, P. Belhumeur and D. W. Jacobs. "In search of Illumination Invariants", *CVPR*, 254-261, 2000. 1, 2, 3, 4, 5, 6, 7
- [6] N. Dalal and B. Triggs. "Histograms of Oriented Gradients for Human Detection," *CVPR*, I:886-893, 2005. 2, 3
- [7] D. K. Hammond and E. P. Simoncelli. "Nonlinear Image Representation via Local Multiscale Orientation". *Courant Institute Technical Report*, TR2005-875, 2005. 2
- [8] T. Igarashi, K. Nishino, and S.K. Nayar. "The Appearance of Human Skin," Technical Report, Dept. of Computer Science, Columbia University CUCS-024-05, Jun, 2005. 1, 2, 7
- [9] K. Jonsson, J. Kittler, Y. Li and J. Matas. "Support vector machines for face authentication", *BMVC*, 543-553, 1999. 2, 3
- [10] Y. H. Kwon and N. da Vitoria Lobo. "Age classification from facial images". *CVIU*, 74:1-21, 1999. 2, 7
- [11] A. Lanitis, C. J. Taylor, and T. F. Cootes. "Toward Automatic Simulation of Aging Effects on Face Images". *IEEE Trans. PAMI*, 24(4):442-455, 2002. 2
- [12] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation", *IEEE Trans. SMC B*, 34(1):621-8, 2004. 2
- [13] D. Lowe, "Distinctive Image Features from Scale-Invariant Key-points," *IJCV*, 60(2), pp. 91-110, 2004. 2, 3
- [14] A. Martinez. "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class". *IEEE Trans. PAMI*, 24:748-763, 2002. 4
- [15] B. Moghaddam, W. Wahid, and A. Pentland. "Beyond Eigenfaces: Probabilistic Matching for Face Recognition". *Intl. Conf. on Face and Gesture Recognition*, 30-35, 1998. 2, 3
- [16] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation", *IEEE Trans. PAMI*, 19(7):696-710, 1997. 2, 4
- [17] P. J. Phillips. "Support vector machines applied to face recognition", *NIPS*, 2:803-809, 1999. 1, 2, 3, 4, 5, 6
- [18] J. B. Pittenger and R. E. Shaw. "Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception". *J. of Exp. Psychology: Human Perception and Performance*, 1(4):374-382, 1975. 1, 7
- [19] N. Ramanathan and R. Chellappa, "Face Verification across Age Progression", *IEEE Trans. on Image Processing*, 15(11):3349-3361, 2006. 2, 3, 4, 5, 6
- [20] N. Ramanathan and R. Chellappa, "Modeling Age Progression in Young Faces", *CVPR*, I:387-394, 2006. 2
- [21] J. Suo, F. Min, S. Zhu, S. Shan, and X. Chen, "A Multi-Resolution Dynamic Model for Face Aging Simulation", *CVPR*, 2007. 2
- [22] N. Tsumura, H. Haneishi, and Y. Miyake. "Independent component analysis of skin color image". *J. of Optical Society of America A*, 16:2169-2176, 1999. 1, 2, 7
- [23] N. Tsumura, N. Ojima, K. Sato, M. Shiraiishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake. "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin". *ACM Trans. Graph. (SIGGRAPH 2003)*, 22(3):770-779, 2003. 7
- [24] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995. 3
- [25] L. Zelnik-Manor and M. Irani, "Event-Based Analysis of Video", *CVPR*, I:123-130, 2001. 2
- [26] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld. "Face recognition: A literature survey". *ACM Comp. Surv.*, 35(4):399-458, 2003. 2
- [27] S. K. Zhou, B. Georgescu, X. Zhou and D. Comaniciu, "Image Based Regression Using Boosting Method," *ICCV*, I:541-548, 2005. 2