Perspective Distortion Modeling, Learning and Compensation

Joachim Valente
Google
Mountain View, CA
joachim.valente@google.com

Stefano Soatto
UCLA Vision Lab
University of California, Los Angeles
soatto@ucla.edu

Abstract

We describe a method to model perspective distortion as a one-parameter family of warping functions. This can be used to mitigate its effects on face recognition, or synthesis to manipulate the perceived characteristics of a face. The warps are learned from a novel dataset and, by comparing one-parameter families of images, instead of images themselves, we show the effects on face recognition, which are most significant when small focal lengths are used. Additional applications are presented to image editing, videoconference, and multi-view validation of recognition systems.

1. Introduction

The "dolly zoom" is a cinematic technique whereby the distance to a subject is changed along with the focal length of the camera, while keeping its image size constant. It is also known as "vertigo effect," from Hitchcock's classic movie, and exploited by artists to manipulate the subject's perceived character (Fig. 1 top). As evidenced by psychophysical experiments [4, 6, 20, 21], the subject can appear more or less attractive, peaceful, good, strong, or smart depending on the distance to the camera and its focal length.

Figure 1: *(Top) sample images from our focal-distorted face dataset. It is worth emphasizing that there is no artificial warp or optical aberration, and the perceived difference among the various samples is due solely to the distance. (Bottom) sample images used as dictionary samples.*

Just as it affects perception, perspective distortion can affect the performance of any face recognition system. Our *first goal* in this manuscript is to quantify such an effect (Table 1). This is done by testing different face recognition algorithms on images captured under different focal settings than those used for training. This requires a dataset of images of the same subjects taken from different distances. Given the absence of such a dataset in the public domain, we designed and collected a novel one.

Having quantified the effect, our *second goal* is to model perspective distortion, and to learn the model parameters from the training set. It is worth emphasizing that perspective distortion is not an artificial warp or an optical aberration, but a complex deformation of the domain of the image due to the combined physical effects of distance and focal length. It depends on the shape of the underlying face, which is typically unknown, and can involve singularities and discontinuities.[1] Nevertheless, it can be approximated by a one-parameter family of shape-dependent domain deformations. This model enables hallucination of perspective distortion, even without knowledge of the underlying shape.

We illustrate this task by interactively manipulating the perceived distance from the camera. In particular, we demonstrate "focal un-distortion" of videoconference and videochat images, that are often perceived as unattractive due to the short focal length of forward-looking cameras in consumer devices.

Our *third* and final goal is to exploit the structure of our model to render face recognition systems insensitive to perspective distortion. This is done by performing comparisons between image families, rather than between images themselves. We validate this method by testing the same face recognition systems studied in our first goal, where each family is represented by a canonical element computed via pre-processing.

1.1. Related Work

An application of this work is to face recognition, a field too vast to properly review here (see [29] for a survey of

[1] For instance, the ears of the subject are visible on the right in Fig. 1 but not on the left.

978-1-4673-6759-2/15/$31.00 ©2015 IEEE

the state-of-the-art as of a decade ago, and [1] for a more recent account). Since our goal is not to introduce a new face recognition algorithm, but to devise a method for any face recognition system to deal with perspective distortion, we select two representative algorithms in section 3.1. One is chosen for simplicity, the other because representative of the state-of-the-art.

More specifically, our work aims to reduce nuisance variability. A nuisance is a phenomenon that affects the data but should ideally not affect the task. Most prior work on handling nuisances in face recognition focused on illumination [11, 23, 2, 13] and pose variation [13, 5], as well as partial occlusion [28], age [18, 15] and facial expressions [1]. To the best of our knowledge, variability due to optics has not been studied in a systematic way, and while its effects on recognition is not as dramatic as illumination or pose variability, it nevertheless can exceed intra-individual variability and thus lead to incorrect identification, especially at short focals.

Many face datasets for recognition are publicly available. The *FERET database* [22], the *AR-Face database* [16] or the *Extended Yale Face Database B* [14] are among the most widely used to benchmark face recognition algorithms. A more thorough review is done in [1]. Despite the number of available datasets, to the best of our knowledge, none tackles the problem of optical zoom variability.

Additionally, our method requires the distance from the subject in the training set to be known or estimated. [9] tackles the problem of estimating this distance by solving the camera pose via Effective Perspective-n-Point. We however do not leverage 3D modeling and use a different method reminiscent of deformable templates instead (section 4.3). Using this estimate to improve face recognition in presence of perspective distortion is also suggested there.

The psychophysical effects of perspective distortion have been studied in [4]. It is shown to be a crucial factor affecting how a subject is perceived, notably how trustworthy, competent and attractive she looks like. The idea of using some kind of quantification of the perspective distortion, to manipulate the perceived personality, is also mentioned. Inspired by paintings from the Renaissance that use several centers of projection at once to control the viewer's perception, [21] studies how the same effect can be achieved with photographs and shows compelling experiments by combining multiple images of the same scene (a human) taken from different viewpoints, using an image editing software.

[25] describes a system to solve perspective distortion in videochat applications. The method differs from ours in that it relies on matching a 3D face template to the image, and generate a reprojected image as if viewed from a farther viewpoint.

1.2. Organization of This Paper and Its Contributions

In section 2 we describe the dataset we have collected to test the hypothesis that warping due to perspective distortion affects the performance of face recognition. There we further explain the reasons that motivate it, and detail the protocol used.

In section 3 we quantify the impact of perspective distortion on face recognition by comparing the performance of several algorithms when the test image was captured from the same distance as training images and when it was captured from a different distance. We show that the effect is negligible when the distance used in both sets is above half a meter, but significant otherwise.

In section 4 we begin addressing the issue of managing nuisance variability due to perspective distortion. The derivation we propose is generic, in the sense that it applies to any one-parameter group transformation, and in fact even higher-dimensional groups, provided that the dataset spans a sufficiently exciting sample of the variability. Other examples of applications that we have not considered in this work, but where our method could in principle be applicable, include aging and expression, but not pose changes that induce self-occlusions.

We present our results in section 5, both qualitatively (i.e. visually) and quantitatively (i.e. showing numerical improvements on face recognition success rate). There, we also show an application to un-warping of videoconference and videochat images, to illustrate the synthesis component (as opposed to recognition) of our method. Finally in section 6 we discuss possible extensions and applications.

2. Dataset

For testing the hypothesis that perspective distortion affects face recognition, we have generated a protocol and constructed a dataset that comprises 12 images each for over 100 subjects. Most subjects are in their twenties, Caucasian or Asian, with about 47 % females. The dataset spans 7 focal lengths and 5 different expressions for each subject, and is captured against a green screen with photographic studio quality but otherwise uncontrolled illumination.

2.1. Focal-Distance Relation

Throughout this work, we assume that the distance between the subject and the center of projection (COP) is varied along with focal length so that the face occupies the same area on the image plane. More precisely, under a simplistic optical model, for an aspect ratio of 3:2, this correspondence is given by

$$d = f\frac{\sqrt{13}hK}{2\gamma_{35}} \tag{1}$$

where d is the distance from the subject to the COP, f is the focal length of the lens, h is the height of the face (typically around 19 cm), K is the crop factor of the image sensor and γ_{35} is the diagonal of a full-frame 35 mm (36 mm $\times$ 24 mm), i.e. $\gamma_{35} = 43.3$ mm.

Based on this relation we will use the terms "focal" and "distance" interchangeably, although the source of variability is really the distance. The term "focal" will be preferred because easier to control during the construction of the dataset.

2.2. Dataset Requirements

As we explain in section 4, our method relies on averaging the dependency on the shape of the underlying face, which improves with the number of samples in the dataset. Of our set, 33 % are to be used in the learning phase, and 67 % in testing.

Also, our method models the warping by learning it on face images where perspective distortion is the only nuisance. Therefore illumination is assumed to be constant within the training set, pose is frontal and expression neutral. The images we collect span from wide-angle (10 mm or distance of 12.7 cm) to telephoto (70 mm or distance of 88.6 cm) in a fine-grained fashion (7 focals in our case).

We also need an additional 5 pictures of each subject with different expressions to serve as dictionary (on the 67 % subjects not used during learning). The 7 focal-varying images will serve as test samples.

2.3. Protocol

Each individual was asked to sit on a stool lit on both sides by a 70 W softbox RPS Studio RS-4070 to reduce the effects of cast shadows. Behind them was a green screen to remove background variability.

The camera used was a Canon EOS 30D. For wide-angles we used a Canon lens EF-S 10-22 mm f/3.5-4.5 USM and for medium-range we used a Canon lens EF 25-70 mm f/2.8L II USM. The sensor's crop factor is $K = 1.6$. All photos were shot at 1/60, f5.6, ISO 400 with a white balance fixed at 5000 K. However to ensure uniformity images were further processed to adjust brightness and contrast.

In a first stage subjects were asked to remove their glasses and if needed to put their hair up so as not to hide the eyes and eyebrows. They had to look towards the camera with a frontal pose and neutral expression, but the latter was not strictly enforced, resulting in some minor expression variability. Seven photos were taken in sequence with focals 10 mm, 17 mm, 22 mm, 24 mm, 34 mm, 50 mm and 70 mm.

Then in a second stage they were asked to smile, to vary expressions, to look at a fixed object, resulting in about 30° out-of-plane rotation, to show a neutral frontal expression, and finally to make a "funny face" (akin to the "joker" expression in the IMM Face Database [17]).

In a post-processing step, all images were normalized and aligned with respect to the similarity group by placing the eyes in canonical position, as customary. Fig. 1 shows the resulting 12 samples for one of the 100 subjects used in this work.

3. Impact of Perspective Distortion on Face Recognition

In this section we examine how the particular variability due to perspective distortion influences recognition success rate. Although many algorithms are designed to be insensitive to various sources of variability, we show that in practice extreme distortions lead to incorrect identifications.

3.1. Face Recognition Algorithms

We will consider two families of face recognition systems. The first one (EIGENDETECT) is chosen for simplicity, based on the assumption that a linear subspace captures the within-class variability, as suggested in [26, 24]. The resulting "eigenfaces" then capture the principal components of the space spanned by the samples in the training database.

The second algorithm is considered representative of the state-of-the-art and based on sparse representation coding (SRC) [28]: given learnt faces $(I_i)_{i=1}^n$ put side by side in a dictionary-matrix A, solve the ℓ^0 minimization problem

$$\min \|x\|_0 \text{ s.t. } Ax = I \,. \tag{2}$$

This NP-complete problem is relaxed to an ℓ^1 minimization which naturally yields a sparse vector x. Lastly we compute the per-subject residuals r_k for all labels k, defined as the norm of the difference between Ax and $A\hat{x}_k$ where $\hat{x}_k$ is x for components that correspond to subject k and 0 elsewhere. The output is the subject with lowest residual. Since no code is provided, we implemented our own version matching the same success rate on standard datasets claimed by the authors.

SRC actually projects images on a low-dimensional subspace (e.g. $\mathbb{R}^{120}$) both for speed issues and efficiency reasons. This projection can be done in several ways, including downsampling (SRC+DOWNSAMPLE), masking to isolate a part of the face (SRC+MASK) or using "randomfaces" (projection using a random matrix).[2] In the mask version we isolated the right eye and the mouth in order to study the class of algorithms that only rely on local features.

Both EIGENDETECT and SRC work well on simple datasets like the *AT&T Laboratories Cambridge Face Dataset* (respectively 94.38 % and 95.62 %) but they differ on challenging ones like the *Extended Yale Face Database*

[2] However this randomness introduces excessive variance between runs and therefore is not used in this work.

Focal length	EIGEN	SRC+D	SRC+M
10 mm	52.24 %	82.09 %	41.79 %
17 mm	77.61 %	91.04 %	76.12 %
22 mm	79.10 %	94.03 %	77.61 %
24 mm	91.04 %	98.51 %	82.09 %
34 mm	86.57 %	100 %	89.55 %
50 mm	88.06 %	98.51 %	89.55 %
70 mm	86.57 %	100 %	85.07 %

Table 1: *Success rate for three face recognition algorithms (*EIGENDETECT, SRC+DOWNSAMPLE, SRC+MASK*) for each focal length,* 70 mm *being the reference focal length. The learning set is composed of 5 images of each individual with different expressions. The success rate is defined as the number of correctly identified subjects over the number of subjects.*

B [14] (respectively 38.38 % and 90.98 %). Our goal is to show that managing perspective distortion *improves* performance, so the actual performance figure is irrelevant other than for serving as a baseline. Indeed, we will see that both are affected by perspective distortion, especially from short distances.

3.2. Experiments

We used the 5 expression-varying images as dictionary samples (neutral, smiling, angry, looking left and "joker"). Those photos were shot with a focal of 70 mm (thereafter called the *reference focal*). Then we ran 7 recognition tasks, one for each focal length, over the last 67 % subjects of the dataset (the first 33 % being reserved for face warping modelization). We repeated the experiment for the three algorithms considered. The results are summarized in table 1.

Success rate is at most slightly affected for focals close to the reference focal, but dramatically drops with short focals. A wide-angle (10 mm) produces distortions that significantly decrease recognition rate, even for state-of-the-art algorithms (e.g. 41.79 % instead of the nominal 89.55 % for SRC+MASK).

4. Learning Perspective Distortion

In this section we describe a method to hallucinate image domain deformations due to changes in frontal distance. In a first step we suppose that the initial focal is known (*e.g.,* from EXIF metadata). Then we solve the problem where the initial focal is unknown. In more general terms, the method allows to generate the family spanned by a single data point under a one-parameter group transformation, without other knowledge

(3)

4.1. Formalization

4.1.1 Image Formation

With a simplified formalism that does not involve illumination, pose and noise, the image of a face taken with focal f can be written:

$$I_f(x) = I_{f_0}(w_f(x)), \;\; x \in D \tag{4}$$

where w_f can be viewed as a *warp* from the image lattice D to itself and f_0 is the reference focal. A derivation of this formalism is given in the supplementary material.

Our goal in this section is thus: *given an image of a face* $I : D \to \mathbb{R}^3$*, corresponding to a known or unknown focal* f_0*, find the set of functions* $\{w_f : D \to D\}$ *modeling perspective distortions for any focal* f*.*

4.1.2 Representation of a Face

The warps w_f depend on the shape of the face but not its albedo. For this reason we can discard the albedo information in our representation of a face. We only wish to represent the shape S. Explicit reconstruction could be employed here, even though the absence of viewpoint variability makes it entirely dependent on priors [3, 12, 19]. To avoid that, and for simplicity, we consider the warp a function of the hidden variable S, represented by a few sample points within. Active appearance models (AAM) [8, 7] can then be employed to fit a template on unseen faces. The points fitted via AAM are thereafter called *landmarks*. In practice we used $N = 64$ landmarks, delineating the eyebrows, the eyes, the nose and nostrils, the mouth and the outline of the face (see supplementary material). As customary, we remove the affine component (the mean) but rather than doing so across the entire dataset, we index the mean by focal length:

$$I_f \equiv X_f - \overline{X_f} = \Delta X_f \in \mathbb{R}^{2N} \tag{5}$$

where $X = [x_{1x}\; x_{1y}\; \dots\; x_{Nx}\; x_{Ny}]^\top$ and $\overline{X_f}$ is the average face at focal f.

4.1.3 Assumptions on Warps

To go further we need to make basic assumptions of regularity on the warps w_f. Namely we assume that, as a function of ΔX, a warp is a diffeomorphism[3] from $\mathbb{R}^{2N}$ to itself. We can then write the linear approximation:

$$\begin{aligned} \Delta X_f &= w_f(\Delta X_{f_0}) \\ &= w_f(0) + Dw_f(0)^\top \Delta X_{f_0} + \mathcal{O}(\|\Delta X_{f_0}\|^2) \end{aligned} \tag{6}$$

[3] In reality it is sufficient for the warp to be differentiable on $\mathbb{R}^{2N}$.

where $\|\cdot\|$ is some norm on $\mathbb{R}^{2N}$. This approximation is valid so long as faces are "close" to the average face, which should be the case in practice.

By letting $b_f \triangleq w_f(0)$ and $A_f \triangleq Dw_f(0)^\top$ we obtain the following affine approximation:

$$\Delta X_f \approx A_f \Delta X_{f_0} + b_f \ . \tag{7}$$

4.2. Learning the Model

4.2.1 Face Warping as a Quadratic Minimization Program

Eq. (7) gives a convenient way to warp any face taken at focal f_0 to its counterpart at focal f. Unfortunately we cannot compute A_f and b_f because they depend on the unknown function w_f. However, since they do not depend on the face itself, we wish to learn them using a sufficient number of samples.

To that end we want to minimize the quantity

$$\sum_{i=1}^{n_T} \|A_f \Delta X^i_{f_0} + b_f - \Delta X^i_f\|^2$$

with n_T being the number of training samples and the norm being the Euclidian norm. However this problem is typically under-constrained because there are $2N(2N+1)$ free variables and each subject contributes $2N$ constraints. To avoid overfitting it is necessary to regularize the elements of A and b. We naturally want to encourage a matrix A close to the identity and b close to zero, because this corresponds to w_f being the identity, and even though a face undergoes important changes that motivate this work, it should stay close to itself through perspective distortion. Note that we need to learn a matrix A and a vector b for each pair of parameters (f_1, f_2). We thus propose to solve the following quadratic minimization program:

$$A_{f_1\to f_2},\ b_{f_1\to f_2} = \underset{A,\ b}{\operatorname{argmin}}\ q(A,\ b) \tag{8}$$

where

$$q(A,\ b) = \sum_{i=1}^{n_T} \|A\Delta X^i_{f_1} + b - \Delta X^i_{f_2}\|^2 + \lambda\|A - I\|^2 + \mu\|b\|^2 \ . \tag{9}$$

Lagrange multipliers λ and μ are selected via grid search, using 67 % of the training data for learning and 33 % for cross-validation. Once λ and μ are selected, we learn A and b again over the entire training data. In practice we used $\lambda = 10^5$ and $\mu = 10^{-2}$.

4.2.2 Interpolation Between Focals

The quadratic program (8) enables the transformation from any parameter f_1 to any other parameter f_2 for which we have data. Obviously data is only collected for a small sample of focal lengths.

Provided that the sampling is fine enough, and that the sensitivity of $A_{f_1\to f_2}$ and $b_{f_1\to f_2}$ to source focal f_1 and destination focal f_2 is smooth, bilinear interpolation can be used to approximate $A_{f\to f'}$ and $b_{f\to f'}$ for any focals (f, f'). Should the sampling be too coarse, one can resort to finer methods, such as cubic spline interpolation.

4.3. When the Source Focal is Unknown

So far we have seen how to hallucinate an image $I_{f'}$ of a face at any focal f' given the image I_f, provided we know the source focal f. In typical applications we may not know this focal and therefore need to infer it. Formally, we seek a function $\phi : \mathbb{R}^{2N} \to \mathbb{R}$ such that $|\phi(X_f) - f| < \eta$ with high probability for some tolerance $\eta \in \mathbb{R}^+$. The tolerance depends on the sensitivity of A and b to the source and destination focals. Indeed mistaking f_1 for f_2 may be tolerable if $A_{f_1\to f'} \approx A_{f_2\to f'}$ and $b_{f_1\to f'} \approx b_{f_2\to f'}$.

Several approaches can be considered. Provided the focal space is sufficiently densely sampled and the data is clustered by focal length (which seems suggested by [9]), a nearest-neighbor search can be attempted. However our data did not prove clustered enough and a reliable estimate of the focal length could not be obtained. Linear SVM approaches also proved insufficient. To deal with the non-linearity of the data, we instead trained a neural network with one hidden layer containing 4 nodes.[4] This leads to an RMS error of 13.17 mm on the testing data, which is surprisingly accurate given that beyond a threshold, even a trained human cannot achieve such precision.

4.4. Comparison of Families for Perspective Distortion Mitigation

We saw in section 3 that both a basic and a state-of-the-art algorithms occasionally fail when shown faces taken from an unusual standpoint. To address this issue, based on the interpretation of perspective distortions being a one-parameter group transformation, we propose to *compare families spanned by images, rather than images themselves.* This idea is common in applications where the data is acted upon by a group [10].

A distance between families can be defined by minimizing over all possible group actions: If (I_1, I_2) are two images that we want to compare, and $[I]$ is the family spanned by I under the action of the group, then we can define a distance between families via

$$d([I_1], [I_2]) = \min_{I'_1\in[I_1],\ I'_2\in[I_2]} d_0(I'_1, I'_2) \tag{10}$$

where d_0 is a base distance in the data space. This however requires solving an optimization problem at decision time.

[4] More complex architectures, e.g. three hidden layers with 32, 16 and 8 nodes, also gave good results but took longer to train.

Alternatively, one can exploit the fact that each family is an equivalence class, which can be represented by any of its elements. So long as it is possible to select a unique "canonical element," one can simply compare canonical elements (eq. 11). This does not entail any optimization, and this is the approach we take, with the canonical element being the mapping of an image to reference focal length. This can be seen as a pre-processing step, after which the warped image can be fed to any standard face recognition system. In practice the focal estimation takes $0.3\,\mathrm{s}$ and the actual warp $2.0\,\mathrm{s}$ for a 256×256 frame on consumer hardware.

$$d([I_1],[I_2]) = d_0(\hat{I}_1,\hat{I}_2)\ . \tag{11}$$

5. Experimental Assessment

5.1. Qualitative Results

As suggested in [4], extrapolation of perspective distortions can be applied directly to image editing. The warp described in section 4 can easily be extrapolated by letting the source and destination landmarks be respectively control points and their images, and using a thin-plate spline [27] to obtain a dense warp. We implemented this solution in a Matlab GUI application (see Fig. 2) that allows warping an input image into its hallucinated version at any focal in the range [10 mm ; 70 mm].

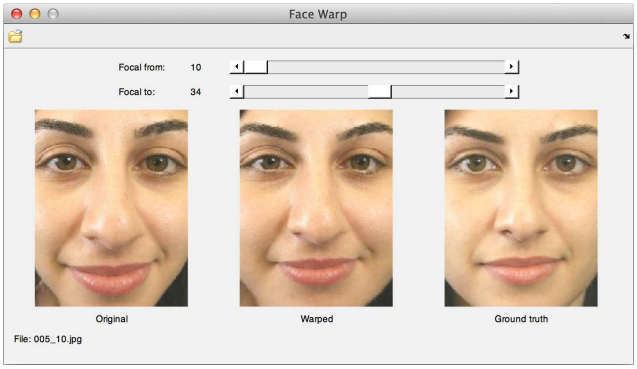


Figure 2: *Face warping GUI. The handles allow to correct the source focal and to control the destination focal. The first panel is the input, the middle panel is the warped face image. When a ground truth face is available it is displayed on the third panel.*

In Fig. 3 we show an application to un-distortion of videoconference streams. In this proof-of-concept demonstration, it is assumed that a detector/tracker yields a smooth estimate of the location of the eyes. Landmarks are fitted using AAMs. The distance to the screen (and hence the "focal") is simply estimated using the distance between the eyes, since the focal of the camera is known, and the face is then warped to the desired viewing distance. This application enables mitigating the undesirable effects of the typical optics employed in forward-looking cameras on mobile devices and tablets.

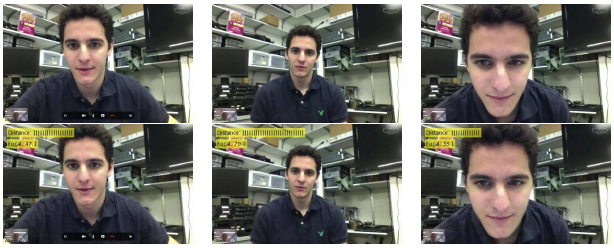


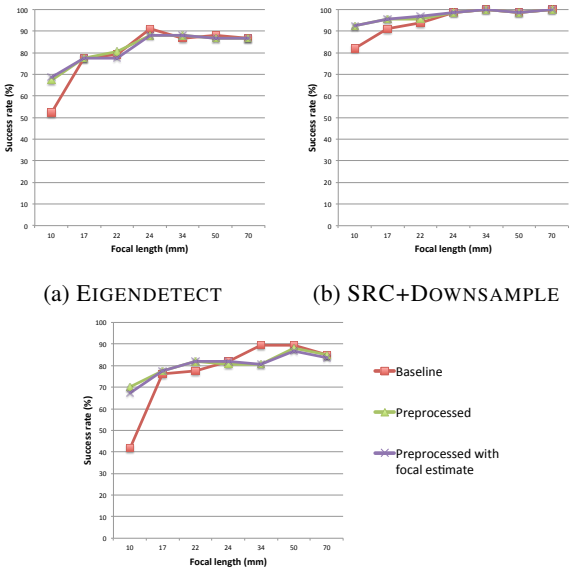
Figure 3: *Application of face unwarping to videoconference streams. (Top) original frames 1, 115 and 403. (Bottom) unwarped versions. Since the focal is known (*$30\,\mathrm{mm}$ *in 35-mm equivalent), the image uncropped and the face frontal, the distance from the subject is estimated using the distance between the eyes, and is then converted to an estimated "focal" using formula 1. The face is then warped to* $f = 63\,\mathrm{mm}$ *which corresponds to a viewing distance of* $50\,\mathrm{cm}$*. Not counting the detection of the eyes and the fitting of AAMs, the application runs at an average rate of* $4.1\,\mathrm{s}$ *per* 432×270 *frame, time mostly spent for resampling in Matlab's affine transformation function and for thin-plate spline interpolation. The warp is only applied to the face area and smoothly vanishes on its edges by fixing control points on them before using thin-plate spline interpolation.*

5.2. Managing Perspective Distortion in Face Recognition and Validation

To illustrate the mitigation of perspective distortion in face recognition, we conducted two experiments. In the first one we pre-processed the images by warping them from their true focal length (known in our dataset) to the reference focal length. In the second experiment we do not suppose the focal length known and instead estimate it as explained in section 4.3. Fig. 4 summarizes improvement of success rate by comparing the three experiments: without pre-processing, with pre-processing when focal is known and with pre-processing when focal is estimated.

The most noticeable results appear for the extreme focal length $f = 10\,\mathrm{mm}$. Because of huge distortions happening at this distance, algorithms perform at their worst. Our method compensates for these distortions and allows to achieve higher success rates. Above a certain threshold, perspective distortion becomes negligible and, as expected, our method only produces negligible random fluctuations. Note that our focal estimate is reliable enough to give results that are almost as good as when the focal is known.

In a final proof-of-concept experiment (Fig. 5), we take the opposite approach where the focal length is known and controlled by the system. Because the warp induced by perspective distortion is shape-dependent, it is possible to capture multiple images at different focal lengths, rescale them, and then test the compatibility of the resulting deformation



(a) EIGENDETECT (b) SRC+DOWNSAMPLE

(c) SRC+MASK

Figure 4: *Success rate of face recognition algorithms with and without pre-processing.*

with the shape of the underlying scene. This would allow validation of the identity of a face in a way that a single-image based recognition system cannot do (even the best face recognition system based on a single view cannot discriminate between an image of a person and an image of an image of a person). Practically, the application estimates the warping between images taken from different distances and validates or invalidates the output of the underlying single-view recognition system.

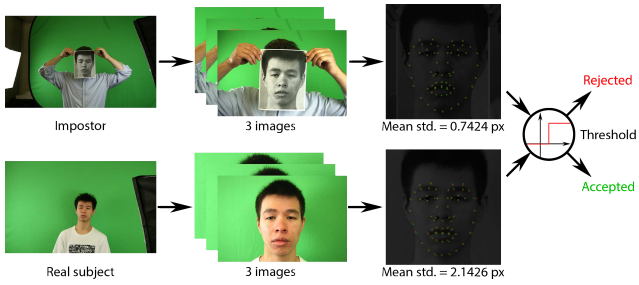


Figure 5: *Multi-view validation of an underlying single-view recognition system. In this scenario, an impostor uses a photograph pretending to be some authorized subject. The camera controls its own viewing distance and focal length and triggers the shutter from different distances. After scaling and processing, the standard deviation of each landmark trail, averaged over all landmarks, can be thresholded to unveil the impostor. Single-view approaches would inevitably fail here.*

5.3. Limitations

The effects of perspective distortion in face recognition are modest for long focals. Certainly they are not as deleterious as out-of-plane rotations, occlusions, and illumination changes, but nevertheless significant, as they affect the performance of face recognition systems, especially at close distances where such deformations exceed inter-class variability.

It would be tempting to extend the method presented in 4.3 to estimate the distance from the COP to the subject. This could be of interest in image forensics. However the effects of perspective distortion become negligible for distances beyond a few meters, and our method would not be of any use for such a purpose.

The application of the warps for synthesis purposes (Fig. 2, 3) requires the location of the face to be known to high accuracy. When the focal needs to be estimated, an accurate location of the fiducial points is also required (because this involves non-linear steps sensitive to small variations). As a result a proper implementation of a system like the one in Fig. 3 would require on-line accurate face detection and tracking and possibly other pre-processing to warp the face to fronto-parallel, and would fail altogether in the presence of significant out-of-plane rotation that yields self-occlusions. Lastly, one would probably want to segment the face from the background to avoid warping the latter.

The videoconference demonstration in Fig. 3 may seem superfluous in actual scenarios where participants are typically far from the camera. However it addresses a real-world, large scale problem when applied to personal videochat contexts, or in "selfie" mobile applications, in which one cannot back off from the camera more than an arm's length [25].

6. Conclusion

We study the effects of varying distance in frontal face images. While such variations have significant perceptual impact, and have been exploited by artists for centuries [21], an explicit modeling and a quantitative assessment of this phenomenon and its impact on face recognition have not been attempted before.

It is also possible to employ the system for synthesis purposes to modify the appearance of a photograph or a video as if it was taken from a different distance, thereby manipulating a person's perceived qualities.

The methodology developed could be extended to other families of one-parameter transformations, assuming that they yield differentiable and differentially-invertible warps, which is not the case in the presence, for instance, of occlusions. This includes self-occlusions from out-of-plane rotation.

References

[1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, 2007. 2

[2] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):721–732, 1997. 2

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999. 4

[4] R. Bryan, P. Perona, and R. Adolphs. Perspective distortion from interpersonal distance is an implicit visual cue for social judgments of faces. *PloS one*, 7(9):e45301, 2012. 1, 2, 6

[5] X. Chai, S. Shan, X. Chen, and W. Gao. Locally linear regression for pose-invariant face recognition. *Image Processing, IEEE Transactions on*, 16(7):1716–1725, 2007. 2

[6] E. A. Cooper, E. A. Piazza, and M. S. Banks. The perceptual basis of common photographic practice. *Journal of vision*, 12(5), 2012. 1

[7] T. F. Cootes, G. J. Edwards, C. J. Taylor, et al. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001. 4

[8] G. J. Edwards, T. F. Cootes, and C. J. Taylor. Face recognition using active appearance models. In *Computer VisionâĂŤECCVâĂŹ98*, pages 581–595. Springer, 1998. 4

[9] A. Flores, E. Christiansen, D. Kriegman, and S. Belongie. Camera distance from face images. In *Advances in Visual Computing*, pages 513–522. Springer, 2013. 2, 5

[10] U. Grenander. *Elements of pattern theory*. JHU Press, 1996. 5

[11] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *Audio-and Video-Based Biometric Person Authentication*, pages 10–18. Springer, 2003. 2

[12] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang. Automatic 3d reconstruction for face recognition. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 843–848. IEEE, 2004. 4

[13] F. J. Huang, Z. Zhou, H.-J. Zhang, and T. Chen. Pose invariant face recognition. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 245–250. IEEE, 2000. 2

[14] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):684–698, 2005. 2, 4

[15] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs. A study of face recognition as people age. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. 2

[16] A. M. Martinez. The ar face database. *CVC Technical Report*, 24, 1998. 2

[17] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, may 2004. 3

[18] U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5):947–954, 2010. 2

[19] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 296–301. IEEE, 2009. 4

[20] P. Perona. A new perspective on portraiture. *Journal of Vision*, 7(9):992–992, 2007. 1

[21] P. Perona. Far and yet close: Multiple viewpoints for the perfect portrait. *Art & Perception*, 1(1-2):105–120, 2013. 1, 2, 7

[22] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, 2000. 2

[23] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pages 157–164. IEEE, 2003. 2

[24] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *JOSA A*, 4(3):519–524, 1987. 3

[25] B. Super, B. Augustine, J. Crenshaw, E. Groat, and M. Thiems. Perspective improvement for image and video applications, Aug. 19 2010. US Patent App. 12/772,605. 2, 7

[26] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991. 3

[27] G. Wahba. *Spline models for observational data*, volume 59. Siam, 1990. 6

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009. 2, 3

[29] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003. 1