Exploration of the Correlations of Attributes and Features in Faces

Matthaeus Schumacher[1] and Volker Blanz[1]

[1] Institute for Vision and Graphics, University of Siegen

Abstract— **What are the attributes and features of faces that allow humans or machines to make most reliable inferences from visible to occluded regions of the face, or from shape to texture and vice versa? While both the Human Visual System and many example-based algorithms rely on correlations, these are implicit and difficult to visualize. This paper identifies and visualizes the most reliable correlations using a canonical correlation analysis (CCA) of faces in a 3D Morphable model. We investigate correlations between shape and texture, but also between shape of mouth and shape of eyes / lower and upper facial shape / overall shape and eye or mouth, and we separate intrinsic correlations from random correlations in the training set. By projecting the CCA axes on semantic attributes such as "large eyes" or "wide lips", they can be partly translated into verbal descriptions.**

Using an algorithm that fills in missing information in faces, such as occluded regions, based on PCA or CCA, and a subsequent assessment of perceived similarity, we evaluate the benefit of CCA over PCA. There is no evidence of CCA being superior, which means that PCA captures correlation sufficiently and is not affected by spurious random correlations in the limited training set.

I. Introduction

When we look at a human face in a front view image, most of us feel confident to guess what the face would look like in a side view, and sometimes we may be surprised if individuals have an unexpected profile when they turn their heads. The same is true for our ability to "fill in" missing regions in images of faces: if we see a person with sunglasses, we may guess what they look like when they take off the glasses, or if we see only the eyes of a someone wearing a motorbike helmet, we form a mental image of their head shape.

In computer vision, similar tasks have been addressed with methods such as *Hallucinating Faces* [1]: Statistical representations of the visual appearance of faces can be used to fill in missing areas in images or missing detail in low quality images. For a recent survey, see [2]. The problem if inferring depth from front view images has been investigated in computer vision and graphics by [3], and a comparison between computational methods and human expectation has been presented in [16].

Computational methods rely mostly on first order correlations between coordinates and colors of facial feature points in datasets of face images or scans. [16] have shown that this approach is consistent with the behavior of human participants. However, it remains unclear what exactly these correlations are, and how strong and reliable they are.

In this paper, we strive to isolate the most relevant correlations of global or local attributes of faces from a dataset of 3D scans in a 3D Morphable Model (3DMM, [3]). Principal Component Analysis (PCA) is a standard technique to exploit correlations in data. However, visualized principal components may mislead us into false conclusions on what exactly is correlated in faces, and what isn't: Consider a set of 2D vectors $(x, y)^T$ in a symmetrical normal distribution. The principal component with highest variation may be any 2D vector, for example the vector $(1, 1)^T$, and from this we may conclude that x and y are correlated, which in fact they aren't. Only by looking at the second component (which would be an orthogonal vector $(1, -1)^T$) we would see that this is not true. For high-dimensional data, it is difficult to distinguish true correlations from false ones just by looking at the principal components: The first component makes faces smaller and more round, but do we know if other components, combined, account for the opposite effect?

While it is easy to calculate the correlation between attributes of faces in a data set, our task is more difficult: Find the pair of attributes in two modalities (shape versus texture, front versus side view, upper versus lower half of the face, eyes versus mouth) that have highest correlations. The background of this question is: if we are to make inferences from one to the other, what are the attributes we should rely on? And what are the rules that humans may have learned and that they apply when they imagine new views that they haven't seen?

In this paper, we adapt *canonical correlation analysis* (CCA) to explore and visualize correlations between different part of a face or between different modalities. The CCA was introduced by Hotelling [11] and is a common statistical method that estimates linear correlations between two multidimensional variables. In the last decades it has been widely used in several scientific fields such as economics [9], medical studies [4] and even in classification of malt whiskys [14]. But also in computer vision and pattern recognition, CCA was used for solving different tasks. [8] applied CCA for learning filters for multidimensional signal processing, and [13], for example, used CCA to locate pixels in video frames that are correlated with sound of the recorded scene. Since CCA only handles linear correlations, [15] introduced a Kernel CCA that estimates non-linear correlations, and [19] used this kernel based method to recognize facial expressions.

In psychology, an *interactive activation model*, developed by Burton et al [5] and [6], allows a computational description and simulation of the human cognitive systems in terms of various face recognition tasks.

A. 3D Morphable Model

The 3D Morphable Model of 3D faces (3DMM, [3]) is a statistical model that captures the range of natural faces in terms of 3D shapes and textures. It is derived from a dataset of 200 3D scans of faces, including 100 female and 100 male faces. 199 faces are caucasian, one female face is asian. The crucial step is to establish dense point-to-point correspondence of all faces with a reference face. Then, shapes and textures of all m individual faces $i \in \{1, ..., m\}$ in the database are represented by shape and texture vectors [3]

$$\mathbf{S}_i = (x_1, y_1, z_1, x_2, \ldots, x_n, y_n, z_n)^T \quad (1)$$
$$\mathbf{T}_i = (r_1, g_1, b_1, r_2, \ldots, r_n, g_n, b_n)^T \quad (2)$$

formed by concatenating coordinates and colors of all n vertices of the reference model. So a given point such as the tip of the nose will be represented by the same vertex for all individual faces. In our model $m = 200$ and $n = 75,972$. Furthermore, arithmetic means are calculated to have zero-mean shape and zero-mean texture vectors: $\mathbf{S}_i = \overline{\mathbf{s}} + \mathbf{s_i}$ and $\mathbf{T}_i = \overline{\mathbf{t}} + \mathbf{t_i}$ with the arithmetic mean $\overline{\mathbf{s}}$ and $\overline{\mathbf{t}}$ and the zero-mean shape vector $\mathbf{s_i}$ and texture vector $\mathbf{t_i}$.

In this face-space representation, linear combinations of shape and texture vectors generate morphs of the database faces, and will therefore have plausible and natural face-like appearance:

$$\mathbf{S} = \overline{\mathbf{s}} + \sum_{i=1}^{m} \alpha_i \mathbf{s_i} \quad (3)$$

$$\mathbf{T} = \overline{\mathbf{t}} + \sum_{i=1}^{m} \beta_i \mathbf{t_i} \quad (4)$$

B. Attribute Vector

The attribute vector [7] is an easy to handle method for manipulating the appearance of faces in one specific/ defined direction. Thus, it is possible to change only one facial characteristic, such as the overall shape of a face, and retain all other characteristics, such as the shape of the mouth or the eyes, entirely.

Since attribute vectors are defined in the same face space representation as the shape and texture vectors, both number of vertices and (more important) the dense point-to-point correspondence of all vertices is maintained for each attribute vector. Due to this property, an addition or subtraction of attribute vectors to shape or texture vectors are possible. Hence the manipulation process is implemented by adding or subtracting multiples of an attribute vector to a shape or texture vector:

$$\mathbf{s_{changed}} = \mathbf{s_i} + \mathrm{d} \cdot \mathbf{a_{s,k}}. \quad (5)$$

Here $\mathbf{s_i}$ is a shape vector and $\mathbf{a_{s,k}}$ is an attribute vector for shape describing one specific attribute k. d expresses how strong the characteristic should change. An example of how the attribute vector manipulates the appearance of a face is shown in Figure 1 For this instance, let $\mathbf{a_{s,cheek}}$ be an attribute vector for shape, describing whether the shape of the cheeks are skinny or puffy. Adding or subtracting multiples of $\mathbf{a_{s,cheek}}$ to any shape vector alters the facial shape regarding this attribute, but keeps all other characteristics such as the shape of the mouth or the eyes unchanged (see first row in Figure 1). The second row in Figure 1 shows how the overall shape of an input face can be altered by utilizing an attribute vector. Although this example shows only changes concerning the 3D facial shape, the attribute vector is also applicable to facial texture. For the generation of such attribute vectors, two processing steps are necessary.

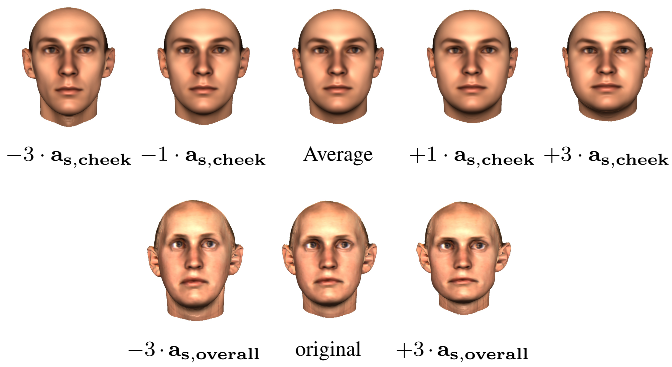


Fig. 1: Example of two different attribute vectors: The first row shows the manipulated average face by adding or subtracting the attribute vector $\mathbf{a_{s,cheek}}$ that describes the shape of the cheeks. Subtracting multiples of this vector results in skinny cheeks whereas adding multiples of the vector result in puffy cheeks. The second row shows the manipulation of an input face with the attribute vector $\mathbf{a_{s,overall}}$ that describes the overall facial shape. In this example adding the attribute vector alters the facial shape to a round facial shape and subtracting to a rectangular facial shape. Note that all other facial characteristics are not changed by the attribute vector.

First, let $\mathbf{s_i}$ be a set of shape sample vectors, and $b_i \in \mathbb{R}$ be the ratings of a given attribute for these 3D faces. The ratings b_i can be either given as ground truth (this could be the age or any other measurable attribute such as the nose length or the width of the mouth) or a subjective rating selected by the user [7]. In a second step, the attribute vector $\mathbf{a_s}$ is computed by solving an optimization problem, that can be converted to a simple weighted sum:

$$\mathbf{a_s} = \frac{1}{m} S\mathbf{b} = \frac{1}{m} \sum_{i=1}^{m} b_i \mathbf{s_i} \quad (6)$$

with the data matrix $S = (\mathbf{s_1} \ldots \mathbf{s_m})$ and $\mathbf{b} = (b_1 \cdots b_n)$. To calculate an attribute vector $\mathbf{a_t}$ for texture, the same can be done with a texture data matrix $T = (\mathbf{t_1} \cdots \mathbf{t_n})$. For this paper, the set of shape and texture vectors is the whole database of 200 3D laser scans from the 3DMM (see Section I-A).

II. Correlation Estimation

A. Attribute Mapping Function

The main goal of this paper is the exploration of correlation in facial data. More precisely we want to figure out if statistical relations between different facial parts (e.g. between eyes and mouth, upper and lower part of the face) or between different modalities (e.g. between RGB color information and 3D shape) exist. So is it possible to draw conclusions from the shape of the mouth to the shape of the eye or from facial color to shape of facial parts or the general shape and vice versa. We want to explore if worded statements like “male people with small eyes have probably an overall rectangular facial shape” or “people with fair skin will probably have narrower lips than people with darker skin” can be formulated automatically from a statistical analysis.

To solve this task, a description method to measure global facial features such as overall shape and even more any partial characteristics of faces like the specific shape of nose, eyes or cheeks, is necessary first. Furthermore, this measurement should map the strength of each characteristic to a single value, to have a descriptive tool for comparison different input faces concerning the intensityof the related characteristic. The following notation refers to shape first, but applies to texture in the same way. For this, let $f(\mathbf{s_i}) = l_i$ be an *attribute mapping function* that maps the shape vector $\mathbf{s_i} \in \mathbb{R}^n$ of an input face i to a single value $l_i \in \mathbb{R}$ that rates the facial shape regarding a defined facial characteristic. The attribute mapping function $f(\mathbf{s})$ should be applicable to simple characteristics such the width of the nose (which can be measured by a trivial distance calculation between two vertices) as well as to more complex characteristics such as the specific shape of facial parts like the cheeks or the eyes (which requires more complex calculations for mapping to a single value). Due to the small number of input heads ($m = 200$), we restrict the mapping to a linear function f that can be implemented as a scalar product. Therefore, the attribute vector concept described in Section I-B is used, since it handles the demanded constraints entirely. In this paper we use the attribute vector $\mathbf{a_{s,k}}$ for rating the strength of a facial characteristic by projecting a shape vector $\mathbf{s_i}$ onto $\mathbf{a_{s,k}}$. This projection is the scalar product of $\mathbf{a_{s,k}}$ and $\mathbf{s_i}$. So the attribute mapping function $f(\mathbf{s_i})$ for a specific attribute vector can be written as

$$f(\mathbf{s_i}, \mathbf{a_{s,k}}) =< \mathbf{s_i}, \mathbf{a_{s,k}} >= l_i. \tag{7}$$

The value l_i expresses the strength of the characteristic defined by $\mathbf{a_{s,k}}$ for an input face represented by a shape vector $\mathbf{s_i}$. For example, let the addition of multiples of $\mathbf{a_{s,k}}$ to a shape vector modifies the overall shape towards an angular shape and the subtraction towards a round facial shape. Then values of l_i greater than zero denotes an angular face and values less than zero a round overall shape. Moreover, the scale is continuous, so it is possible to rate and compare different strength of angularity or roundness.

By concatenating m shape vectors to a matrix, it is possible to rate several input faces (simultaneously) for one facial attribute:

$$f(S, \mathbf{a_{s,k}}) = S^T \mathbf{a_{s,k}} = \mathbf{l_{s,k}} \tag{8}$$

with

$$S = (\mathbf{s_1} \dots \mathbf{s_m}) \quad \text{and} \quad \mathbf{l_{s,k}} = (l_{1,k} \dots l_{m,k})^T \tag{9}$$

Here, S is a *shape matrix* with m shape vectors as column vectors. The *label vector* $\mathbf{l_{s,k}}$ represents the strength of the shape characteristic k for each of the m shape vectors in S.

If two different attribute vectors $\mathbf{a_{s,1}}$ and $\mathbf{a_{s,2}}$ are projected onto the same shape matrix S, the relation of elements in $\mathbf{l_{s,1}}$ and $\mathbf{l_{s,2}}$ is consistent in the following sense: the first entry in both label vectors is related to the first input shape vector, the second entry in both label vectors to the second input shape vector and so on. This property is crucial for further calculation of facial correlation in Section II-B.

B. Exploring facial Correlation between Shape and Texture

In this section we focus on facial shape and RGB information to illustrate the method for exploring facial correlations between these two modalities. We utilize the attribute mapping function and the corresponding label vector. But unlike the previous sections, where a predefined attribute vector is used, we go the other way around by estimating an unknown attribute vector, that describes the correlations between shape and texture.

Let $S = (\mathbf{s_1}, \dots, \mathbf{s_m})$ be a shape matrix with $m = 200$ zero-mean shape vectors (as columns) and let $T = (\mathbf{t_1}, \dots, \mathbf{t_m})$ be a texture matrix with the corresponding zero-mean texture vectors. Then $f(S, \mathbf{a_{s,1}})$ calculates the label vector $\mathbf{l_{s,1}}$ that rates every input shape (vector $\mathbf{s_i}$) regarding an unknown facial shape characteristic described by attribute vector $\mathbf{a_{s,1}}$, and $f(T, \mathbf{a_{t,1}})$ calculates the label vector $\mathbf{l_{t,1}}$ that rates every input texture (vector $\mathbf{t_i}$) regarding an unknown texture characteristic represented by attribute vector $\mathbf{a_{t,1}}$. The index 1 for both attribute vectors and the label vector denotes that those vectors describe the direction with the largest correlation. Since the position of shape and texture vectors are consistent in T and S (shape vector s_i and texture vector t_i of face i are at the same position in S respectively T), the relation of all entries in both label vectors are also consistent.

$$\begin{pmatrix} l_{s,1,1} & \leftrightarrow & l_{t,1,1} \\ l_{s,1,2} & \leftrightarrow & l_{t,1,2} \\ \vdots & & \vdots \\ l_{s,1,m} & \leftrightarrow & l_{t,1,m} \end{pmatrix} = \begin{pmatrix} f(\mathbf{s_1}, \mathbf{a_{s,1}}) & \leftrightarrow & f(\mathbf{t_1}, \mathbf{a_{t,1}}) \\ f(\mathbf{s_2}, \mathbf{a_{s,1}}) & \leftrightarrow & f(\mathbf{t_2}, \mathbf{a_{t,1}}) \\ \vdots & & \vdots \\ f(\mathbf{s_m}, \mathbf{a_{s,1}}) & \leftrightarrow & f(\mathbf{t_m}, \mathbf{a_{t,1}}) \end{pmatrix}$$

Now the goal is finding those two attribute vectors $\mathbf{a_{s,1}}$ and $\mathbf{a_{t,1}}$ that minimizes the angle θ between the corresponding label vectors $l_{s,1}$ and $l_{t,1}$ (in an optimal solution the difference between $l_{s,1}$ and $l_{t,1}$ would be zero). This leads to a maximization of

$$\frac{< S^T \mathbf{a_{s,1}}, T^T \mathbf{a_{t,1}} >}{\|S^T \mathbf{a_{s,1}}\| \|T^T \mathbf{a_{t,1}}\|}. \tag{10}$$

With zero-mean shape and texture vectors (see Section I-A) it is similar to maximizing the Pearson correlation coefficient and Equation (10) can be written as

$$r_1 = corr(S^T \mathbf{a_{s,1}}, T^T \mathbf{a_{t,1}}), \tag{11}$$

where r_1 is the correlation coefficient and $\mathbf{a_{s,1}}$ and $\mathbf{a_{t,1}}$ are the attribute vectors with the largest correlation.

This maximization problem can be solved by using the canonical correlation analysis.

C. Canonical Correlation Analysis (CCA)

The problem formulation of CCA is similar to Equation (10) when S and T are considered as two random variables, and $l_s = S^T \mathbf{a_s}$, $l_t = T^T \mathbf{a_t}$ as a linear combination of basis vectors $\mathbf{a_s}$ and $\mathbf{a_t}$[8]. Then Equation (10) to be maximized can be written as

$$\begin{aligned} r &= \frac{E[\mathbf{a_{s,1}}^T S T^T \mathbf{a_{t,1}}]}{\sqrt{E[\mathbf{a_{s,1}}^T S S^T \mathbf{a_{s,1}}] E[\mathbf{a_{t,1}}^T T T^T \mathbf{a_{t,1}}]}} \\ &= \frac{\mathbf{a_{s,1}}^T C_{st} \mathbf{a_{t,1}}}{\sqrt{\mathbf{a_{s,1}}^T C_{ss} \mathbf{a_{s,1}} \mathbf{a_{t,1}} C_{tt} \mathbf{a_{t,1}}}} \end{aligned} \tag{12}$$

where C_{ss} and C_{tt} are the covariance matrices of $\mathbf{a_s}$ and $\mathbf{a_t}$, and $C_{st} = C_{ts}^T$ are the cross-covariance matrices, respectively. Canonical correlation analysis finds two sets of basis vectors, such that the correlation between the projections of the random variables onto these basis vectors are maximized [8]. With the total covariance matrix

$$C = \begin{pmatrix} C_{ss} & C_{st} \\ C_{ts} & C_{tt} \end{pmatrix} \tag{13}$$

the canonical correlations between S and T are calculated by solving the eigenvalue equations

$$\begin{cases} C_{ss}^{-1} C_{st} C_{tt}^{-1} C_{ts} \mathbf{a}_s = r^2 \mathbf{a}_s \\ C_{tt}^{-1} C_{ts} C_{ss}^{-1} C_{st} \mathbf{a}_t = r^2 \mathbf{a}_t. \end{cases} \tag{14}$$

r^2 are the squared correlation coefficients, and $\mathbf{a}_s$ and $\mathbf{a}_t$ are the normalized attribute vectors (the normalized CCA basis vectors). The number of non-zero solutions is limited to the smallest dimensionality of S and T. The solutions are sorted in descending order concerning the correlation coefficient. Thus, the attribute vectors $\mathbf{a_{s,1}}$ and $\mathbf{a_{t,1}}$ of the first solution describe the correlation with the largest correlation coefficient r_1, the attribute vectors $\mathbf{a_{s,2}}$ and $\mathbf{a_{t,2}}$ of the second solution describe the correlation of the second largest correlation, and so on. The solutions of the eigenvalue problems are related, so only one equation has to be solved

$$\begin{cases} C_{st} \mathbf{a}_t = r \lambda_s C_{ss} \mathbf{a}_s \\ C_{ts} \mathbf{a}_t = r \lambda_t C_{tt} \mathbf{a}_t \end{cases} \tag{15}$$

with

$$\lambda_s = \lambda_t^{-1} = \sqrt{\frac{\mathbf{a_t}^T C_{tt} \mathbf{a_t}}{\mathbf{a_s}^T C_{ss} \mathbf{a_s}}} \tag{16}$$

Due to the high dimensionality of the attribute vectors ($3 \cdot n \cdot$ with $n = 75,972$) in relation to the number of input heads ($m = 200$), the *small sample size* (SSS) problem occurs [10], [17]. In this case, the CCA is unfeasible, since it always finds a solution that results in maximum correlation with a correlation coefficient of $r_k = 1$.

By using a principal component analysis (PCA) the SSS problem can be avoided [17]. For this, the distribution of database faces (see Section I-A) can be described in terms of arithmetic means, unit-length eigenvectors and standard deviations for shape and texture. With this, we can rewrite Equation (3) and (4) in a new basis (see [3]):

$$\mathbf{S} = \bar{\mathbf{s}} + \sum_{i=1}^{m-1} \alpha_i \mathbf{u_{s,i}}, \tag{17}$$

$$\mathbf{T} = \bar{\mathbf{t}} + \sum_{i=1}^{m-1} \beta_i \mathbf{u_{t,i}}. \tag{18}$$

As the attribute vectors are defined in the same face space representation as the shape and texture vectors, they can also be represented by a linear combination of the principal components $\mathbf{u_{s,i}}$ and $\mathbf{u_{t,i}}$:

$$\mathbf{a_{s,k}} = \sum_{i=1}^{m-1} \alpha_{k,i} \mathbf{u_{s,i}} = U_s \boldsymbol{\alpha_k}, \tag{19}$$

$$\mathbf{a_{t,k}} = \sum_{i=1}^{m-1} \beta_{k,i} \mathbf{u_{t,i}} = U_t \boldsymbol{\beta_k}, \tag{20}$$

with $U_s = (\mathbf{u_{s,1}} \ldots \mathbf{u_{s,m-1}})$. and $U_t = (\mathbf{u_{t,1}} \ldots \mathbf{u_{t,m-1}})$ Equation (11) can be written as

$$\begin{aligned} r_1 &= corr(S^T U_s \boldsymbol{\alpha}_1, T^T U_t \boldsymbol{\beta}_1) && (21) \\ &= corr(S_\alpha \boldsymbol{\alpha}_1, T_\beta \boldsymbol{\beta}_1) && (22) \end{aligned}$$

with the data matrices $S_\alpha = S^T U_s$ and $T_\beta = T^T U_t$. Now the attribute vector $\mathbf{a_{s,k}}$ for shape is represented by the coefficient vector $\boldsymbol{\alpha_k}$ and the attribute vector $\mathbf{a_{t,k}}$ for texture by the coefficient vector $\boldsymbol{\beta_k}$. This reduces the complexity for the CCA calculation as well, since the coefficient vectors, with a maximum size of $m - 1 = 199$, are much smaller than the original attribute vectors. However, the size of the coefficients vectors are de facto much smaller than 199. The optimal number of components is evaluated in the next section.

D. Correlation Validation

As pointed out in Section II-C, the dimensionality reduction due to principal component analysis avoids the small sample size problem and makes it possible to solve the canonical correlation analysis problem numerically. In the following, we demonstrate that the number of principal components has to be reduced further since using all $m = 199$ principal components would lead to correlation coefficients equal to 1 for all solutions even on random data.

In our evaluation, the correlations between facial shape and texture were calculated by CCA with different numbers of PCA dimensions. We started the estimation with 5 PCA components and increased the number in steps of 5. The largest correlation coefficient r_1 for each number of components is plotted with the blue line in Figure 2.

This graph shows that using more than 80 principal components leads to correlation coefficients close to 1. However, we cannot be sure that the estimated attribute vectors describe real and informative correlations, as opposed to random ones. In large datasets, it will always be possible to find solutions with a large correlation coefficient, even if these correlations describe random effects. To eliminate this, we permuted the order of the data vectors in one of the input data matrices. In our case, the order of shape vectors $\mathbf{s_i}$ in the shape matrix S were altered and the order of the texture vectors $\mathbf{t_i}$ in the texture matrix T are left unchanged. Note that it is not the order of vertices in the shape and texture vectors, but the assignment of vectors to individual faces that we altered, so the shape vector s_i of sample face i is no longer mapped to the correct texture vector t_i anymore. Afterwards the correlations between the modified shape and the unchanged texture matrices are recalculated with CCA. Again, we implemented the estimation with different numbers of PCA dimensions. The result is shown in the red line of Figure 2.

The result shows that the correlation coefficient of the permuted data is lower than the coefficient of the unchanged data (see blue line in Figure 2). It also shows that in higher dimensions (higher number of PC components used), the two curves converge: If more than 85 components are used, the correlation coefficient r_1 of the largest correlation becomes 1, even for trivial random datasets.

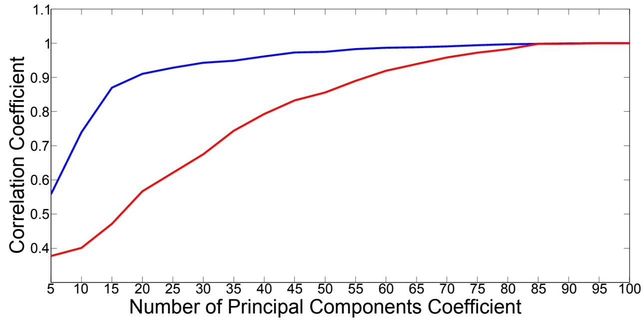


Fig. 2: Influence of spurious, random correlations in the dataset: The blue curve shows the largest correlation coefficient r_1, depending on the dimensionality of input vectors (number of principal components) between shape and texture, based on the fact that for each individual face i we have a shape and a texture vector. By a permutation of the order of $\mathbf{t}_i$, we destroy the mapping between individual shapes and textures. The red curve shows that for high-dimensional data vectors, CCA still finds (meaningless) correlations.

From these findings, we draw the following conclusion: The estimated facial correlations are non-random correlations if the proper number of coefficients is chosen, since the difference between the two curves is substantial in the range of 15 to 40 components. If not stated otherwise, we use 35 number of principal components in this paper as tradeoff between the magnitude of the correlation coefficient (blue line in Figure 2) and distance to the random correlation coefficient (red line). Note that an analytical criterion for the statistical significance of correlations, along the lines of a t-Test, would be difficult in this case for two reasons: first, we have very high dimensional data and a relatively low sample size (200 faces), and second, it is inherently difficult to separate signal from noise in this type of data. We are not referring to spatial noise on the surface of the face, but the randomness of facial features in the ensemble of human faces. Even if it was feasible to apply methods that attempt to separate "true" from "random" sources of variations to our problem, such as Probabilistic PCA [18], these methods would make strong assumptions. In contrast, our Monte-Carlo analysis with random permutations provides a valid and reliable test of the importance of correlations in paired data vectors.

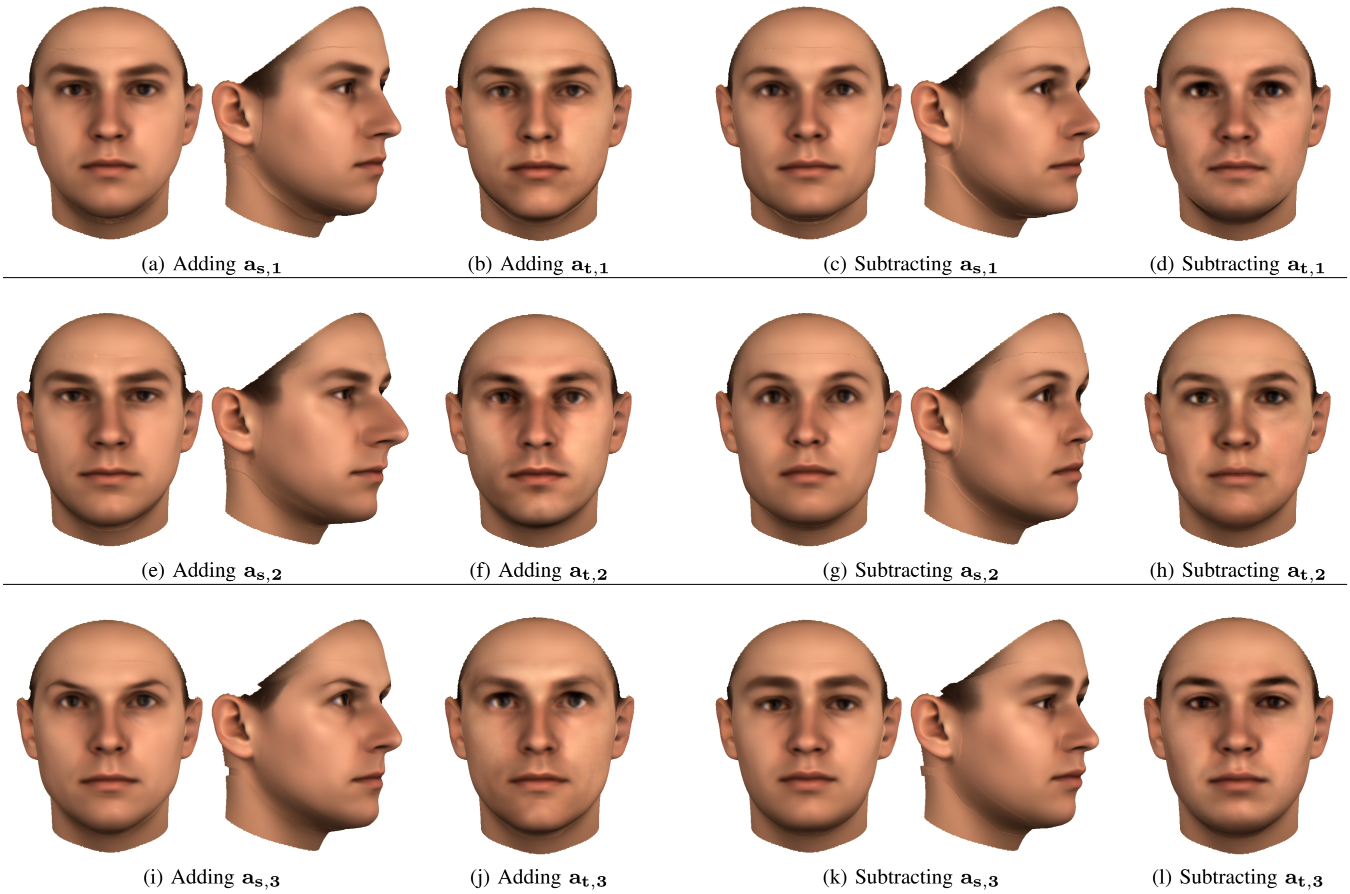
III. Visualization of Correlations

The combination of the 3D Morphable model of faces (see Section I-A) and the CCA (Section II) makes it easy to explore the correlations in faces. As described in Section II-C, CCA calculates pairs of basis vectors, such that the correlation between the projections of the input data onto these basis vectors are maximized (see Equation 14). Since the number of non-zero solutions is limited to the dimensionality of the input data S_α and T_β, and we are using 35 principal components (Section II-D), CCA calculates 35 pairs of basis vectors $\boldsymbol{\alpha_k}$ and $\boldsymbol{\beta_k}$ with $k \in 1, ..., 35$. These basis vectors are the PCA coefficient vectors for shape and texture and can be interpreted as attribute vectors by using Equation (19) for shape and Equation (20) for texture.

Sorted in a descending order with respect to the correlation coefficients, the first pair of attribute vectors $\mathbf{a_{s,1}}$ and $\mathbf{a_{t,1}}$ visualizes the largest correlation between shape and texture, and the second pair $\mathbf{a_{s,2}}$ and $\mathbf{a_{t,2}}$ with $r_2 < r_1$ the second largest correlation. Due to this representation, we can use the attribute vector concept for visual inspection of correlations by adding multiples of these vectors to any face (as described in Section I-B). It is important to keep in mind that the pairs of attribute vectors $\mathbf{a_{s,k}}$ and $\mathbf{a_{t,k}}$ are related, so we show them as a pairwise manipulation side by side in Figure 3. Note, that the attribute vectors of one modality ($\mathbf{a_{s,k}}$, $k = 1, 2, ...$) are not pairwise orthogonal, since CCA enforces a more indirect criterion of independence of components. Figure 3 shows the first three pairs of attribute vectors with the largest correlation coefficient in three rows. The vectors $\mathbf{a_{s,k}}$ with $k \in 1, 2, 3$ are added (a) or subtracted (c) to the average face shape while texture remains unchanged, and in separate images (b, d), the related attribute vectors for texture ($\mathbf{a_{t,k}}$ with $k \in 1, 2, 3$) are added (b) to or subtracted (d) from the average face texture (here, the shape is not modified). Note that the relative sign of $\mathbf{a_{s,k}}$ and $\mathbf{a_{t,k}}$ is important here, unlike the signs of principal components in standard PCA, so (a) and (b) form a pair of attributes, and (c) and (d) are the opposite pair. We used 35 principal components for the correlation estimation and the value of the 3 largest correlation coefficients are: $r_1 = 0.9487$, $r_2 = 0.9322$ and $r_3 = 0.9202$. In the calculations, we consider only vertices of the inner part of the face, and ignore areas such as the neck, the forehead and the ears. In the following figures, these areas are rendered with the average facial shape and texture.

Figure 3 indicates that in terms of facial shape, the shape of the nose, the eyebrows and the eyes and the thickness of the lips are correlated with the color or brightness of the eyelashes and the lips, and a beard shadow. More precisely, subtracting $\mathbf{a_{s,2}}$ makes the shape of the nose smaller and finer, as well as the the eyebrows thinner and more curved. Also the eyes are more circular and the lips thicker. Regarding the related attribute vector for texture $\mathbf{a_{s,2}}$, subtraction reduces the beard shadow, darkens the color of the eyelashes and makes the color of the eyebrows more continuous. Also the color of the lips is more pale, which is perhaps one of the more unexpected correlations that we found. Some of the correlations can be explained by the typical differences between male and female faces that were found previously in analyses of the differences of male and female 3D scans [3].

With the method described in this paper, correlations between any modality or sub-region of human face scans can be investigated in the same way as we described for shape and texture. For example, we calculated the correlations between the shape of facial front and side information: In this case, the input matrix for the frontal information is formed only by the x (left-right) and y (vertical) coordinates of the shape vectors s_i (see Section I-A), and the second input matrix for the side information only with the z (depth) coordinates. All other subsequent calculations are the same as the estimation of shape and texture correlation.

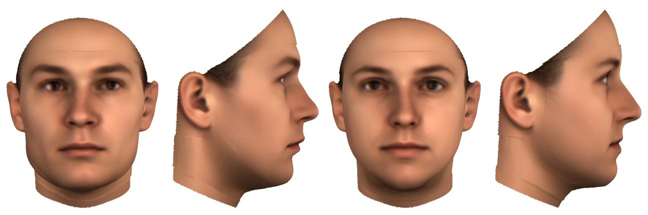


(a) Adding $\mathbf{a_{s,1}}$ (b) Adding $\mathbf{a_{t,1}}$ (c) Subtracting $\mathbf{a_{s,1}}$ (d) Subtracting $\mathbf{a_{t,1}}$

(e) Adding $\mathbf{a_{s,2}}$ (f) Adding $\mathbf{a_{t,2}}$ (g) Subtracting $\mathbf{a_{s,2}}$ (h) Subtracting $\mathbf{a_{t,2}}$

(i) Adding $\mathbf{a_{s,3}}$ (j) Adding $\mathbf{a_{t,3}}$ (k) Subtracting $\mathbf{a_{s,3}}$ (l) Subtracting $\mathbf{a_{t,3}}$

Fig. 3: Visualization of correlations between facial shape and texture information: Due to addition and subtraction of the calculated attribute vectors for shape $\mathbf{a_{s,k}}$ and for texture $\mathbf{a_{t,k}}$ to a face, it is possible to visualize the estimated correlation. Here, the 3 pairs of attribute vectors with the largest correlation coefficient are used, to illustrate this mechanism by applying this on the average facial shape and texture. (3a) and (3c) shows this for the first attribute vector for shape $\mathbf{a_{s,1}}$, and (3b) respectively (3d) the related attribute vector for texture $\mathbf{a_{t,1}}$. (3e), (3g), (3f), (3h) illustrates the second largest correlation and (3i), (3k), (3j), (3l) the third largest correlation.



(a) + $\mathbf{a_{front,1}}$ (b) + $\mathbf{a_{side,1}}$ (c) - $\mathbf{a_{front,1}}$ (d) - $\mathbf{a_{side,1}}$

Fig. 4: Visualization of correlations between frontal and side information: The first and the third example shows the addition or subtraction of the attribute vector $\mathbf{a_{front,1}}$ with frontal information (x (left-right) and y (vertical) coordinates) and the second and fourth example shows the attribute vector $\mathbf{a_{side,1}}$ with side information (z (depth) coordinates). All examples are rendered with average texture.

Figure 4 shows another example of facial correlations we have evaluated. It visualizes the correlation between facial front (x and y) and side (z) coordinates. For this example, we considered only the shape of the inner parts of the face again, and ignored the neck, the forehead, the ears and the texture. The figures show the average facial shape and texture in these areas.

A. CCA Attributes Mapped to Semantically Meaningful Characteristics

The visualization of correlations (Section III) has shown that the estimated pairs of attribute vectors do not describe only one specific facial characteristic, but rather several combinations of different characteristics. In order to explore which facial characteristics are in the correlated attributes and to obtain verbal descriptions, we propose a method for automated exploration. This is achieved by projecting the estimated attribute vectors (for example $\mathbf{a_{s,k}}$ and $\mathbf{a_{t,k}}$ for correlation between shape and texture) to predefined attribute vectors that describe only one semantically meaningful facial characteristic each. 50 such attribute vectors are generated with the method from Section I-B and manual labelling of the database faces with respect to overall shape of the face or the cheeks, the shape of the mouth, the eyes or the eyebrows or one specific texture characteristics, such as the brightness of the eyes, the lips or the eyebrows.

	α_1	α_2	α_3
narrow/wide eyes	0.1094	-0.0923	**0.348**
convex/concave nose	0.0469	**0.3198**	0.0952
male/female	-0.0719	**-0.2552**	0.0572
round/angular	**-0.2124**	-0.0738	0.1488
length of nosebridge	0.0441	0.0116	**-0.3363**
straight/curved eyebrows	-0.0712	-0.1991	**0.4021**
thin/thick of eyebrows	0.2828	**0.2403**	**-0.379**

Fig. 5: Comparison between the shape coefficients $\boldsymbol{\alpha_k}$ of the calculated attribute vectors $\mathbf{a_{s,k}}$ and the coefficients of the predefined attributes.

	β_1	β_2	β_3
dark/bright eyes	0.162	0.0937	**0.3593**
beard shadow	-0.1374	**0.2143**	0.0798
light/dark eyebrows	**0.3364**	0.3046	**-0.6342**
male/female	-0.0012	**-0.3391**	0.0361

Fig. 6: Comparison between the texture coefficients $\boldsymbol{\beta_k}$ of the calculated attribute vectors $\mathbf{a_{t,k}}$ and the coefficients of the predefined attributes.

Now a comparison of the predefined vectors with the estimated basis vectors calculated by CCA is possible. Since the correlations are calculated with Equation (22), the estimated pairs of attribute vectors are already represented by the face-space coefficients $\boldsymbol{\alpha_k}$ and $\boldsymbol{\beta_k}$. The predefined attribute vectors can be converted into this representation as well by projecting each of the 35 principal component $\mathbf{u_{s,i}}$ or $\mathbf{u_{t,i}}$ onto each vector. To compare the calculated with the predefined attribute vectors, the scalar product between the coefficients serves as the rating criterion. So, for the estimated shape attribute vector $\mathbf{a_{s,k}}$ and a predefined attribute vector $\mathbf{a_{s,eyes}}$ that describes the shape of the eyes,

$$rating_{k,eye} = \frac{< \boldsymbol{\alpha_k}, \boldsymbol{\alpha_{eye}} >}{\|\boldsymbol{\alpha_k}\| \|\boldsymbol{\alpha_{eye}}\|}, \tag{23}$$

with $\boldsymbol{\alpha_k}$ as a coefficient vector for $\mathbf{a_{s,k}}$ and $\boldsymbol{\alpha_{eye}}$ as the coefficient vector for $\mathbf{a_{s,eye}}$.

With this method we evaluated the correlations between facial shape and texture as well as several other combinations (e.g. between frontal and side information, between eyes and mouth). In the following, we take a closer look at the correlation between shape and texture, using the set of 50 predefined attribute vectors.

The analysis is restricted to the most reliable non-random correlations according to the Monte-Carlo simulation II-D, so we use only the sets of attribute vectors with a correlation coefficient r_k greater than the largest correlation coefficient of the permuted datasets. In case of correlations between shape and texture (using 35 principal components), 11 pairs of attribute vectors ($\mathbf{a_{s,k}}$ and $\mathbf{a_{t,k}}$ with $k \in 1, .., 11$) are used, since $r_1 1 = 0.7717$ is the last correlation coefficient greater than the highest correlation coefficient $r_{1,permuted} = 0.7438$ of the permuted input data.

B. Results of CCA Projection

To illustrate this method, Figure 5 and 6 lists the ratings for the exploration of correlations between shape and texture. In this table, only a selection of the most informative predefined attribute vectors are listed, as well only the pairs of attribute vectors with the 3 largest correlations (for a visualization of these 3 vectors pairs see Figure 3).

The sign of the rating values has to be treated like the labels l_i used in the attribute mapping function (see Equation (7 in Section II-A). So, consider $\mathbf{a_{s,round/angular}}$ as an defined attribute vector, so that the addition modifies the overall shape towards an angular shape and the subtraction towards a round facial shape and let $\mathbf{a_{s,1}}$ be the estimated attribute vector for shape with the highest correlation. Than a negative rating value denotes that the correlation of a round overall shape is described within the first pairs of attribute vectors.

Figure 5 and 6 shows that the automatic exploration of the correlation between shape and texture is consistent with the renderings (see Section III) and further illustrates more relations. Concerning the second pair of estimated attributes ($\boldsymbol{\alpha_2}$ for shape and $\boldsymbol{\beta_2}$ for texture, second columns in Figure 5 and 6) the highest ratings are consistent with the visual appearance. The method indicates a correlation between gender and the shape of the nose. A beard shadow seems to be related with a concave nose, as well as thick eyebrows. It also shows that these findings are related to a male appearance. Also a slop chin, with a convex nose thin eyebrows occur more often among females apparently.

The correlation of the third pair of attribute vectors (($\boldsymbol{\alpha_3}$ for shape and $\boldsymbol{\beta_3}$ for texture in Figure 5 and 6) suggests that the width of the eyes is related to the brightness of the eyebrows, as well as to the brightness of the eyes. That is also consistent with the visualization in Figure 3.

IV. CCA Prediction of Occluded Areas

If predictions from visible to invisible structures of faces are based on correlation, and CCA with our Monte-Carlo simulation helps to identify reliable correlations and avoid random ones, we would expect that our CCA components provide better predictions than standard PCA, as used for example by [16].

In our experiments, we considered several correlations between facial parts: between frontal and side, between the entire face and eyes, between the entire face and mouth, between upper and lower part, shape and texture. We calculated a PCA for each facial part and used CCA to find pairs of correlated attribute vectors in the subspaces spanned by 35 principal components. Similar to the inference technique described in [16], we used *multivariate linear regression* (MLR [12]) to infer from one facial part to the other. For example, the linear coefficients of the upper part (with respect to its PCA basis vectors) can be predicted from the linear coefficients of the lower part.

To compare the PCA with the CCA, our approach was to build linear combinations of the estimated attribute vectors to generate new faces. We used only the sets of attribute vectors with a greater correlation coefficient than the largest correlation coefficient of the permuted datasets (see Section

III), which in our case is in the range between 8 and 13 pairs. Since the attribute vectors are not pairwise orthogonal, a linear combination of attribute vectors is not possible directly. Instead, the attribute mapping function is applied to calculate the labels for all pairs of face parts. Then, MLR is trained to map the labels from one part to those of the other. For prediction of unknown data, the mapping gives us the labels for the invisible part, and a pseudoinverse calculation defines the coefficients of the non-orthogonal attribute vectors that reproduce these target labels up to a minimal least-squares error.

Overall, we ran 12 experiments with different tasks. In one experimental setup the ground truth of the whole face and the whole face with the prediction of the lower part by CCA and by the PCA-based method was shown. The task was to rate which of the predictions are closer to the ground truth. In another setup only two images are shown. For example the upper part of the face completed with the prediction of the lower part by CCA and by the PCA-based method. Now the task was to rate which prediction is more plausible, without knowing the ground truth.

These two tasks were run with prediction of the eye region (including eyes and eyebrows) and mouth region from the remaining face regions, as well front view to side view. We also modified the experiments by adding a third PCA-based prediction method LinVert as in [16].

In none of our pilot studies (3 to 4 participants each, 200 trials) did we find any trends towards preferences to any of the prediction methods. Since it is difficult to establish the absence of an effect experimentally, further measurements with more participants made little sense. Still, we conclude at this point that CCA is unlikely to be superior to the PCA-based methods in this setting. On the positive side, PCA seems to capture correlations sufficiently and is not affected by spurious random correlations in the limited training set.

V. Discussion

The results presented in this paper shed new light on the chances and limitations of inferences from visible to invisible structures in faces, both by the human visual system and by computer graphics or vision. We identified the most highly correlated dimensions in the face space of shapes and the space of textures, respectively, or in vector spaces built from disjoint facial regions. To the best of our knowledge, we were the first to apply CCA to human face data. Our Monte-Carlo simulation (Section II-D) helped to eliminate random correlations and find the true ones in our dataset by reducing the CCA problem to an appropriate, lower dimensional subspace.

We had expected a substantial improvement of the predictive power of our model based on CCA, as opposed to PCA, and we had hoped to be able to verify this in an experiment that compares our computational predictions with the expectations of human observers. It is slightly disappointing, yet not less instructive and worth reporting, to find that no improvement was found: Even though simple, PCA-based prediction tends to rely both on true and on spurious correlations, the result looks just as similar to the ground truth, and just as plausible to human observers. This corroborates the findings of [16] that indicated that human expectation is in line with a PCA based prediction in the case of guessing profiles from front views of faces.

Given random pairs of high-dimensional sample vectors $(\mathbf{s}_i, \mathbf{t}_i)$, you will always find directions that are highly correlated. This is what the Monte-Carlo simulation (Section II-D) quantified. However, it is unlikely to obtain *the same randomly* correlated directions in different training sets, so we would not conclude that the human visual system is more like PCA than like CCA. Instead, the differences between PCA-based prediction, CCA-based prediction, human expectation and ground truth seem to be equally far in different directions and within the range of residual unpredictiveness of faces.

References

[1] S. Baker and T. Kanade, "Hallucinating faces", *in Int. Conf. on Automatic Face and Gesture Recognition*, 1999.

[2] Y.Liang, J.H. Lai, W.S. Zheng and Z. Cai, " A survey of face hallucination", *in 7th Chinese Conference, CCBR*, 2012, pp 83-93.

[3] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces", *in SIGGRAPH*, 1999, pp 187-194.

[4] S. Becker, Mutual information maximization: models of cortical self-organization, *Network: Computation in Neural Systems*, vol. 7, 1996, pp 7-31.

[5] A.M. Burton, V. Bruce and R.A. Johnston, Understanding face recognition with an interactive activation model, *British Journal of Psychology*, vol. 81, 1990, pp 361-380.

[6] A.M. Burton and V. Bruce, Naming faces and naming names: Exploring an interactive activation model of person recognition, *Memory*, vol.1, 1993, pp 457-480.

[7] V. Blanz, I. Albrecht, J. Haber, and H.-P. Seidel, Creating face models from vague mental images, *Computer Graphics Forum*, vol. 25, 2006, pp 645-654.

[8] M. Borga, Learning Multidimensional Signal Processing, Linkping Studies in Science and Technology. Dissertations No. 531, Linkping University, Sweden; 1998.

[9] J. Hasbrouck and D. Seppi, Common factors in prices, order flows, and liquidity, *Journal of Financial Economics*, vol. 59, 2001, pp 383-411.

[10] Z.-Q. Hong and J.-Y. Yang, Optimal discriminant plane for a small number of samples and design method of classifier on the plane, *Pattern Recognition* , vol. 24, 1991, pp 317-324.

[11] H. Hotelling, Relations between two sets of variates, *Biometrika*, vol. 28, 1936, pp 321-377.

[12] A.J.Izenman, Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis*, vol. 5, 1975, pp 248-264.

[13] E. Kidron, Y.Y. Schechner and M. Elad, Pixels that sound, *Computer Vision and Pattern Recognition*, vol. 1, 2005, pp 88-95.

[14] F.J. Lapointe and P.Legendre, A classification of pure malt scotch whiskeys, *Applied Statistics-Journal of the Royal Statistical Society*, vol. 43, 1994, pp 237-257.

[15] T. Melzer, M. Reiter and H. Bischof, Appearance models based on kernel canonical correlation analysis, *Pattern Recognition*, vol. 36, 2003, pp 1961-1971.

[16] M. Schumacher and V. Blanz, Which facial profile do humans expect after seeing a frontal view? A comparison with a linear face model, *ACM Transaction on Applied Perception*, vol. 9, 2012, pp 1-16.

[17] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng and D.S. Xia, A new method of feature fusion and its application in image recognition, *Pattern Recognition*, vol. 38, 2005, pp 2437-2448.

[18] M.E. Tipping and C.M. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society, Series B*, vol. 61, 1999, pp 611-622.

[19] W.Zheng, X. Zhou, C. Zou and L. Zhao, Facial expression recognition using kernel canonical correlation analysis (kcca), *IEEE Transactions on Neural Networks* vol. 17, 2006, pp 233-238.