ELSEVIER

Image and Vision Computing 18 (1999) 9–19

image and vision COMPUTING

www.elsevier.com/locate/imavis

3D shape and 2D surface textures of human faces: the role of “averages” in attractiveness and age

A.J. O’Toole[a,*], T. Price[a], T. Vetter[b,1], J.C. Bartlett[a], V. Blanz[b]

[a]School of Human Development, The University of Texas at Dallas, GR4.1, Richardson, TX 75083-0688, USA
[b]Max Planck Institut fur biologische Kybernetik, Spemannstr. 38, Tubingen 72076, Germany

Received 16 April 1998; accepted 27 March 1999

Abstract

Recent work in the psychological literature has indicated that attractive faces are in some ways “average” [J.H. Langlois, L.A. Roggman, Attractive faces are only average, Psychological Science, 1(2) (1990) 115–121] and that the apparent age of a face can be related to its proximity to the average of a computationally derived “face space” [A.J. O’Toole, T. Vetter, H. Volz, E.M. Salter, Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age, Perception, 26 (1997) 719–732]. We examined the relationship between facial attractiveness, age, and “averageness”, using laser scans of faces that were put into complete correspondence with the average face [T. Vetter, V. Blanz, Estimating coloured 3D face models from single images: an example based approach, in: H. Burkhardt, B. Neumann (Eds.), Proceedings of the Fifth European Conference on Computer Vision, Freiburg, Germany, 1998, pp. 499–513]. This representation enabled selective normalization of the 3D shape versus the surface texture map of the faces. *Shape-normalized* faces, created by morphing the texture maps from individual faces onto the average head shape, and *texture-normalized faces*, created by morphing the average texture onto the shape of each individual face, were judged by human subjects to be both more attractive and younger than the original faces. The study shows that relatively global, psychologically meaningful attributes of faces can be modeled very simply in face spaces of this sort. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Human face; Morphing; Principal component analysis; Correspondence; Image description

1. Introduction

The relationship between human image perception and artificial image manipulations is a central problem for many image processing applications. One aspect of this problem concerns the relationship between some relatively global, but psychologically meaningful perceptual descriptors of objects, and the image properties that underlie these percepts. For example, human faces can be described in a number of global ways that are likely to be the result of a combination or *configuration* of image based features: e.g. “attractive”, “generous”, “mean-looking” or “thirty-something”. In the present study, we focus on understanding two of these dimensions in ways that will allow us to change images selectively along specific perceptual dimensions while keeping other dimensions constant (e.g. to beautify or age a face image without changing the identity of the person). One application of this approach is to the problem of image search in databases, for which the mapping of human image descriptions onto formal image representations can substantially increase the efficiency of the search.

There is good evidence in the literature that at least one component of being “attractive” is related to being “average” [1]. The primary psychological evidence for this claim comes from a study showing that “composite images” made by averaging together the faces of several individuals are judged by human subjects to be more attractive than the original unaveraged faces. Although controversial for a number of technical [4–7], and theoretical [8,9] reasons, which we discuss below, the attractiveness of “average” faces is an interesting result both for psychological and computational models of face recognition.

As has been discussed in great detail elsewhere, the attractiveness of averages versus extremes has important implications for theories of mate selection based on evolutionary biology [1,4,8,10–13]. Much less considered, however, are the implications of this finding for human memory or more precisely for understanding the “recognizability” of average

* Corresponding author. Tel.: + 1-972-883-2486; fax: + 1-972-883-2491.
E-mail addresses: otoole@utdallas.edu (A.J. O’Toole); vetter@mpik-tueb.mpg.de (T. Vetter)
[1] Also corresponding author.

PII: S0262-8856(99)00012-8

faces. By recognizability we mean simply the accuracy with which a face can be correctly recognized when seen before, and correctly judged "novel" when it has not been seen before. In fact, psychological work has indicated, somewhat counterintuitively, that attractive faces are recognized by human subjects less accurately than are unattractive faces [14]. These findings [5,11,14] link facial attractiveness to very well established findings relating the perceived typicality/distinctiveness of a face and its recognizability, e.g. [15,16]. People recognize typical or "average" faces less accurately than distinctive faces. This phenomenon has important implications for the expected accuracy of eyewitness identifications for *individual faces*. Simply put, some faces are more likely to be falsely recognized than others, and so, the credibility of eyewitness identifications will vary systematically with the typicality and attractiveness of the face to be identified.

One unexplored application of computational models of face recognition concerns the ability to predict the accuracy with which human subjects will recognize *individual* faces, i.e. predict which faces humans will identify correctly and which faces may generate identification errors. Quantifying the information that makes a face attractive or typical, thus, has possible applications to this problem. The problem, however, poses challenges for computational models of face recognition for two reasons. First, it is likely that the information that makes a face either typical or (non-equivalently) attractive is related, at least in part, to the configuration of features in a face, rather than exclusively to any single feature [9]. Second, both of these facial attributes make implicit reference to a population of faces. For example, faces are likely to be considered typical or attractive relative to a reference group that may involve the sex, race, and age of a face [17].

A computational model for manipulating the attractiveness or distinctiveness of a face in ways that are perceptually salient for human subjects should, therefore, be sensitive to the statistical structure of a set of faces. Such a model should also be based on a representation in which the configural structure of a face can be manipulated in a global and relatively natural way. A "face space" representation, commonly used in both the psychological [18] and the computational literatures [2,19–22] meets these requirements. In its generic form a face space entails the following notions: (1) faces can be thought of as points in a multidimensional space; (2) the axes of this space represent a set of features on which the faces are encoded; and (3) the distance between any two faces in this space is a measure of the similarity between the faces [18,23].

In the more quantitative literature, face spaces are typically implemented by using principal components analysis (PCA) [19–22] of a covariance matrix made using a set of face images. This yields a set of feature axes (principal components, PCs), which are derived directly from the statistical structure of the set of faces. Individual faces are points/vectors in this space and thus can be described by their coordinates in the space, or in other words, by their values on each "feature axis" or PC. Finally, it is worth noting that when PCA is applied to a physical measure of faces, such as pixels, surface values, or fiducial point location codes,[2] the resultant PCs are of the same form and the faces can be expressed as weighted linear combinations of these "features"/PCs. As such, alterations to faces that are made by operating on their coordinates in this space can be viewed (if they are image-based), constructed (if they are surface based), or synthesized (if they are derived from a fiducial code, or other code in which the faces are represented in a comparable, corresponded/aligned coordinate system, [3,24]). A primary question in the psychological literature over the past few years concerns the kind of representation (e.g. image, surface or some combination) that is best for modeling human perception and memory for objects and faces (cf. [25,26]).

It has been posited in the psychological literature that the distinctiveness of a face is related to its distance from the average face in a generic face space [18]. In fact, his is the primary manipulation used by most automatic caricature generators that operate in quantitatively-based face spaces. The goal of a caricature generator is to increase the "distinctiveness" of a face by exaggerating features that help to differentiate the face from other faces, e.g. Refs. [27,28]. More formally, automatic caricatures usually work as follows. First, a measure of the average value of a set of "features" across a large number of faces is computed. These features are defined, usually, as a set of facial landmark locations or "fiducial points" (e.g. corners of the eye and other points that are reasonably easy to localize/match on all faces). It is worth noting that this is a representation of the 2D configural structure of the face because it captures the spatial layout of the facial features in the projected 2D facial image. Next, to create a caricature of an individual face, a measure of the deviation of the face from the average 2D configuration is computed. Finally, "distinctive" or unusual features of the face are exaggerated and the face is redrawn with the exaggerated features to produce the caricature. The basic manipulation of an automatic caricature-generator, therefore, is to "move" the face away from the center of a face space based on the 2D (projected) configural stucture of a face.

In recent years, computational face spaces have been used by psychologists to ask questions about the nature of human representations of faces. The logic behind this approach is straightforward. Computational face spaces derived from different kinds of face representations (2D pixel-based images [19–22], 3D surfaces from laser scans [24]) may make different predictions about the similarity/confusability of faces. More formally, the distance between two faces in a face space based on 2D pixel-based images may be very

[2] Fiducial point codes comprise the locations of a set of facial landmarks, e.g. corners of the eyes. The number and variety of encoded fiducial points can vary widely across different applications.

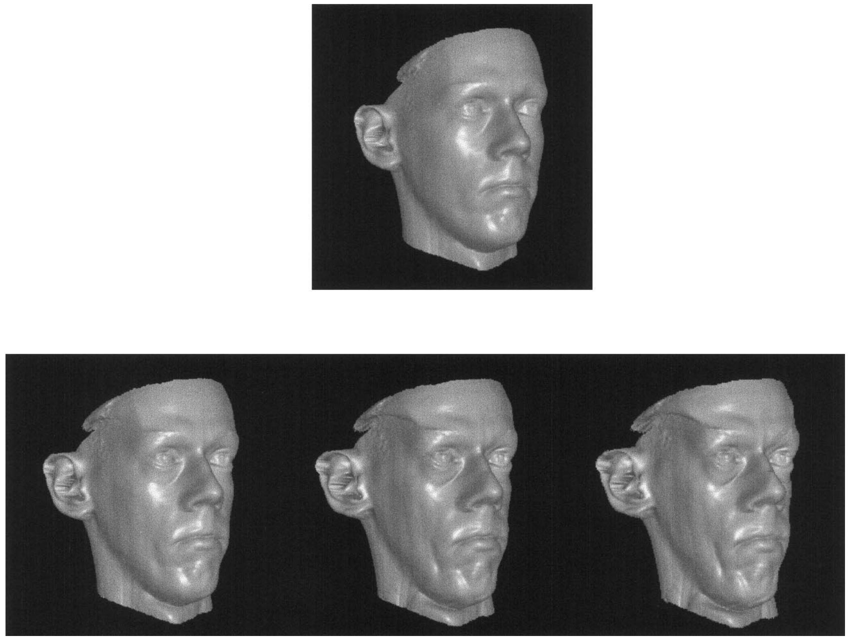

Fig. 1. 3D caricatures of a 26-year-old male (row 1). Increasing the level of exaggeration, i.e. distance from the mean in the face space, increases the apparent age of the face (row 2: left to right).

different than the distance between faces in a three-dimensionally based face space. It is possible then to test hypotheses about human representations of faces by varying the nature of the features used to create computational face spaces and by using psychological data to evaluate the adequacy of different face space representations as models for human perception and memory.

Before proceeding, it is worth illustrating briefly that computationally derived face spaces can differ both quantatively and qualitatively in the predictions they make about perceptual variations in facial appearance. For example, recent work illustrates that the application of an automatic caricature algorithm to faces represented by their 3D structure alters the age of a face more than its distinctiveness [2] (see Fig. 1).[3] In that study, faces were represented as vectors in a PCA-based face space derived from a low level encoding of the 3D head structure. The caricature algorithm operated as follows: (a) a face vector in this space was multiplied by a scalar, x ($x > 1$ yields a caricature; $x < 1$ yields an anticaricature)[4]; and (b) the caricature was created by recombining the PCs/eigenheads according to their new coordinates. Using this representation, it is clear that the direction of the face (vector) in this space represents the identity of the face. All of the faces pictured in Fig. 1 are actually on the line that connects the average face to the veridical face (and continues). As can be seen, all the faces retain the identity of the original. The face in the first row is the actual head scan of a 26-year-old male. The three faces in row 2 are increasing levels of caricature. Here it can be seen that the distance from the average, i.e. the length of the vector, represents its distinctinctiveness, and in this particular case, its age as well [2].

A generic caricature applied to a computationally derived face space based on a 3D representation of faces produced a very salient change in the age of faces. Applied to a 2D configural representation of faces, a similar trajectory in the face space produced more salient changes in the distinctiveness of faces. Thus, when implementing simple algorithms for manipulating the appearance of faces, the nature of the features underlying the face space has important perceptual consequences.

In the present study, we explored the question of quantifying and manipulating facial attractiveness and age in the context of the average(s) of a physical face space. Additionally, we have based the face space on a more complete representation of faces than has been used previously. The representation combines both the 3D stucture of the faces and the overlying two-dimensionally based texture map. An example laser scan stimulus appears in Fig. 2, with the combined surface and texture map rendered from the front, the pure surface map in the center, and the texture map on the right. Further, in our representation, the faces are in complete correspondence with each other, i.e. are aligned so that the positions of the discrete features overlap [3]. We describe this procedure shortly. For present purposes, this approach overcomes two shortcomings of the averaging manipulation used to make composite faces [4–7]. These shortcomings confound the nature of the information being manipulated.

[3] It is worth noting also that the averaged faces in Ref. [6] appeared younger than the veridicals.

[4] Note that because the face space was based on 3D deformation fields, which we discuss shortly, the origin of this space was the average face.

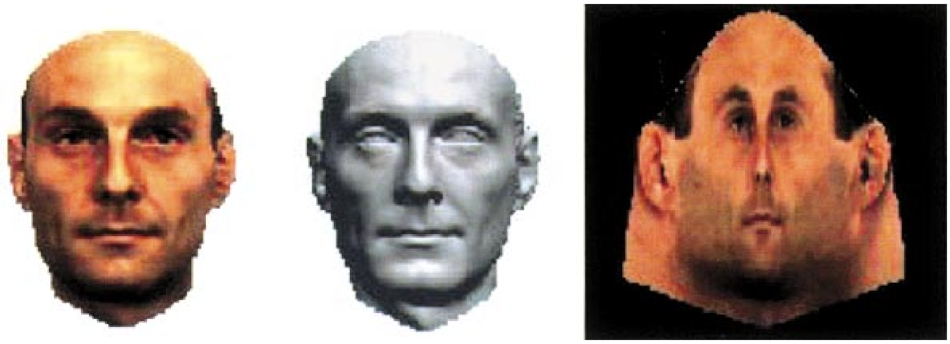

Fig. 2. Laser scan data, 3D structure and texture combined and rendered from the front (left), 3D head surface data (middle) and texture map (right).

The first shortcoming is that the composite procedure allows for the possibility of blurring out facial blemishes and other small imperfections in the face when averaging faces. This might provide an alternative explanation for the results reported in Ref. [1], (though see also Refs. [5,6,11]). The second shortcoming relates to the first but concerns the more general problem of blurring as it affects the alignment of facial features prior to the averaging. Specifically, in simple arithmetic averaging, the exact positions of the features (eyes, etc.) are not aligned prior to the averaging procedure, and hence, may be blurred in the final averaged face image.

In the present study, we used an automated "correspondence" algorithm applied simultaneously to both the 2D and 3D information from laser scans of human faces [3]. Using this representation we were able to ask more precise questions about the relationship between "averageness" and the perceived attractiveness of human faces. The purpose of a correspondence algorithm in this context is to put all faces into a comparable coordinate system before "moving them" toward the average. Finding a common coordinate system in morphing and automated caricature generator procedures [27,29] is done usually by a human operator who hand-locates between 50 and 300 fiducial (and supplemental) points on the face, prior to warping and interpolating. For example, the lower lip of a face might be represented by the locations of 12 points, the left and right corners of the mouth and 10 equally spaced intervening points.

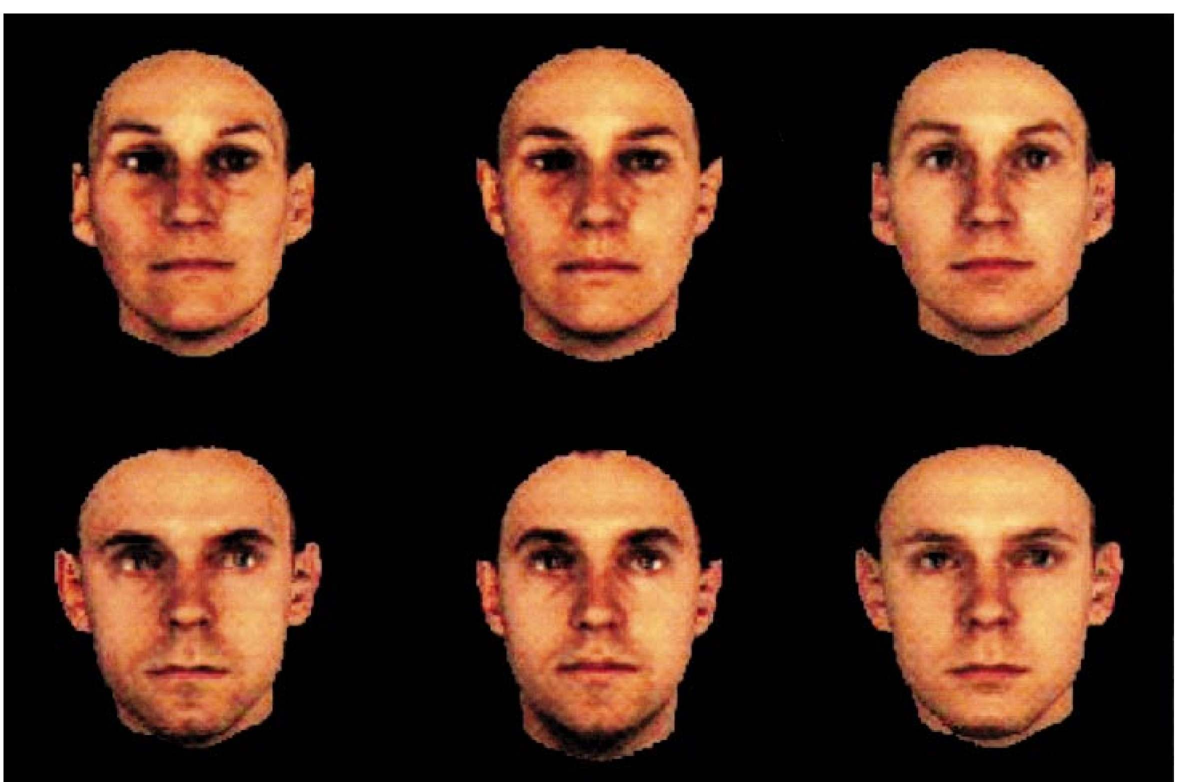

Fig. 3. Original face (left), 3D, shape-normalized (middle) and texture-normalized (right), female and male faces.

The automated correspondence algorithm used in this study casts the matching problem into its more general computer vision form in which one attempts to match all of the data points in two images/surfaces, rather than just a subset of the fiducial points. This is the approach taken most commonly in solving the classical correspondence problems in stereopsis and motion analyses. Although this problem is far from solved in a perfectly general form, a great deal of progress has been made recently on the problem with faces. Specifically, several methods based on elaborated optic flow algorithms [30] have been applied successfully to the task of automating a correspondence finding procedure for images of human faces [31–33]. These approaches have been extended successfully to laser scan data for human heads [3]. Full details of how the correspondence algorithm works can be found in Refs. [3,34] and an outline of the implementation details for this paper are summarized in Appendix A.

For present purposes, each face is represented as a "deformation field" from the average, which can be divided into the parts/subspaces based on: (a) the 3D head structure from the laser scan; and (b) the overlying 2D surface texture, which is mapped point-for-point onto the head surface [3,34].[5] In short, what is represented from each individual face is how it differs from the average in terms of its 2D texture and 3D shape.

This fully corresponded representation of the combined surface and texture information enables us to address the shortcomings of the composite approach in a straightforward and complementary way. A face can be moved toward the average by simply drawing a line between the face and the average and then "moving" the face toward the average, reconstructing it at its new coordinates (i.e. combining the PCs linearly using the new coordinates as weights). Thus, the problem of blurring blemishes and small imperfections can be solved by moving the face within the subspace defined only by the face shapes, leaving the original texture in tact.[6] Likewise, the shape can be retained and the surface image texture can be moved toward the average.

More formally, we created shape-normalized faces by morphing the texture maps from individual faces onto the average head shape and texture-normalized faces by morphing the average texture onto the shape of each individual face. In a face space model based on the shape and texture deformation fields of faces, the average face lies at the origin of the space. These normalization procedures, therefore, amount to a simple operation of zeroing out the face's coordinates in the subspace corresponding to either the texture or shape of the face. Samples of these stimuli appear in Fig. 3. The left column contains two normal faces, the middle column contains the shape-normalized versions of the faces and the right column shows the texture-normalized versions of the faces.

A final question we considered in the present study concerns whether the perceived age of a face decreases as the face is moved toward the average/center of the face space. Suggestions to this effect have been reported previously [6] for the composite procedure. In that study, averaged faces appeared younger than their originals. The authors, however, did not find a correlation between the attractiveness and estimated age of the *unaveraged faces*.[7] The work with 3D representations has also indicated that moving faces toward the center of a face space has a rather dramatic effect on the apparent age of a face [2].

The purpose of the present study was very straightforward. We wished to measure the relative contributions of 3D shape averaging versus 2D texture averaging to the findings that one component of attractive faces is related to being average. We also wished to examine explicitly, the contribution of "de-aging" to the effect. In the first experiment, we asked human subjects to judge the attractiveness of the original, shape-normalized, and texture-normalized faces. These judgments were supplemented in Experiment 2 with estimates of the ages of the original and altered faces. Finally, we applied partial correlation techniques to the problem of assessing the independence of the effects of shape- and texture normalization on the age and attractiveness of faces.

2. The stimuli

2.1. Description of laser-scanned heads

Laser scans (Cyberware™) of 100 heads of young adults (50 male and 50 female) comprised the stimulus data base. The mean age of faces in the database was 26.9 years (standard deviation = 4.7 years). The subjects were scanned wearing bathing caps, which were removed digitally. The laser scans provided surface map data consisting of the lengths of 512 × 512 radii from a vertical axis centered in the middle of the subject's head to "sample" points on the surface of the head. This is a cylindrical representation of the head surface, with surface points sampled at 512 equally spaced angles around the circular slices of the cylinder, and at 512 equally spaced vertical distances along the long axis of the cylinder. Additionally, further pre-processing of the heads was done by making a vertical cut behind the ears, and a horizontal cut to remove the shoulders. A subset of 48 (24 males and 24 females) was selected randomly from this

[5] We wish to note here that the use of the term "2D" with the surface texture should be qualified somewhat. Although the information captured in the texture is inherently 2D, due to the fact that the laser scanner uses ambient light, the texture is normally viewed wrapped around a head surface, from a specific viewpoint, and under specific lighting conditions.

[6] Although previous work [11] has suggested that these blemishes cannot provide a full account of the effects reported in Ref. [1], that study used line drawings as the control for blurring and so a number of additional features of the image may also have altered.

[7] Though it should be noted that there was only minimal age variability in the faces used there and hence it may not have been possible with such a small range to detect a correlation.

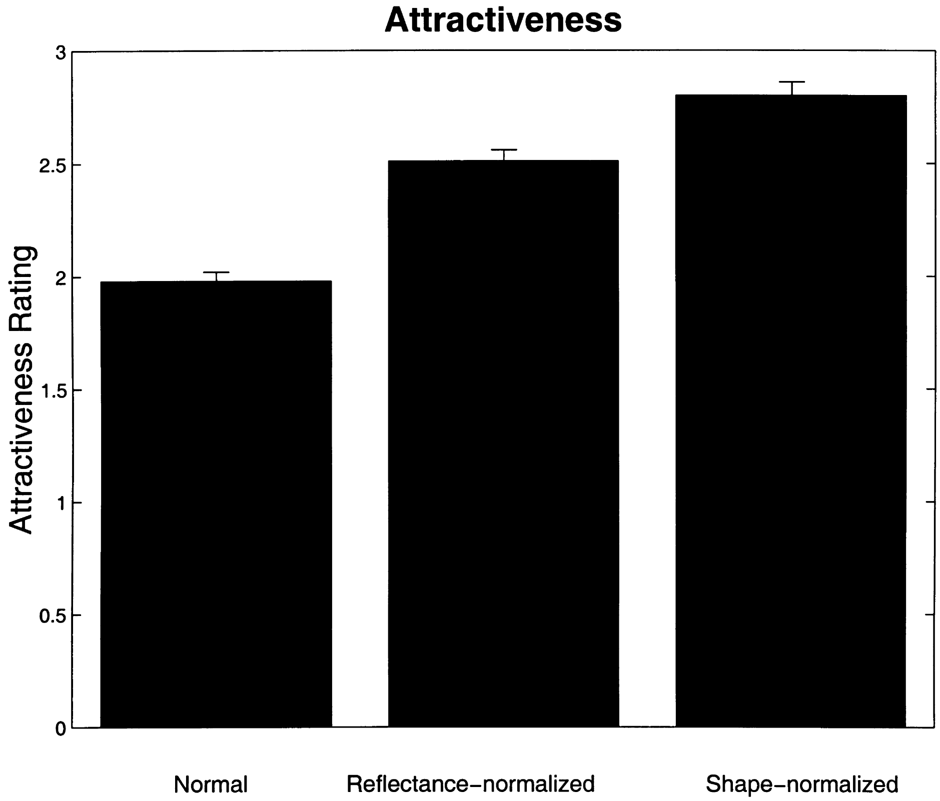

Fig. 4. Attractiveness ratings for the normal (left), shape-normalized (middle) and texture-normalized (right) faces.

data base to serve as stimuli in the experiments reported here.

2.2. The correspondence problem

The procedures applied to solving the correspondence problem for this particular set of laser scan stimuli are complex but the basic principles have been described in detail elsewhere [3,34]. Additionally, to make this manuscript self-contained, we describe the implementation details of the algorithm in Appendix A.[8] For present purposes, the basic idea is to match the data points in each individual face with the corresponding feature points in the average face, with the goal of representing each face as a "deformation" field from the average. Thus each data point in the face representation contains a pointer to the analogous data point in the average. This was done by applying optic flow algorithms optimized in this case to deal with the continuous surface and texture data found in faces [34].

2.3. 3D shape and 2D texture normalization

Three sets of faces were made from these original laser scans. Two sets of stimuli were made from the original surface and texture maps of 48 faces. Texture-normalized faces were created by wrapping the average texture map onto the surface map of each individual face. Shape-normalized faces were made by mapping the texture maps of each individual face onto the average shape. Each resulting face was rendered from the frontal viewpoint (Fig. 3).

3. Experiment 1—facial attractiveness

3.1. Procedure

Thirty-six volunteers (17 males and 19 females) from the University of Texas at Dallas (UTD) participated in this experiment. Most of these volunteers were undergraduate students compensated with a research credit for a core course in the psychology curriculum. Observers read instructions which indicated the purpose of the experiment and were told to rate the attractiveness of each face presented to them on a scale of 1–5, with 1 being least attractive and 5 being most attractive. Each subject viewed the full set of 144 (normal, shape-normalized, and texture-normalized versions of each of the 48 individual faces). The face remained visible until an attractiveness rating was given. The experiment was conducted on a Macintosh computer programmed using PsyScope [36].

3.2. Results

The mean attractiveness ratings for each subject on each type of face were computed. These data appear in Fig. 4. As indicated by the figure, attractiveness ratings were

[8] A version of Appendix A appears also in Ref. [35].

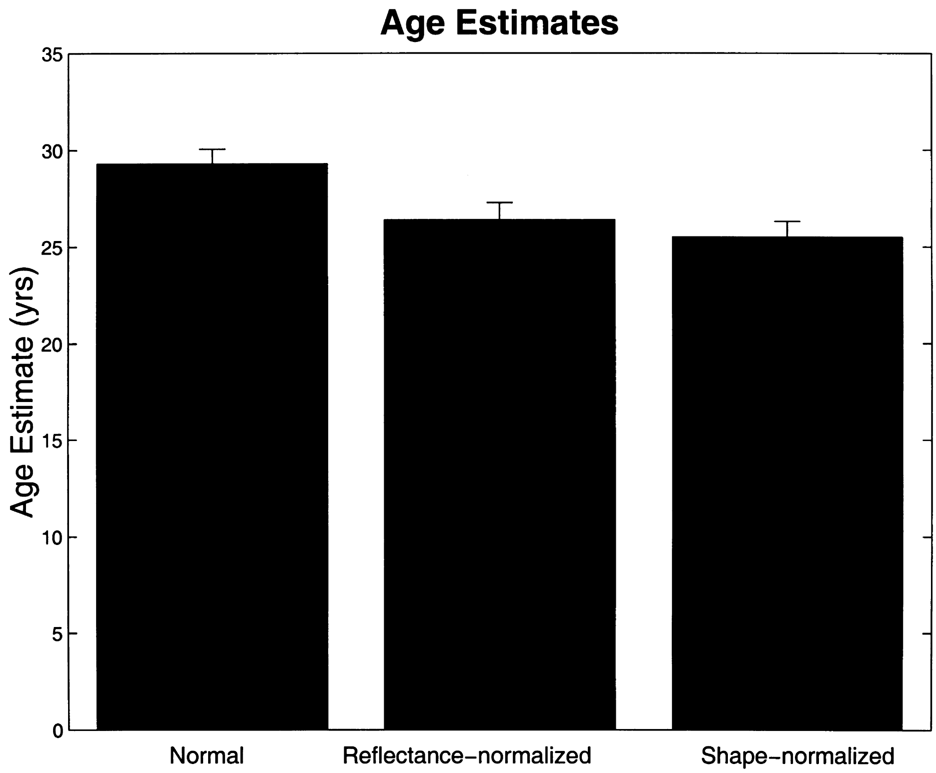

Fig. 5. Age estimates for the normal (left), shape-normalized (middle) and texture-normalized (right) faces.

highest for the shape-normalized faces, followed by the texture-normalized faces and then the normal faces. These data were submitted to a three-factor analysis of variance with face type and face gender as within-subjects independent variables and subject sex as a between-subjects variables. We found a main effect of face type, $F(2,68) = 134.24$, $p < 0.0001$. No other factors or interactions approached significance.

3.3. Discussion

These results indicate clearly that normalizing the faces with respect both to their 3D structure and their 2D texture increased the attractiveness of faces. This replicates the principal findings of Ref. [1] and extends them in several ways. First, the shape-normalization affected the attractiveness more than did the texture-normalization. The fact that the shape-normalized faces, which retain their image-based blemishes and imperfections, were considered more attractive than the originals indicates that the "blurring" of small imperfections in the face images cannot account entirely, or even primarily, for the previous results. In fact, the shape-normalized faces which retained their image-based imperfections were considered more attractive than the texture-normalized faces.

Next, both our human subjects and ourselves noticed spontaneously that the normalized faces seemed to appear younger than the originals (cf. also Ref. [6]). In Experiment 2, we examined this question formally by collecting age estimates on the normal, shape- and texture-normalized faces which we could then relate to the attractiveness ratings collected in Experiment 1.

4. Experiment 2—age estimation

4.1. Procedure

Twenty-eight volunteers (15 females and 13 males) participated in this phase of the experiment. Observers were assigned to one of three groups to estimate the ages of normal, shape-normalized or texture-normalized faces. They did this by typing in an age estimate on the computer keyboard. The face remained visible until the estimate was made. Due to the more demanding and time consuming nature of this phase of the experiment we counterbalanced the testing such that each subject saw only one group of faces (normal, texture-normalized, or shape-normalalized) and judged the age for each face in that particular set.

4.2. Results

The mean age estimate ratings for each subject on each face were computed and divided according to the face type rated. These data appear in Fig. 5 and indicate that age estimates for both the shape- and texture-normalization were younger than for the normal faces. Shape-normalization decreased the apparent age of the face by about 4 years and texture-normalization decreased the apparent age by about 3 years. The pattern of effect is similar to that seen for the attractiveness ratings, with the shape-normalized faces judged youngest, the texture-normalized next, and the originals judged oldest. More formally, these data were submitted to a three-factor analysis of variance with face sex as a within-subjects independent variable and face type and subject sex as between-subjects variables. We found a main effect of face type, $F(2,22) = 3.38$, $p < 0.05$, with age estimations decreasing from normal to texture-normalized to shape-normalized. The sex of the face was also significant, $F(1,22) = 45.41$, $p < 0.0001$, however, this could be due to the actual age variation between males and females in the group of faces presented. No other factors or interactions proved significant.

5. Combined analysis—attractiveness and age

The results of Experiment 2 indicated that both the shape and texture normalization manipulations decreased the apparent age of the faces. Given that these manipulations also increased the attractiveness of the faces, and that the pattern of these effects were similar, we were interested in assessing the extent to which the de-aging effect could account for the increased attractiveness.

To determine this we applied a partial correlation technique to the combined data from Experiments 1 and 2. This worked as follows. For each face, we used the age estimates supplied by the human subjects for individual faces in Experiment 2 as a predictor for the attractiveness judgments supplied by the subjects in Experiment 1. The error of these estimates (the residuals) were then analyzed with an ANOVA, using face rather than subject as the unit of measure. The pattern of means for the residuals, with the age component partialed out, was the same as that seen for the raw data, with the shape-normalized faces judged most attractive, the texture-normalized faces next, and finally the original faces. Face type was again statistically significant, $F(2,92) = 47.42$, $p < 0.0001$, indicating that even taking into account the de-ageing effects of the manipulations, the attractiveness of the faces was increased by moving them toward the average.

6. General discussion

The relationship between human image perception and artificial image manipulations is a central problem for many image processing applications. An understanding of this will allow us to change images selectively along even relatively abstract specific perceptual dimensions. For the problem of image search in databases, the mapping of human image descriptions onto formal image representations can increase the efficiency of the search.

In the present study, we used a computationally defined face space based on a representation of how the faces differ in their 3D shape and 2D texture from the average face. Our primary manipulation consisted of altering the length of the face vectors in a selected subspace of the general face space. This manipulation is opposite to that carried out normally in automated caricature generators. Faces increased in attractiveness and decreased in apparent age with shape- or texture normalization. Additionally, we showed that although the normalization procedure simultaneously affects both the age and attractiveness of the faces, the perception of these two facial attributes was not synonymous.

We think these results are important for three reasons. First, although there is much data in the psychological literature to suggest that humans process faces configurally rather than as a set of features, computational models have not always considered representational systems in this light. Altering the global information in faces in ways that change psychologically meaningful facial attributes like attractiveness and typicality/distinctiveness can be done very simply using an appropriate face space representation. In this work, increasing the attractiveness of faces can be seen as a kind of inverse operation to caricaturing, instead of increasing the distance of an individual face to the average we replace parts of the face representation by average values.

Second, the present results help to clarify the relationship between some relatively abstract attributes of faces that are of some consequence for understanding human memory for faces. To recognize a face, one needs to encode the information that makes it different from all other faces in the world. Faces vary, however, in the extent to which they differ from other faces in the world. The present results indicate that there is a relationship between the attractiveness of faces and their closeness to the average face. Previous work has indicated that these shape- and texture-normalized faces are recognized by human subjects less accurately than are the original faces [35].

Third, where age is concerned, the results further clarify the importance of paying careful attention to the nature of the underlying face space representation. In a purely three-dimensionally based face space, age was the primary perceptual correlate for face vector length [2]. Using a combination of the texture and shape, both attractiveness and age related to vector length, with the age component being far less potent here than it was for a purely 3D representation of faces [2].

Finally, we wish to note clearly that although "average" is in some ways attractive, the present results do not suggest that it is the only source of attractiveness. Many previous studies have shown convincingly that atypical aspects of faces can be perceived as attractive, most notably [8]. The present work shows only that at least some aspects of the averageness of faces can be linked reliably to the attractiveness and age of faces. Combined with other data on the recognizability of these shape and texture-normalized faces [35], it links a global face descriptor that humans use quite comfortably, i.e. attractiveness, to the accuracy of human memory for faces. It would be of great interest to see if this kind of global measure would be equally useful for predicting the performance of computational models of face recognition at the level of individual faces.

Acknowledgements

A.O. gratefully acknowledges support from NIMH grant 1R29MH5176501A5, the Alexander von Humboldt Foundation, and Texas Instruments Inc.

Appendix A. 3D correspondence algorithm

In order to construct a general flexible 3D face model that allows for computing an average face as well as for exchanging shape and texture between different faces, it is crucial to establish correspondence between a reference face and each individual face example. For all vertices of the reference face, we have to find the corresponding vertex location on each face in the dataset. If, for example, vertex j in the reference face is located on the tip of the nose, with a 3D position described by the vector components X_j, Y_j, Z_j in S_{ref}, then we have to store the position of the tip of the nose of face i in the vector components X_j, Y_j, Z_j of S_i. In general, this is a hard problem, and it is difficult to formally specify what correct correspondence is supposed to be. However, assuming that all face data sets are roughly aligned and that there are no categorical differences such as some faces having beards and others not, an automatic method is feasible for computing the correspondence (the algorithm is described in more detail in Refs. [3,34]).

For matching points on the surfaces of two 3D objects we modified an existing optical flow algorithm developed for 2D images.

A.1. Optical flow algorithm

In video sequences, in order to estimate the velocities of scene elements with respect to the camera, it is necessary to compute the vector field of optical flow, which defines the displacements $(\delta x, \delta \mathrm{y}) = (x_2 - x_1, y_2 - y_1)$ between points $p_1 = (x_1, y_1)$ in the first image and corresponding points $p_2 = (x_2, y_2)$ in the second image. A variety of different optical flow algorithms have been designed to solve this problem (for a review see Ref. [37]). Unlike temporal sequences taken from one scene, a comparison of images of completely different scenes or faces may violate a number of important assumptions made in optical flow estimation. However, some optical flow algorithms can still cope with this more difficult matching problem, opening up a wide range of applications in image analysis and synthesis [31].

In a previous study [33], we computed correspondence between face images using a coarse-to-fine gradient-based method [38] applied to the Laplacians of the images and followed an implementation described in Ref. [30]. The Laplacian of the images were computed from the Gaussian pyramid adopting the algorithm proposed by [39]. For every point x,y in an image $I(x,y)$, the algorithm attempts to minimize the error term $E = \sum (I_x \delta x + I_y \delta y - 1)^2$ for δx, δy, with I_x, I_y being the spatial image derivatives of the Laplacians and δI the difference of the Laplacians of the two compared images. The coarse-to-fine strategy starts with low resolution images and refines the computed displacements when finer levels are processed. The final result of this computation $(\delta x, \delta y)$ is used as an approximation of the spatial displacement of each pixel between two images.

A.2. 3D face representations

The extension of this optical flow algorithm to the 3D head data is straightforward due to the fact that the cylindrical representation of a head surface is analogous to images: Instead of grey-level values in image coordinates x,y, here we store the radius values and the color values for each angle ϕ and height h. A parameterization of a 3D head in cylindrical coordinates, therefore, consists of two 'images', one representing the geometry of the head and the other containing the texture information. In order to compute the correspondence between different heads, both texture and geometry were considered simultaneously. The optical flow algorithm as described earlier had to be modified in the following way. Instead of comparing a scalar grey-level function $I(x,y)$, our modification of the algorithm attempts to find the best fit for the vector function

$$\vec{F}(h, \phi) = \begin{pmatrix} \mathrm{radius}(h, \phi) \\ \mathrm{red}(h, \phi) \\ \mathrm{green}(h, \phi) \\ \mathrm{blue}(h, \phi) \end{pmatrix}$$

in a norm

$$\left\| \begin{pmatrix} \mathrm{radius} \\ \mathrm{red} \\ \mathrm{green} \\ \mathrm{blue} \end{pmatrix} \right\|^2$$

$$= w_1 \mathrm{radius}^2 + w_2 \mathrm{red}^2 + w_3 \mathrm{green}^2 + w_4 \mathrm{blue}^2.$$

The coefficients $w_1, \ldots, w_4$ correct for the different contrasts in range and color values, assigning approximately the same weight to variations in shape as to variations in all color channels taken together.

For representing the geometry, radius values can be replaced by other surface properties such as Gaussian curvature or surface normals.

The displacement between corresponding surface points is captured by a correspondence function

$$C(h, \phi) = \begin{pmatrix} \delta h(h, \phi) \\ \delta \phi(h, \phi) \end{pmatrix}.$$

A.3. Interpolation in low-contrast areas

It is well known that in areas with no contrast or with strongly oriented intensity gradients, the problem of optical flow computation cannot be uniquely solved based on local image properties only (aperture problem). In our extension of the algorithm to surfaces of human faces, there is no structure to define correct correspondence on the cheeks, along the eyebrows and in many other areas, and indeed the method described so far yields spurious results there.

The ambiguities of correspondence caused by the aperture problem can be resolved if the flow field is required to be smooth.

In our algorithm, smoothing is performed as a separate process after the estimation of flow on each level of the coarse-to-fine approach. For the smoothed flow field $(\delta h'(h,\phi), \delta\phi'(h,\phi))$, an energy function is minimized using conjugate gradient descent such that on the one hand, flow vectors are kept as close to constant as possible over the whole domain, and on the other hand as close as possible to the flow field $(\delta h(h,\phi), \delta\phi(h,\phi))$ from the computation described above. The first condition is enforced by quadratic potentials that increase with the square distances between each individual flow vector and its four neighbors. These interconnections have equal strength over the whole domain. The second condition is enforced by quadratic potentials that depend on the square distance between $(\delta h'(h,\phi), \delta\phi'(h,\phi))$ and $(\delta h(h,\phi), \delta\phi(h,\phi))$ in every position (x,y). These potentials vary over the parameter domain. If the gradient of color and radius values, weighted in the way described above, is above a given threshold, the coupling factor is set to a fixed, high value in the direction along the gradient, and zero in the orthogonal direction. This allows the flow vector to move along an edge during the relaxation process. In areas with gradients below threshold, the potential is vanishing, so the flow vector depends on its neighbors only.

After all individual faces of the training set have been matched to a reference face, the average 3D shape as well as the average surface texture map can be computed. Additionally, correponding values of surface texture of different faces can be exchanged.

References

[1] J.H. Langlois, L.A. Roggman, Attractive faces are only average, Psychological Science 1 (2) (1990) 115–121.

[2] A.J. O'Toole, T. Vetter, H. Volz, E.M. Salter, Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age, Perception 26 (1997) 719–732.

[3] T. Vetter, V. Blanz, Estimating coloured 3D face models from single images: an example based approach, in: H. Burkhardt, B. Neumann (Eds.), Proceedings of the Fifth European Conference on Computer Vision, Freiburg, Germany, 1998, pp. 499–513.

[4] T.R. Alley, M.R. Cunningham, Averaged faces are attractive, but very attractive faces are not average, Psychological Science 2 (2) (1991) 123–125.

[5] J.H. Langlois, L.A. Roggman, L. Musselman, S. Acton, picture is worth a thousand words: reply to 'on the difficulty of averaging faces', Psychological Science 2 (5) (1991) 354–357.

[6] J.H. Langlois, L.A. Roggman, L. Musselman, What is average and what is not average about attractive faces, Psychological Science 5 (4) (1994) 214–220.

[7] J.B. Pittenger, On the difficulty of averaging faces: comments on Langlois and Roggman, Psychological Science 2 (5) (1991) 351–353.

[8] D.I. Perrett, K.A. May, S. Yoshikawa, Facial shape and judgements of female attractiveness, Nature 368 (1994) 239–242.

[9] A.J. O'Toole, K.A. Deffenbacher, D. Valentin, K. McKee, D. Huff, H. Abdi, The perception of face gender: The role of stimulus structure in recognition and classification, Memory and Cognition 26 (1998) 146–160.

[10] S.W. Gangestad, R. Thornhill, R.A. Yeo, Facial attractiveness, developmental stability, and fluctuating asymmetry, Ethology and Sociobiology 15 (1994) 73–85.

[11] G. Rhodes, T. Tremewan, Averageness exageration and facial attractiveness, Psychological Science 7 (2) (1996) 105–110.

[12] G. Rhodes, A. Sumich, G. Bryatt, Are average configurations attractive only because of their symmetry, Psychological Science 10 (1999) 52–58.

[13] R. Thornhill, S.W. Gangestad, Human facial beauty: averageness, symmetry, and parasite resistance, Human Nature 4 (3) (1993) 237–269.

[14] L.L. Light, S. Hollander, F. Kayra-Stuart, Why attractive people are harder to remember, Personality and Social Psychology 7 (2) (1981) 269–276.

[15] L. Light, F. Kayra-Stuart, S. Hollander, Recognition memory for typical and unusual faces, Journal of Experimental Psychology: Human Learning and Memory 5 (1979) 212–228.

[16] T. Valentine, V. Bruce, The effects of distinctiveness in recognising and classifying faces, Perception 15 (1986) 525–536.

[17] A.J. O'Toole, K.A. Deffenbacher, D. Valentin, H. Abdi, Structural aspects of face recognition and the other-race effect, Memory and Cognition 22 (1994) 208–224.

[18] T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition, Quarterly Journal of Experimental Psychology 43A (1991) 161–204.

[19] T. Kohonen, Self-organization and Associative Memory, 3, Springer, Berlin, 1989.

[20] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, D. Valentin, Low dimensional representation of faces in higher dimensions of the face space, Journal of the Optical Society of America 10 (1993) 405–411.

[21] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, Journal of the Opthamalogical Society of America 4 (3) (1987) 519–524.

[22] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.

[23] K.A. Deffenbacher, T. Vetter, J. Johanson, A.J. O'Toole, Facial aging, attractiveness, and distinctiveness, Perception 27 (1998) 1233–1244.

[24] T. Vetter, N.F. Troje, Separation of texture and shape in images of faces, Journal of the Optical Society of America A14 (1997) 2152–2161.

[25] H.H. Bülthoff, S. Edelman, Psychophysical support for a two-dimensional view interpolation theory of object recognition, Proceedings of the National Academy of Science (1992) 60–64.

[26] I. Biederman, Recognition-by-components: a theory of human image understanding, Psychological Review 94 (1987) 115–117.

[27] S.E. Brennan, The caricature generator, Leonardo 18 (1985) 170–178.

[28] P.J. Benson, D.I. Perrett, Synthesising continuous-tone caricatures, Image and Vision Computing (1991) 123–129.

[29] P.J. Benson, D.I. Perrett, Visual processing of facial distinctiveness, Perception (1994) 75–93.

[30] J.R. Bergen, R. Hingorani, Hierarchical motion-based frame rate conversion, Technical Report, David Sarnoff Research Center, 1990.

[31] D. Beymer, A. Shashua, T. Poggio, Example-based image analysis and synthesis, Technical Report, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1993.
[32] D. Beymer, T. Poggio, Image representations for visual learning, Science 272 (1996) 1905–1909.
[33] T. Vetter, T. Poggio, Linear object classes and image synthesis from a single example image, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 733–742.
[34] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d-faces, in: Rockwood (Ed.) SIGGRAPH 99 Conference Proceedings, Addison Wesley, ACM SIGGRAPH, Los Angeles, 1999.
[35] A.J. O'Toole, T. Vetter, V. Blanz, Three-dimensional shape and two-dimensional reflectance contributions to face recognition over changes in view-point, Vision Research, 1999.
[36] J.D. Cohen, B. McWhinney, M. Flatt, J. Provost, Psyscope: a new graphic interactive environment for designing psychology experiments, Behavior Research Methods, Instruments and Computers 25 (1993) 257–271.
[37] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision 12 (1994) 43–77.
[38] J.R. Bergen, P. Anandan, K. Hanna, R. Hingorani, Hierarchical model-based motion estimation, in: Proceedings of the European Conference on Computer Vision, 1992, pp. 237–252.
[39] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Transactions on Communications 31 (1983) 532–540.