Quantification of facial expressions using high-dimensional shape transformations

Ragini Verma, Christos Davatzikos, James Loughead, Tim Indersmitten, Ranliang Hu, Christian Kohler, Raquel E. Gur, Ruben C. Gur

Abstract

We present a novel methodology for quantitative analysis of changes in facial display as the intensity of an emotion evolves from neutral to peak expression. The face is modeled as a combination of regions and their boundaries. An expression change in a face is characterized and quantified through a combination of non-rigid (elastic) deformations, i.e., expansions and contractions of these facial regions. After elastic interpolation, this yields a geometry-based high-dimensional 2D shape transformation, which is used to register regions defined on subjects (i.e., faces with expression) to those defined on the reference template face (a neutral face). This shape transformation produces a vector-valued deformation field and is used to define a scalar valued regional volumetric difference (RVD) function, which characterizes and quantifies the facial expression. The approach is applied to a standardized database consisting of single images of professional actors expressing emotions at predefined intensities. We perform a detailed analysis of the deformations generated and the regional volumetric differences computed for expressions. We were able to quantify subtle changes in expression that can distinguish the intended emotions. A model for the average expression of specific emotions was also constructed using the RVD maps. This method can be applied in basic and clinical investigations of facial affect and its neural substrates.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Expression analysis; Emotion intensity quantification; Shape transformation; Elastic deformations; Facial action coding system (FACS)

1. Introduction

Facial expression is the main means for human and other species to communicate emotions and intentions (Darwin, 1872). Such expressions provide information not only about the affective state, but also about the cognitive activity, temperament, personality and psychopathology of an individual (Darwin, 1872; Schmidt and Cohn, 2003). Facial expression analysis has been increasingly used in basic research on healthy people and in clinical investigations of neuropsychiatric disorders including affective disorders and schizophrenia. Lacking objective methods for quantifying facial change, research in this area has focused on emotion recognition capabilities of patients compared to healthy...
controls (Sakheim et al., 1982; Salem et al., 1996). Since brain disorders likely affect emotional expressions, and not only perception (Gainotti et al., 1993; Sakheim et al., 1982), it would be of value to develop methods of expression quantification and analysis that can compare how emotions are expressed by an individual suffering from a brain disorder compared to healthy controls. The problem of expression quantification is rendered extremely challenging by individual physical facial differences such as wrinkles, skin texture etc. These may confound efforts to quantify differences in expressiveness, degree of facial mobility, frequency and rate of expression, all of which could be associated with brain disorders. A major neuropsychiatric disorder characterized by deficits in emotional expressiveness is schizophrenia, where “flat affect” is a hallmark of the illness (Rimm, 1984).

Methods of expression rating currently employed are time consuming and depend on subjective judgment of raters. We propose an efficient and objective expression quantification scheme which is reproducible, robust, automated and fast. In order to develop objective measures of changes in facial expressions, it is necessary to: (1) quantify the change in expression as the emotion changes in intensity from mild to peak; (2) quantify the difference in facial expression of the same emotion between patients and healthy people; (3) construct a model of an expression using healthy people as a standard against which the expression of patients with different disorders can be tested. Quantification of fine-grained structural changes in the face is necessary to capture the subtlety of human expression changes, and this requires advanced morphometric tools. We created a model for each of four universally recognized expressions—anger, fear, sadness and happiness. However, our proposed method of expression quantification is general and can be applied to any face undergoing an expression change.

We treat the face as a combination of regions and their boundaries. A neutral face is chosen as a template and a face with expression, to be compared to and quantified against the template, is chosen as a subject. Corresponding regions are identified on each of these faces. We then compute an elastic transformation which warps the template to the subject face, mapping the corresponding boundaries to each other (Davatzikos, 2001). These regions are chosen such that all the distinctly identifiable features of the face are separated and accounted for (see Fig. 1). Such a shape transformation can be defined for each expression and at varying intensities of expression. The differences can be analyzed on a region-wise basis by comparing the corresponding shape transformations. These transformations quantify the volumetric differences among faces with expression. The resultant deformation field characterizes how the neutral face deforms to the face with expression. These deformation fields can be averaged to generate a model for each expression, which can be used for the visual presentation of the quantitative measurements of expression.

1.1. Literature survey

Facial expressions have been investigated as a tool for understanding the regulation of emotions in health and disease and investigate its neural substrates. Facial expression analysis consists of two sub-problems: expression recognition and expression quantification. Each of these requires facial modeling. Expression recognition involves classifying the expression as one of the several possible emotions (Ekman and Rosenberg, 1997). In expression quantification, the intensity of the emotion needs to be quantified on a region-wise basis, to understand how much each region contributes to an expression as well as to its intensity. Automatic expression analysis has attracted attention in computer vision literature because of its importance in clinical investigations, but the efforts have been focused on expression recognition (Bartlett et al., 1999; Black and Yacoob, 1997; Cohn et al., 1999; Essa and Pentland, 1997; Lien et al., 2000; Tian et al., 2001; Terzopoulos and Waters, 1990; Yacoob and Davis, 1994; Zhang, 1999). We propose a powerful method for expression quantification that is applicable region-wise and can be extended to automate the recognition of facial expressions.

Expression recognition methods can be categorized on the basis of the data used for the analysis (single images: Martinez, 2002; video sequences: Essa and Pentland, 1997; Lien et al., 2000; Tian et al., 1999, 2001; Yacoob and Davis, 1996) and the methodology adopted for facial modeling. Both 2D and 3D models (Karouzis et al., 2004; Terzopoulos and Waters, 1990) have been applied. Expression recognition requires modeling the face in terms of features and then coding each expression as a unique combination of values of these features. In facial feature extraction, there are mainly two types of approaches: geometric feature-based methods, which are local, and appearance-based methods, which are global. In geometric feature-based methods (Yuille et al., 1992; Zhang, 1999), a large number of landmark facial feature points are extracted, on which the features are computed. In video-based methods, these feature points are tracked (Black and Yacoob, 1997). The values of these features are used as an input to a classification system and the...
We therefore define a geometry-based shape transformation system (FACS) (Ekman and Friesen, 1978). FACS, the leading method for measuring facial movement in behavioral science, is an anatomically based coding system for recording appearance changes caused by the activation of individual facial muscles. The measurement units of FACS are action units (AUs), which are described in terms of muscular actions responsible for each facial movement. Therefore, using FACS, each facial expression can be described in terms of the activated AUs. FACS is performed manually by highly trained experts. There are several automated versions of FACS, which automate the process of studying facial action (Bartlett et al., 1999; Cohn et al., 1999; Donato et al., 1999; Essa and Pentland, 1997; Lien et al., 2000; Tian et al., 2001). The approach of all these systems is video-based. Features identified on the face (e.g., outline of lips, eyes, eyebrows, cheeks and furrows) in the first frame of the video sequence, are tracked through the video using optical flow (Yacoob and Davis, 1996) or other tracking techniques. These features can be modeled as a collection of feature points (Cohn et al., 1999; Tian et al., 2001) or deformable templates (Terzopoulos and Waters, 1990). The change in the states of these features is used to determine the action unit activated, which in turn identifies the emotion. Most of these approaches use a manual pre-registration step to initialize the features as they vary from expression to expression and also between individuals. These methods do not attempt to quantify the difference of the same emotion between two faces.

In the 3D approach, the faces are modeled using an anatomical basis of their muscle structure (Karpouzis et al., 1999; Terzopoulos and Waters, 1990) or by using multiple views (Gur et al., 2002). These models are used to study the changes in each of the facial regions (Indersmitten and Gur, 2003).

Such methods address expression recognition as a classification problem. However, for expression quantification, the faces need to be compared on a region-wise basis as each region may uniquely contribute to an expression. Faces are complex combinations of regions that deform (expand and contract) non-rigidly and elastically as the expression changes. Expression recognition methods have concentrated only on some specific facial regions and these methods fall short of characterizing the full elastic deformation between two faces. In feature-based methods, it is computationally expensive to compute and analyze features at all points, and for global methods there is insufficient information to quantify expressions on a region-wise basis. For successful quantification, we need a deformable registration capable of providing more than simply a combination of rotation, translation and scaling, but rather a technique that will map one face to the other non-rigidly (not just a combination of rotation, translation and scaling) and produce a dense deformation field which accounts for all points on the face.

Several techniques that afford quantitative morphological analysis of complex structures such as the brain and heart, which transform non-rigidly over time, have been developed in medical imaging (Christensen et al., 1997; Collins and Evans, 1996; Maintz and Viergever, 1998). These techniques spatially normalize, i.e., elastically register deformable, elastic body parts of a subject to a template. We propose to use one such technique called shape transformations to model facial deformations. Shape transformations provide a powerful way to obtain a detailed and localized structural characterization of complex objects such as the brain (Bajcsy and Kovacic, 1989; Davatzikos, 2001; Davatzikos et al., 1996). The principle of this approach is that morphological characteristics of two individuals or populations can be contrasted by comparing the corresponding transformations (Thompson, 1917). As these shape transformations map morphological characteristics of a subject to a common reference system of a template, they facilitate comparisons and pooling of data from various sources. Their highly non-linear nature makes them optimal for facial expression analysis and gives them high flexibility in adjusting the shape of one region to another. However, shape transformations that are merely intensity-based (Collins and Evans, 1996; Thompson, 1917) do not account for the underlying anatomy and the geometrical shape of facial regions, and hence cannot be applied for large uniform regions such as cheeks and foreheads. We therefore define a geometry-based shape transformation (Davatzikos, 2001; Davatzikos et al., 1996; Subsol et al., 1996) between neutral faces and faces with expression.

1.2. Our work in perspective

In our expression quantification method, the shape transformation maps a neutral face taken as a template, to a face
with expression, which is the subject. Our shape transformation is based on point correspondences determined on distinct region boundaries along with some landmark points, demarcated on the subject and template face. We compute an elastic transformation, which warps the template to the subject face, mapping the corresponding boundaries to each other and elastically interpolating the enclosed region (Davatzikos, 2001). Comparison of the properties of these shape transformations helps to quantify recognized emotions. Our specific contributions are:

- We present a method for quantifying expression changes between various intensities, e.g., mild, medium, and peak, and across individuals expressing the same or different emotion, by using high-dimensional shape transformations. The current literature on face expression recognition is able to recognize only peak emotion, i.e., when all the action units are activated, and not grades of that emotion.

- We obtain regional volumetric difference maps using the shape transformations, which quantify the deformation between two faces. We design a model for average of each expression based on these regional volumetric difference maps. Such models designed using expressions of healthy people are useful for examining individual differences and in clinical investigation, as they can be used as guides for a visual-based expression analysis. Such analysis, in conjunction with clinical methods, can be helpful for comparison of expressions of individuals with brain disorders. We can also predict the expression of a neutral face using the deformation field generated by the shape transformation. The shape transformation and the quantification map will be described in detail in Section 2.

- We can perform our analysis on single images and do not need video sequences. This makes the proposed method widely applicable. Existing standardized databases used in clinical investigations, mainly consist of single images and hence cannot be analyzed with any of the video-based techniques available in the literature, which also require changes between subsequent frames to be small for tracking to be possible.

- Our method is general and applicable to all emotions at all levels of intensity.

In Section 2, we describe the method of computing the shape transformation, and creation of the quantification maps. Experiments carried out to validate the method and demonstrate its applicability are described in Section 3. Section 3 also describes the implementation details of our method. In Section 4, we discuss the results of the experiments and identify future directions that we propose to follow.

2. Methodology of expression quantification

In this section, we first give an overview of our proposed approach. We then detail the mathematical design and computation of the shape transformation.

2.1. Approach

We choose a neutral face (face with no expression) as a template. All the other expressions of the same individual are categorized as subjects, to be analyzed against this template. A subject face is one that has expression. The expression could vary in intensity from mild to medium to peak.

Our general approach can be outlined in the following steps:

1. Identify regions on the template that characterize various facial features. These regions are demarcated by marking their boundaries (cf. Fig. 1). Some landmark points are also marked; curve segments in between are parameterized via constant-speed parameterization.

2. For each region picked on the template, identify the corresponding region on the subject.

3. Compute the elastic transformation from one face to the other, so that the demarcated regions of the face from the subject are mapped to their counterparts in the template.

The shape transform provides us with the information regarding the deformation produced by the expression change. Also, we define a regional volumetric differences function, which provides a numeric value for each pixel on the face. This value quantifies the expansion and contraction of the region on a pixel-wise basis. Additional implementation details of the method are described in Section 3.2.

2.2. Regional analysis of the face through the shape transformation

The shape transformation used in this paper is adapted from the work of Davatzikos (2001). Let \( \Omega_s \) and \( \Omega_t \) denote the subject space and template space, respectively:

\[
\Omega_s = \{ F_s | \text{Face with neutral expression} \}
\]

\[
\Omega_t = \{ F_t | \text{Face with expressions at varying intensities} \}
\]

\[
\text{mild, medium and peak}
\]

Now we define each template and subject face as a union of regions and their boundaries.

\[
\mathcal{F}_s = \bigcup_i \mathcal{R}_s^i
\]

\[
\mathcal{F}_t = \bigcup_i \mathcal{R}_t^i
\]

where \( \mathcal{R}_s^i \) and \( \mathcal{R}_t^i \) are the boundaries of the corresponding regions \( \mathcal{R}_s^i \) and \( \mathcal{R}_t^i \), respectively.

Fig. 1 shows some of the regions that have been demarcated on the face. Some landmark points (typically two to four) are also selected on the boundaries. The boundaries segments in between two consecutive landmarks are parameterized by a constant-speed parameterization, i.e., by evenly
We adapt the shape transformation computed in Davatzikos (2001) and compute the elastic shape transformation \( S \) which:

- maps the corresponding points between the boundaries of the template to their counterparts on the subject boundaries and
- warps the enclosed regions from the template to the subject, elastically.

Thus \( S \) elastically matches the boundaries and accounts for the elastic (deformable) changes not only on the boundary but also within the enclosed region. The resulting shape transformation \( S \) quantifies the shape properties of the subject face, \( F_s \), with respect to the template face, \( F_t \). Therefore, two faces with different emotion expression can be analyzed on the basis of a point-wise comparison of the shape transformations. In the method, the regions can be overlapping regions; as long as the constraints on the driving points/curves are consistent, they will be interpolated to a dense vector field. However, we choose non-overlapping regions for ease of interpretation.

The shape transformation produces a map of vectors (one vector at each pixel) called a vector field. These vectors provide the direction of movement of the pixel as the result of the deformation caused by the change in expression. Each vector is denoted by a 2-tuple \((dx, dy)\) which denotes the displacement in \( x \) and \( y \) direction that each pixel on the template undergoes, when it is transformed to the face with expression.

On adding the vector \((dx, dy)\) to the position of the template \((x, y)\), we get the deformation \((x + dx, y + dy)\) which the pixel on the template undergoes. The position to which the pixel on the neutral template moves to as a result of the action of this vector denotes the deformed position in the subject after undergoing an expression change. This is known as the deformation field. Additional details of the method can be found in Davatzikos (2001) and Davatzikos et al. (1996).

We obtain two quantities from the shape transformation that we use for our analysis:

1. The scalar field of values of the regional volumetric difference function (RVDF), which is evaluated at each pixel on the face. We define the RVDF as:
   
   \[
   \text{RVDF}(s) = \det(V(\mathbf{s})) = \text{determinant of the Jacobian of } S
   \]

   evaluated at each point \( s \) on the subject

   The RVDF value quantifies the volumetric difference (variability in expansion and contraction) between regions. The map containing the RVDF values for each pixel of the face is called the RVD map of the face. Various inferences may be drawn from the values of the RVD function. If \( \text{RVDF}(s) \geq \text{RVDF}(u) \) for the same region on two faces which have the same expression, at the same intensity, then it quantifies that the region in face 1 has deformed (expanded or contracted) more than the region in face 2, relative to their respective neutral states. It, therefore, quantifies the variability in expressing the same emotion across individuals. If this is on the same face and same region, it indicates and quantifies the change in expression. Fig. 2(c) shows a color visualization of an RVD map.

2. The vector displacement fields of the deformation. These characterize the direction and degree of movement of each pixel of the face during the course of an expression change. These vectors can be used to quantify temporal changes in an expression. Fig. 2(d) shows the deformation of each pixel of the template face 2(a) as a result of the expression change from 2(a) to 2(b).

Fig. 2. Information obtained from the shape transformation: (a) template face, (b) subject face with expression, (c) intensity normalized RVD map, (d) vector deformation field and (e) color map for visualization of RVD maps.
In this paper, the quantification measure will be the RVD map of the shape transformation of a face with respect to a template. However the vector deformation field can be used to provide additional information about the direction of movement of the regions. Implementation issues of the algorithm will be discussed in detail in Section 3.2.

3. Experiments and results

We have carried out experiments in facial expression analysis to highlight the features of our approach, by applying our method to a database created for the purpose of studying expression related disorders (Gur et al., 2002). The experiments are not meant to be exhaustive, but focus on the applicability of our approach in the investigation of affect processing. They demonstrate the flexibility and power of our method to aid in the diagnosis of disorders of emotion processing.

3.1. Database used in study

In order to conduct expression analysis research, investigators have collected databases of emotion, which can contain either posed or evoked expressions (Tian et al., 2001). Posed databases have been acquired using posers (actors and non-actors) (Ekman and Rosenberg, 1997). However, there is evidence that posed emotions are less accurately perceived than evoked expressions (Gur et al., 2002). Therefore, we use only evoked expressions for our analysis. Such a database with posed and the more difficult to acquire evoked expressions, has been described and validated (Gur et al., 2002), and we selected a subset of the available stimuli.

Our methodology is tested in this paper using a subset of the database created for investigations of schizophrenia (Kohler et al., 2004). For this database, a group of actors were photographed while portraying posed and evoked expressions of the emotions of happiness, sadness, fear, disgust and anger. Each of the actors was initially photographed with a neutral face. They were then guided by professional theatre directors through enactments of each emotion using both posed (mechanical, the “English” acting method) and evoked (the Stanislavski or “Russian” acting method) procedures. To generate evoked expressions, the actors were instructed to remember a past experience which had generated the same emotion. Expressions were photographed at three predetermined levels of emotion intensity: mild, medium and peak. The images were captured using a Nikon N9035 mm SLR camera (with Kodak Ektachrome 320T film, exposed at 1/80 s and f/5.6). The images from this camera were scanned into the computer off-line. The details of the method can be found in Gur et al. (2002). The images are of size 1024 × 768. The computational algorithms are independent of the size of the image. The images were acquired in color, however we do not require color for our algorithm.

For this study, we selected evoked images from 30 male and 30 female Caucasian actors between ages of 18 and 72 years (average 39) portraying expressions of happiness, sadness, fear and anger. Disgust was not included due to poor recognition in the validation of the full database (Gur et al., 2002; Kohler et al., 2004). The intensity of expressions ranged from mild to peak. The validity and intensity of the emotions expressed were established by controlled rating procedures. The images were evaluated by expert raters (N = 8) to ensure that the ease of recognition of the target emotion and intensity level were comparable for each image included in the subset. The raters were required to evaluate the intensity of each expression as mild, medium or peak, as well as classify each facial expression as being of anger, fear, happiness or sadness. Percentage values were calculated for each face with expression with respect to the intensity and the expression category. All faces whose expression was identified as being of the correct category of emotion by more than 60% of the raters, were chosen for analysis using our method. Those with higher than 60% score on intensity were chosen as peak expressions for analysis. The thresholds were kept low so that greater variability is incorporated. These controls are essential as evaluation of an expression is valid only if the intended expression is perceived correctly by a large proportion of healthy research participants. The aim of this rating was to determine a subset of facial expressions to be used in our analysis, the responses from which can be validated by the raters. Validation of the database has been performed in Gur et al. (2002), Indersmitten and Gur (2003), Kohler et al. (2004) and Palermo and Coltheart (2004).

3.2. Implementation issues

In this section we present some details of the algorithm which play an important role in the implementation and the results produced for analysis.

3.2.1. Choice of template

In order to be able to compare the difference in expression between two individuals, the RVD maps are computed using the same neutral face as the template face. Any face can be used as a template. However faces with neutral expression provide a natural base against which any change of expression can be compared. Hence we use a neutral face as a template.

In addition, the face must be completely facing the camera, with no sideways or forward/backward head tilt. Optimally, the size of the head should be average, i.e., occupying two-thirds of the image and should fit completely in the frame so that the outer boundary can be clearly marked. The template is the neutral face of a person who is in the average age of the population under study. In our case we have performed the analysis on several templates in the age range of 30–40, the average age of our participants being 39. We have experimentally found that the same template can be used for both genders. We did not have enough data to test this on faces of different ethnic groups, but we propose to incorporate this once additional data is available in the future.
As the neutral template face may be of a different individual, than the subject face with expression, the indications of expansion and contraction may be due to the fact that the size of the face used as a template differs highly from that of the subjects. To obtain consistent results, we normalize the RVD maps obtained for the subjects, by using the RVD map of the neutral face of the individual chosen as subject. This is explained in detail in the experiments associated with Figs. 7 and 8.

3.2.3. Choice of landmark points

The choice of regions does affect accuracy. In densely sampled regions, the estimates of the deformation will be more accurate, although this depends on how complex the deformation itself is. However, dense sampling of regions also increases the human error possible in outlining these regions. We chose the regions so that they cover the whole face and are mutually exclusive and non-overlapping. They were carefully drawn so that they correspond to natural lines on the face, or to clearly identifiable markers on the face, such as the hairline, outline of lips, nose, eyebrows and eyes, etc. The aim of this choice was to facilitate repeatability and to avoid, as much as possible, inter-user variability and the human error involved in identifying and marking these outlines and landmark points. We selected non-overlapping regions for facial expression analysis, in order to best interpret regional changes due to the change in facial expression. We found that the expansion and contraction of overlapping regions were not meaningful, unless one region was completely contained in the other. In this case, the smaller regions are a constraint on the larger region. For example, lips within the outline of the face, shows how a region within the face changes when the face changes. However, if the lips and the nose intersect, it is difficult to interpret the individual changes. Although smaller regions may allow better quantification, for example, dividing the brow would help identify changes in the inner brow and the outer brow separately, however in the absence of natural subdivision of the brow, we will increase the human error in identifying these regions, making the analysis of expansion and contraction in these regions difficult. Also the regions are not normalized with respect to the size of the neighboring regions, their relative size differences do not weight one region more than the other.

The regions defined on the templates are fixed. It is important to identify the regions manually as these region demarcations may vary between individuals and also from one expression to another. The ability to manually identify the regions also provides the user with flexibility to alter the regions picked.

3.2.3. Choice of landmark points

Two to four landmark points need to be marked on the boundaries outlining the regions. These serve as seed points to establish correspondence between the outlines on the two faces. Curves between these landmark points are parameterized using constant-speed parameterization. This parameterization is used to establish full correspondence between the outlines. We have experimentally found that up to four landmark points are sufficient for the size of the contours on the face. In case of larger contours, larger number of points may be marked, however this increases the chance of human error. Similarly for smaller size contours, we use lesser number of landmarks.

3.3. Experimental analysis

We apply our approach to the database described above. Fig. 1 shows some of the regions that have been demarcated on the faces. Our method is flexible as these regions may be altered depending on the analysis requested. The algorithm was tested on images from other databases acquired under different settings and different lighting conditions. It performed with accuracy as long as boundaries could be clearly demarcated.

In Fig. 2, we show the information that is produced by the shape transformation. Fig. 2(a) shows the template neutral face and Fig. 2(b) shows the corresponding subject face expressing fear. We then compute the shape transformation that elastically warps the regions demarcated on the template to the corresponding regions identified on the subject (see Fig. 1 for regions). A positive RVD value indicates an expansion and a negative RVD value indicates a contraction. These are the values used in the analyses. However, these RVD values are normalized to a specific range for visualization of the expression changes in the form of a color map. In our case, we choose the range to be 0-90, as it provides the best demarcation. In doing so, the base value of 0, indicating no change, is shifted to 30. The range for displaying the color map can be changed by the user. Fig. 2(c) shows the color-coded RVD map of RVD values computed at each pixel of the face, the color map for which is in Fig. 2(e). After normalization, an increase in RVD values from the template to the subject indicates the expansion and a decrease indicates the contraction of the corresponding region in the subject. These maps are computed at varying intensities of the same emotion, to study expression changes. In general, darker blues indicate contraction and yellow to red depicts increasing expansion. Fig. 2(d) depicts the deformation field indicating the pixel movements due to the expression change from 2(a) to 2(b). In order to create 2(d), we represented the pixels of the template face as a grid of the same size. Then to each position of the grid, the displacement of the vector field produced as a result of the shape transformation, was applied. This produces the deformed grid shown in Fig. 2(d). In this paper, we will use RVD maps (as shown in Fig. 2(c)) for quantification as these provide a numeric value of the changes at each pixel as a result of the expression change.

3.3.1. Regional changes and movements

The RVD maps can be used to determine movement of facial regions. This is achieved by studying the region as a
combination of expansion and contraction of several regions. In Fig. 2(c), the forehead contracts and the upper lid and the region between the eyes expand. This indicates that the eyebrows have raised. The mouth and upper lip expand. This in conjunction with the fact that the chin and cheeks expand slightly and lower face expands, indicating a jaw drop. The actual expansion and contraction of these regions and the direction of their movement can be verified with the actual changes in face in 2(a) and (b). Other changes in the face can be determined as a combination of expansion and contraction of several regions, as will be explained for Figs. 3–5.

We now describe the detailed experiments performed to test various aspects of the approach.

3.3.2. Quantification of expressions

In Figs. 3–5 we show the RVD maps for the expressions—happiness, sadness and anger. The RVD values for the expression of fear has been shown and discussed in Fig. 2. The RVD maps are color-coded on the basis of the color map shown in Fig. 2(e). The regions identified on the faces have been shown in Fig. 1.

In Fig. 3, we quantify three intensities of happiness. Part (a) shows the face with neutral expression and is the template and parts (b)–(d) are the three subjects representing three intensities of happiness. The RVD maps shown in Fig. 3(e)–(g) depict the pixel-wise RVD values which indicate expansion and contraction. The color map is the same as in Fig. 2(e). In this case, the eyes contract, the mouth expands and the cheeks contract indicating a sideways expansion of the mouth. The lower lids expand depicting a cheek raise. The forehead shows no change and neither does the region between the eyes and the eyebrows. The contraction is indicated by a decrease in RVD values from the template to the subject images. The expansion of the regions is indicated by the increase in RVD values from the template to the subject images.

In Fig. 4, the same analysis is carried out for the expression of sadness. Fig. 4(a) shows the template neutral face and parts (b)–(d) show the subject faces showing various intensities of sadness. The eyes contract, the region between the eyes and the eyebrows expand and the forehead contracts, indicating an eyebrow raise, the upper lip expands, the mouth expands and the cheeks contract. The chin expands and the whole face contracts, indicating a sideways expansion of the lips and the chin.

In Fig. 5, the same analysis is carried out for the expression of anger. Fig. 5(a) shows a neutral face and parts (b)–(d) show the varying intensities of anger. The second row shows the color-coded RVD maps for each of the four images. In this case, images 5(b)–(d) are the subjects, to which the template in 5(a) is elastically warped. The RVD values of pixels in the corresponding regions in the faces in Fig. 5 indicate that the eyes contract, the forehead expands and the regions between the eyes contract, indicating a brow lowering. The cheeks contract and the lower lids expand, indicating a cheek raise. The continuous increase in RVD values of the mouth from (e) to (g) shows that the region continues to expand. There is a continuous decrease in RVD values of the eyes indicating a contraction. This can be validated against the actual images of the expression in 5(b)–(d).

The comparison with FACS reveals that our method is able to identify changes in regions which are also identified as being activated by action units by FACS raters (Kohler et
3.3.3. Stages of deformation leading to the peak

While computing the shape transformation for each of the expressions, we also obtained the deformation fields between the neutral template and the faces with expressions. By studying the deformation fields, we can analyze how the face deforms in stages from the neutral to the peak. This can be seen in Fig. 6, which shows some of the intermediate stages (frames) of a face transforming from a neutral face to an angry face. These intermediate facial expressions are generated by interpolating the deformation fields produced by the shape transformation mapping the neutral face to three intensities of expressions of the same emotion. The interpolation is computed to emulate 25 frames/s, i.e., the increments in the displacement are generated such that the change in expression is at video rate. The intermediate expressions can be used for normalization of expressions and also used for predicting the expression of a person.

3.3.4. Effect of choice of initial template

The template is used as a measurement unit on which all the experiments are based. In order for the RVD maps to be comparable for analysis across individuals, it is important to show that the choice of the template does not affect the results. This means that irrespective of the neutral face chosen, the RVD map of the change of expression from neutral to peak shows the same region-wise changes for the same expression. This has been discussed mathematically in Davatzikos et al. (1996) for shape transformations used for the brain. We have conducted two sets of experiments (cf. Figs. 7 and 8) to demonstrate the practical feasibility of the claim. Fig. 7(b) and (e) show the peak anger expressions for neutral faces in 7(a) and (d). The RVD map of the shape transformation from 7(a) and (b) is shown in 7(c) and from 7(d) to (e) is in 7(f). The RVD maps show that similar patterns of expansion and contraction are obtained even when we use completely different neutral faces as in 7(a) and (d). The values of the measurements of expansion and contraction change, as the template changes, since the area of a region would have different values in different metric systems. However, relative magnitudes should not change, to the extent that there is registration error between the template and the subject. This means that the relative expansion and contraction of the regions do not change. This is demonstrated in Fig. 7. The forehead, mouth and lower eyelids expand. The eyes, upper eyelids and area between the eyes contract. The expansion of the forehead and contraction of the upper eye region, indicate a lowering of the eyebrows. The triangle between the eyes contract, indicating a knitting of brows. The cheeks expand on one side and contract on the other, indicating a facial contortion. The change in the cheeks is partially due to the turn of the head, however the difference in deformation between the cheeks is evident even in the presence of a head turn. All these deformations are typical of an angry face, and are highlighted even with a change in template.

The choice of template presents another issue, that if the template used is of an individual different from those chosen as subjects, the expansion and contraction evident in the RVD maps could be due to the difference in the sizes of the facial features between the subject and the template. In Fig. 8, the template 8(a) and the subject 8(b) are of two different individuals. The RVD map of (b) with respect to (a) is shown in (c). However, the expansion and contraction shown in (c) can be due to the difference in size of these regions between the

![Fig. 6. Stages of deformation from the neutral to the peak expression of anger.](image)

![Fig. 7. Effect of change of template.](image)
individuals, as can be seen in the neutral face of the subject, shown in (d). In order to correct the RVD map in (c) for these size differences of the facial features, we compute the RVD map of (d) with respect to (a), as shown in (e). We then use the shape transformations corresponding to the RVD maps to obtain the corrected shape transformation and hence the RVD maps. The corrected RVD map is shown in (f). In (f), we can see that there is a contraction in eyes and the inner eyebrows and the region between the eyebrows. Although there is an expansion in the region above the eyes, as seen in (c) and (e), this was due to the difference in size between the two faces and it is eliminated in the neutral in (f), except in the area immediately between but not above the eyebrows. There is an expansion in the mouth even after the correction.

As has been explained in Section 3.2, we use a neutral face as a template for all our analyses. The face is of an individual who is approximately the average age of the group under study (39 years for our database). The template should be directly facing the camera.

3.3.5. Average expressions

Our approach can be used to model the average expression for each emotion. This can be obtained by averaging (1) the deformation vector fields and (2) the RVD maps generated by the shape transformation, which are the two forms of information that we obtain from the shape transformation, as explained in Section 2. We take all the images of a particular emotion and intensity and compute the shape transformations that map each of these to a template face. The same template (neutral) face is used for the whole analysis. It may be noted that we can use all faces for independent analysis, however for computing the averages, we used the images that were perceived to be of peak expression by the raters (see Section 3.1). This significantly reduces the number of images that we have available for averaging. The vector deformation fields generated by these shape transformation are averaged using Procrustes techniques in order to produce an average deformation for each of the expressions. That is, we average the $x$ and $y$ displacements of each of the pixels and produce values which indicate the average deformation of the pixels on a template. The average vector field is applied to two neutral faces (9(a)) and (c)). Fig. 9(b) shows the average angry face obtained, when the average deformation corresponding to anger is applied to the neutral face in 9(a). 9(c) and (d) show a neutral template and the corresponding average happy face, respectively. The black regions show the regions of expansion of the neutral face, either a muscle stretch or the appearance of features like teeth, the evidence of which was absent in the neutral face, and whose appearance produces a black patch.

The RVD maps for each of these shape transformations are averaged to produce the average RVD values for each expression. The average RVD maps for the emotions of happiness, sadness, anger and fear in men are shown in Fig. 10. The average RVD maps indicate the regions of expansion and contraction for each expression—happiness (10(a)): the mouth and the region between the eyes expands and cheeks contract indicating sideways stretching, eyes contract, the lower lid expands and the cheeks contract indicating raising of the cheek); sadness (10(b)): uneven expansion and contraction of the face indicating contortion, eyes contract and forehead contracts indicating a lowering of the brow while chin and mouth expand, and there is sideways stretching); anger (10(c)): the mouth opens, the region between the eyes and the brows contract and the eyes contract, the region...
between eyes and the nose contract indicating wrinkling of
the nose); fear (10(d): the mouth opens wide, the eyes and
the regions above them expands, the overall face expands
indicating a jaw drop). The average deformed faces give a good
indication of the regions of the face which are undergoing a
change. These are some of regions in which change is also
indicated by FACS (Ekman and Friesen, 1978; Kohler et al.,
2004). As the number of samples is small and there is a large
variability in the expressions, some subtle differences are not
evident in the averages. Additional information can be ob-
tained through the individual’s RVD map (as in Figs. 2–6 )
and the direction and extent of motion of each facial region
can be obtained through the vector field of deformation (cf.
Fig. 2).

3.3.6. Point-wise t-test on the RVDF images to study
regions of significant expansion and contraction

We have carried out point-wise t-test between men
and women, for the expression of anger, to identify the regions
which show significant deformation difference between the
two groups. We generated RVD maps for facial expressions
of anger of 15 men and 10 women. Point-wise t-test was
carried out on these RVD maps. The t-test was in the direction
of women to men because the RVD maps of the individual
faces, as well as the group averages, indicated that there was
larger deformation (expansion and contraction) in women
than in men. The results for these are shown in Fig. 11, where
the regions marked with white are the regions with low p-
value (<0.01). As seen in women relative to men, there is
a greater contraction of the forehead, nose and the region
between the eyes, and a greater expansion of the mouth, lower
eyelid, upper lip and cheeks. The low p-values indicate that
the significant difference in the deformation between men
and women is restricted to the forehead, mouth, upper lip
and lower eyelid regions.

As the number of images for peak expression of sad-
ness and fear were too few after rating by healthy observers
to make comparison meaningful, no comparison was per-
formed. Also the variability is very high between these im-
taghes of sadness and fear and a statistical comparison will
not be meaningful here either. The experiments we have per-
formed are indicative of the tests which can be conducted on
these RVD maps when the number of images available for
each gender and expression is large.

3.3.7. PCA to show that expressions can be
distinguished

We have carried out PCA (Duda et al., 2001) on the RVD
maps of the images of the four expressions for all the male
actors, accepted after they were rated as peak expression
of the emotion that was enacted. We have used 11 images for
happy expression, 6 for sad, 8 for fear and 15 for anger. Each
of these groups includes one image which was rated to be
of low intensity by the raters. This image was not used for
training, but only for testing. We cannot combine the ex-
pressions for the male and female actors, as there is a large
variability between the groups. We formed four classes, one
pertaining to each emotion, by training on the RVD maps
of the images available for that expression (except the one
rated low) and used the leave-one-out paradigm to check all
the images (including the one rated low intensity), using the
Mahalanobis distance. All the 11 happy faces were classified
correctly. The one with low intensity was classified correctly;
however the difference in voting for classification to the class
of anger and happy was very small. Three out of the six sad
images were correctly classified, but of the remaining three,
We have also shown the applicability of our method for the gender-wise analysis of facial expression data (in Fig. 11). We have been able to identify areas that have significant difference in the deformation between men and women for the expression of anger. The analysis has been limited by the size of the database and the facial images available. With a larger number of samples, it will be possible to perform such an analysis for the other expressions also. We propose to do this in the future. In addition, we will be able to identify regions that are more deformed in one gender compared to the other. This is important for clinical investigations of disorders such as schizophrenia and affective illness, where gender differences are salient (Gur et al., 1992, 2004; Stitziel et al., 1992). Also the quantification of a person’s expression may indicate how the person responds to medication or other factors. In the future, we propose to carry out quantitative and statistical analysis to identify differences between evoked and posed expressions and the variability of expressions across ethnicities and gender. Also, we intend to study the differences in expressibility of the left and right side of the face (Ingersmitten and Gur, 2003; Sackeim et al., 1978). In addition, we will investigate the differences in classification with PCA when faces of women are used for analysis, as opposed to male facial expressions.

Our technique is able to capture very subtle differences in facial expression change. This was evident in the fact that PCA analysis on the four expressions was able to distinguish some of the expressions correctly, even with small sample sizes and large inter-sample variability. For the expressions of sadness and fear, the error rate was high as the sample size was too small and there was a large variability in the expression. As was seen in the case of comparison of anger and happiness, the results improved considerably with the increase in sample size. Based on the results of the PCA, through which we were able to classify some emotions, we propose to extend our method further so that it can be used for recognition/classification of expressions as well. For classification, the measurements of expansion and contraction will be fed to either a voxel-wise analysis or to a classification technique with appropriate normalization for magnitude differences. We are currently investigating this with various alternative learning approaches like support vector machines and methods which require smaller datasets for training. Augmenting the database with more images will also give us a better classification of a test expression. Using these learning approaches, we propose to develop a method of face recognition and quantification, which will be validated against clinical expression rating using FACS.

The analysis in this paper has been carried out on single image captured at varying intensities. In order to capture all the motion dependant subtleties of the emotional expression process (like eye blink or involuntary tics), we propose to extend our technique into a fully automated method applicable to video sequences of expression formation (automated after the regions have been identified in the first frame).
In addition, there is a difference in the degree of deformation from the neutral to the peak, between individuals. We propose to analyze and compare the rate of expression change across individuals by applying our approach to videos of the expression change. The analysis of subsequent frames of the video sequence will help generate a temporal trajectory of the expression change. This will also be used in gender-based investigations. In the future we propose a multi-faceted extension and application of our approach to larger datasets with adequate representation from both genders and different ethnicities. We propose to extend our approach to apply to videos of expression change, in order to obtain a temporal quantification. We will perform extensive statistical and computational analysis and validation of the RVD maps and deformation fields on large datasets.

References


Yacoob Y, Davis L. Recognizing human facial expression. Center for Automation Research, University of Maryland; 1994.

