ELSEVIER

Journal of Neuroscience Methods 168 (2008) 224–238

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Automated video-based facial expression analysis of neuropsychiatric disorders

Peng Wang [a], Frederick Barrett [b], Elizabeth Martin [b], Marina Milonova [c], Raquel E. Gur [d,e,f], Ruben C. Gur [b,e], Christian Kohler [f], Ragini Verma [a,*]

[a] *Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, 3600 Market, Suite 380, Philadelphia, PA 19104, USA*
[b] *Brain Behavior Center, Department of Psychiatry, University of Pennsylvania Medical Center, Hospital of the University of Pennsylvania, 3400 Spruce Street, 10th Floor Gates Building Philadelphia, PA 19104, USA*
[c] *School of Arts and Sciences, University of Pennsylvania Medical Center, Hospital of the University of Pennsylvania, 3400 Spruce Street, 10th Floor Gates Building Philadelphia, PA 19104, USA*
[d] *Department of Psychiatry, University of Pennsylvania Medical Center, Hospital of the University of Pennsylvania, 3400 Spruce Street, 10th Floor Gates Building Philadelphia, PA 19104, USA*
[e] *Department of Neurology & Radiology, University of Pennsylvania Medical Center, Hospital of the University of Pennsylvania, 3400 Spruce Street, 10th Floor Gates Building Philadelphia, PA 19104, USA*
[f] *Neuropsychiatry Section, University of Pennsylvania Medical Center, Hospital of the University of Pennsylvania, 3400 Spruce Street, 10th Floor Gates Building Philadelphia, PA 19104, USA*

Received 16 July 2007; received in revised form 20 September 2007; accepted 20 September 2007

Abstract

Deficits in emotional expression are prominent in several neuropsychiatric disorders, including schizophrenia. Available clinical facial expression evaluations provide subjective and qualitative measurements, which are based on static 2D images that do not capture the temporal dynamics and subtleties of expression changes. Therefore, there is a need for automated, objective and quantitative measurements of facial expressions captured using videos. This paper presents a computational framework that creates probabilistic expression profiles for video data and can potentially help to automatically quantify emotional expression differences between patients with neuropsychiatric disorders and healthy controls. Our method automatically detects and tracks facial landmarks in videos, and then extracts geometric features to characterize facial expression changes. To analyze temporal facial expression changes, we employ probabilistic classifiers that analyze facial expressions in individual frames, and then propagate the probabilities throughout the video to capture the temporal characteristics of facial expressions. The applications of our method to healthy controls and case studies of patients with schizophrenia and Asperger's syndrome demonstrate the capability of the video-based expression analysis method in capturing subtleties of facial expression. Such results can pave the way for a video-based method for quantitative analysis of facial expressions in clinical research of disorders that cause affective deficits.

Keywords: Facial expression; Video analysis; Schizophrenia; Affective deficits; Pattern classification

1. Introduction

Facial expressions have been used in clinical research to study deficits in emotional expression and social cognition in neuropsychiatric disorders (Morrison et al., 1988; Berenbaum and Oltmann, 1992; Kring et al., 1994; Mandal et al., 1998). Specifically, patients with schizophrenia often demonstrate two types of impairments in facial expressions: "flat affect" and "inappropriate affect"(Gur et al., 2006). However, most of the current clinical methods, such as the scale for assessment of negative symptoms (SANS (Andreasen, 1984)), are based on subjective ratings and therefore provide qualitative measurements. They also require extensive human expertise and interpretation. This underlines the need for automated, objective and quantitative measurements of facial expression. We previously reported a method for quantifying facial expressions based on static images

* Corresponding author. Tel.: +1 215 662 7471; fax: +1 215 614 0266.
E-mail addresses: peng.wang@uphs.upenn.edu (P. Wang), fbarrett@bbl.med.upenn.edu (F. Barrett), raquel@bbl.med.upenn.edu (R.E. Gur), gur@bbl.med.upenn.edu (R.C. Gur), kohler@bbl.med.upenn.edu (C. Kohler), Ragini.Verma@uphs.upenn.edu (R. Verma).

0165-0270/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2007.09.030

(Verma et al., 2005; Alvino et al., 2007). However, temporal information plays an important role in understanding facial expressions because emotion processing is naturally a temporal procedure. Therefore, facial expression analysis from static 2D images lacks the temporal component, which is essential to capture subtle changes in expression. Although video-based acquisition has been employed in the examination of facial emotion expression (Kring and Sloan, 2007), currently there is no objective and automated way of facial expression analysis for the study of neuropsychiatric disorders, particularly due to the large volume of data that makes human analysis prohibitive. In this paper, we present a computational framework that uses videos to automatically analyze facial expressions and can be used to characterize impairments in such neuropsychiatric disorders.

The merits of automated facial expression analysis (AFEA) are two-fold: using it can avoid intensive human efforts, and can provide unified quantitative results. There are already many AFEA methods being presented in both clinical and computer vision communities (Gaebel and Wölwer, 1992; Hellewell et al., 1994; Schneider et al., 1990; Pantic and Rothkrantz, 2000; Fasel and Luettin, 2003; Tian et al., 2005). Most of the current AFEA methods focus on the recognition of posed facial expressions with application to human computer interaction tasks, and only a few of them have been applied to clinical studies (Verma et al., 2005; Alvino et al., 2007). In previous work on expression quantification (Verma et al., 2005; Alvino et al., 2007), the expression changes were modeled using elastic shape transformations between the face of a neutral template and the corresponding emotionally expressive face. Again, as most of the current AFEA methods, this approach is based on static 2D images without any temporal component.

In this paper, we present a computational framework that uses videos for the analysis of facial expression changes. This framework explores the dynamic information that is not captured by static images during emotion processing, and provides computationally robust results with potential clinical applicability. Broadly, our computational framework includes the detection of faces in videos, which are then tracked through the video, incorporating shape changes. Based on tracking results, features are extracted from faces to create probabilistic facial expression classifiers. The probabilistic outputs of facial expression classifiers are propagated throughout the video, to create probabilistic profiles of facial expressions. Probabilistic profiles contain dynamic information of facial expressions, based on which quantitative measures are extracted for analysis. As an application of this framework, such quantitative measurements for facial expressions could be correlated with clinical ratings to study the facial expression deficits in neuropsychiatric disorders. To our knowledge, the presented framework is the first to apply video-based automated facial expression analysis in neuropsychiatric research.

The rest of the paper is organized as follows: In Section 2, previous related work is reviewed. Our computational framework is presented in Section 3. The experimental results are provided in Section 4. We discuss the results and conclude in Section 5.

2. Related work

2.1. Clinical facial expression analysis

In clinical research, facial expressions are usually studied using 2D images that are described in two ways: either as a combination of muscular movements or as universal global expressions. The Facial Action Coding System (FACS) has been developed to describe facial expressions using a combination of action units (AU) (Ekman and Friesen, 1978). Each action unit corresponds to a specific muscular activity that produces momentary changes in facial appearance. The global facial expression handles the expressions as a whole without breaking up into AUs. The most commonly studied universal expressions include happiness, sadness, anger and fear, which are referred to as universal emotions. While most of the work has been on static 2D images, the Facial Expression Coding System (FACES) (Kring and Sloan, 2007) has been designed to analyze videos of facial expressions, in terms of the duration, content and valence of universal expressions. However, these methods need intensive human intervention to rate the images and videos of facial expressions. Such rating methods are prone to subjective errors, and have difficulties in providing unified quantitative measurements. There is need for automated, objective and quantitative measurements of facial expressions.

2.2. Automated facial expression analysis

Automated facial expression analysis (AFEA) allows computers to automatically provide quantitative measurements of facial expressions. Several factors have contributed towards making AFEA challenging. First, facial expressions vary across individuals due to the differences of the facial appearance, degree of facial plasticity, morphology and frequency of facial expressions (Tian et al., 2005). Second, it is difficult to quantify the intensity of facial expressions, especially when they are subtle. In FACS, a set of rules are used to score AU intensities (Ekman and Friesen, 1978). However, such criteria are subjective to the rater; therefore, it is difficult to extend the measurements to computer-based facial expression analysis, although there have been methods to automatically detect AUs (Pantic and Rothkrantz, 2000). Many AFEA methods have been developed recently to address such problems (Pantic and Rothkrantz, 2000; Fasel and Luettin, 2003). These methods can be categorized as image-based, video-based and 3D surface-based, according to the data used. Below we summarize some typical image-based and video-based facial expression analysis methods.

2.2.1. Image-based methods

Image-based methods extract features from individual images, and create classifiers to recognize facial expressions. Commonly used are geometric features, texture features, and their combinations. Geometric features represent the spatial information of facial expressions, such as positions of eyes and mouth, the distance between two eyebrows. The geometric features used by Tian et al. (2001) are grouped into permanent

and transient. The permanent features include positions of lips, eyes, brows, cheeks and furrows that have become permanent with age. The transient features include facial lines and furrows that are not present at rest but appear with facial expressions (Tian et al., 2001). The texture features include image intensity (Bartlett et al., 1999), image difference (Fasel and Luettin, 2000), edge (Tian et al., 2001; Lien et al., 1998), and wavelets (Lyons et al., 1999; Littlewort et al., 2006). To recognize subtle facial expressions, both features computed by using principal components and image difference usually require precise alignment, not readily feasible in real world applications. The edge features are often used to describe furrows and lines caused by facial expressions, but are difficult to detect for subtle expressions. Gabor wavelets calculated from facial appearance describe both spatial and frequency information for image analysis, and have shown capability in face recognition and facial feature tracking (Wiskott et al., 1997), as well as facial expression recognition (Lyons et al., 1999; Littlewort et al., 2006). Furthermore, experiments (Bartlett et al., 1999; Zhang et al., 1998) demonstrate that the fusion of appearance features (Gabor wavelets or PCA features) and geometric features can provide better accuracy than using either of them alone. To recognize facial expressions, extracted features are input to facial expression classifiers, such as the Nearest Neighbor classifier (Fasel and Luettin, 2000), Neural Networks (Tian et al., 2001), SVM (Littlewort et al., 2006), Bayesian Networks (Cohen et al., 2003a), and AdaBoost classifier (Littlewort et al., 2006; Wang et al., 2004).

2.2.2. *Video-based methods*

It is claimed that temporal information can improve the accuracy of facial expression recognition over using static images (Cohen et al., 2003b). However, only few video-based methods have been developed to use the temporal information of facial expressions (Littlewort et al., 2006; Cohen et al., 2003b; Yacoob and Davis, 1996; Yeasin et al., 2004; Lien et al., 2000; Chang et al., 2004). In the work of Yacoob et al. (Yacoob and Davis, 1996), each facial expression is divided into three segments: the beginning, the apex and the ending. Rules are defined to determine the temporal model of facial expressions. Such rules are ad hoc, and cannot be generalized to complex environments. In the work of Cohen et al. (2003b), facial expressions are represented in terms of magnitudes of predefined facial motions, so called Motion-Units (MU). A Tree-Augmented-Naive Bayes classifier is first used to recognize facial expressions at the level of static images, and then a multi-level Hidden Markov Model (HMM) structure is applied to recognize facial expressions at the level of video sequences. Yeasin et al. (2004) also present a two-stage approach to recognize facial expression and its intensity in video using optical flow. Another example of using HMM for facial expression analysis can be found in Lien et al. (2000). Besides HMM, the sampling-based probabilistic tracking methods, known as "particle filtering" or "Condensation", are also used to track facial expression in video sequence (Chang et al., 2004, 2005). Manifold subspace features have been applied for video-based facial expression analysis. However, in their methods, a separate manifold is built for each subject, and the subjects appear in both training and testing sequences. It is unclear that such specifically learned manifolds can be generalized to different subjects, since it is observed that their manifolds show different structures (Chang et al., 2004).

An important facet in video-based methods is how to maintain accurate tracking throughout the video sequence. A wide range of deformable models, such as muscle-based models (Ohta et al., 1998), a 3D wireframe model (Cohen et al., 2003b), a facial mesh model (Essa and Pentland, 1995, 1997), a potential net model (Kimura and Yachida, 1997), ASM (Lanitis et al., 1997), and a geometry-based shape model (Verma et al., 2005; Davatzikos, 2001), are used to track facial features in video. Although it has been demonstrated that a sophisticated deformable model can improve facial tracking accuracy, thereby improving facial expression analysis accuracy (Wen and Huang, 2003), there are no comprehensive experiments showing which deformable model is superior to the others.

In summary, video-based methods can capture subtle changes and temporal trends of facial expression, which cannot be achieved by static image-based methods. Due to the large amount of data in videos, a fully automated method for analysis is required. In the following sections, we first present a framework that is able to quantify the facial expression changes in video, and then describe normative data on healthy people, and finally apply the method in two illustrative patients to examine its potential for research in neuropsychiatric disorders.

3. Methods

This section presents our computational framework for facial expression analysis using video data. We provide an overview of the framework in Section 3.1, with further details in subsequent subsections.

3.1. *A framework of quantitative facial expression analysis in video*

Our framework for automated facial expression analysis of video data comprises the following components: (1) detecting landmarks that define the facial shape, and tracking landmarks and hence the facial changes due to expressions; (2) feature extraction based on these landmarks; (3) creation of classifiers based on extracted features, and probabilistic classification at each frame of the video sequence; and (4) probabilistic propagation of facial expressions throughout the video. We first apply a face detector and a landmark detector to automatically locate landmarks in videos. Based on these detected landmarks, the method further extracts geometric features to characterize the face shape changes caused by facial expressions. Geometric features are normalized, which are demonstrated to be robust to skin color and illumination variations, and are input to facial expression classifiers for analysis. Therefore, the third part of the method is the creation of probabilistic classifiers using the extracted features. Offline-trained support vector machines (SVMs) (a type of non-linear pattern classification technique) are employed to obtain the likelihood probability of each facial expression. Since the probabilistic classifiers only describe the facial expressions at individual frames, our framework further

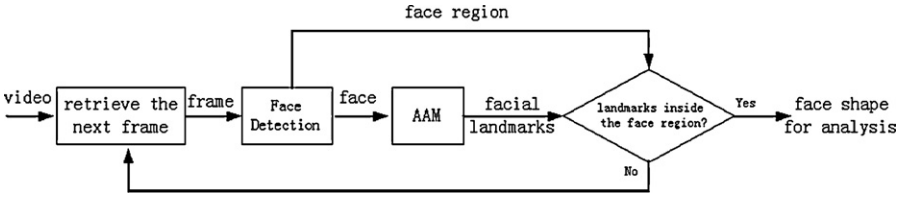


Fig. 1. Landmark detection and tracking in videos.

propagates the measurements at individual frames throughout videos using a sequential Bayesian inference scheme, to obtain a representation of facial expression changes in the whole video in the form of a temporal *probabilistic profile of facial expressions*. The computational framework is general, and applicable to all types of participants, for video-based facial expression analysis. The method is applied to a group of healthy people and representative patients with neuropsychiatric disorders, and measurements extracted from probabilistic profile of facial expressions are expected to distinguish between patients and controls.

3.2. Landmark detection and tracking in video

In this section, we present our landmark detection and tracking method. In the work of Alvino et al. (2007), the face region is manually outlined to obtain the deformation between faces with expression and neutral faces for analysis. However, manual labeling is time-consuming, and subjective to the person who labels the face. Especially in our study, the video of each participant may contain different facial expressions, up to 10,000–20,000 frames. Thus, it is a formidable task to manually mark all the face shapes in the videos. An automated system is desirable to perform the landmark points detection and tracking with minimum human intervention. To automate the process, we first detect the face and facial landmarks in the starting frame of the video using a face detector and an Active Appearance Model (AAM) (Cootes et al., 2001), and then track the landmark points in all the remaining frames. In the meantime, the face detector is running through the video to monitor the tracking, and re-initializes the tracker when participants' faces are out of the frontal view or occluded when the facial expression analysis cannot be performed. The whole scheme is illustrated in Fig. 1.

3.2.1. Face detection

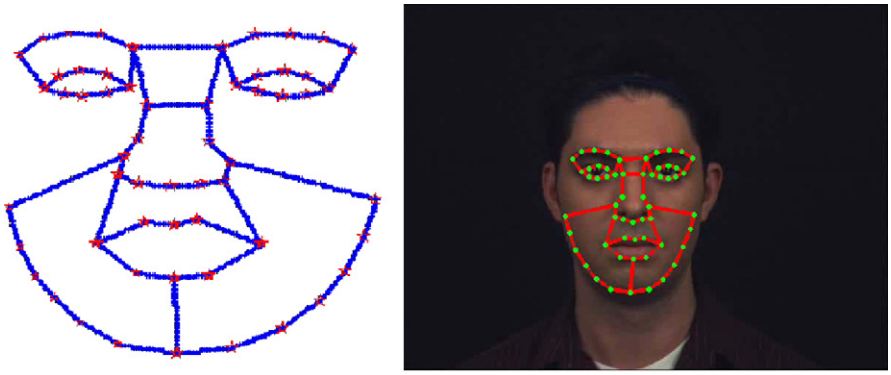
In our method, the face is automatically detected in the first frame of the video. Many face detection methods have been recently developed (Yang et al., 2002). Among current methods, the AdaBoost-based methods achieve excellent detection accuracy as well as real-time speed (Viola and Jones, 2004; Li and Zhang, 2004; Wang and Ji, 2007). Here we have applied AdaBoost algorithm with Haar features, to detect frontal and near-frontal faces (Viola and Jones, 2004). In this method, critical Haar features are sequentially selected from an over-complete feature set, which may contain more than 45,000 Haar wavelet features. Threshold classifiers are learned from the selected features, and are combined by AdaBoost. With a cascade structure (Viola and Jones, 2004), AdaBoost-based frontal face detection methods can achieve real-time speed (i.e., above 15 frames per second) with accuracy comparable to other methods. Note that our face detector aims at detecting only frontal faces, since our facial expression analysis is only applied to frontal faces. Fig. 2 shows face detection result in the first frame of a video.

3.2.2. Landmark detection and tracking

Inside each detected face, our method further identifies important landmarks to characterize facial expression changes. Active appearance model (AAM) (Cootes et al., 2001) locates these landmark points. AAM is a statistical method to model face appearance as well as face shape. In AAM, the face shape is represented by a set of landmarks, and the face texture is the image intensity or color of the whole face region. AAM face model combines the principal components from face texture and shape to formalize a vector, and then apply an additional principal component analysis (PCA) to further reduce the feature dimensionality. AAM models can be learned offline from collected annotated training samples. To locate landmarks in a given image with unknown faces, an efficient method has been developed in Cootes et al. (2001) to identify landmarks in images by minimizing the error between original face and its PCA reconstruction.

In our method, we define the face shape using 58 landmarks, as shown in Fig. 3(a). Among those landmarks, 5 points are defined on each eye brow, 8 points are defined on each eye, 11 points are defined on the nose, 8 points are defined on the mouth, and 13 points are defined on the face outline. The face texture in our AAM is defined as the RGB color values of the face, which are transformed on the mean shape. We collect about 100 face images with manually annotated landmarks, to obtained AAM models. Our implementation of AAM is modified from Stegmann et al. (2003). For given images with unknown faces, our method automatically detects the landmarks

Fig. 2. Face detection at the first frame.



(a) (b)

Fig. 3. Definition of landmarks and their detection: (a) 58 landmarks defined on face; (b) the landmarks detected at the first frame.

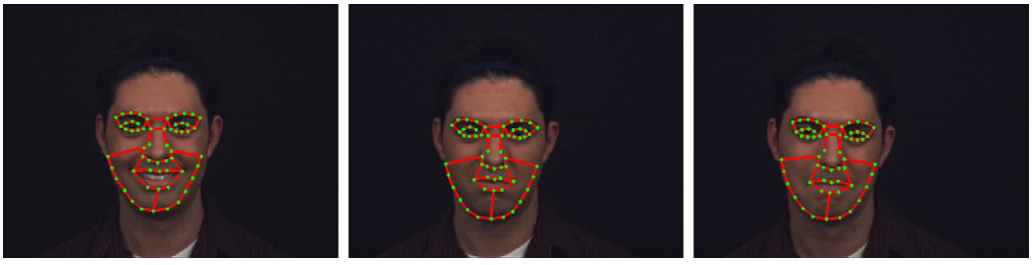


Fig. 4. Landmark tracking results in the video.

using the trained AAM model. Fig. 3(b) shows the detected landmarks at the first frame of video.

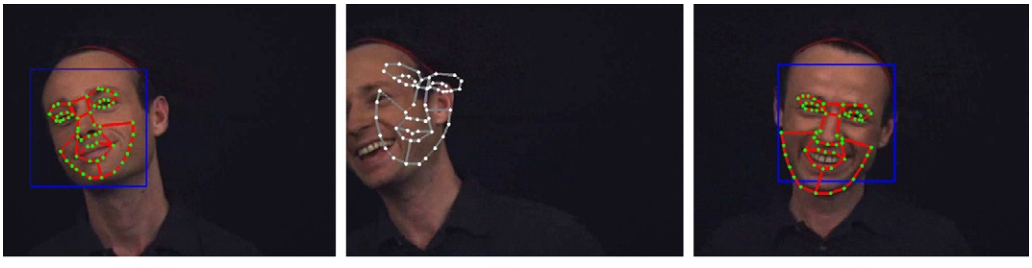
AAM is also used to track the landmark points in the rest of the video. At each frame, the face shape is initialized with the shape at the previous frame, and then AAM is applied to update the face shape at the current frame. Compared to independent landmark detection at individual frames, the AAM tracking speeds up the searching procedure by limiting the searching only around the previous location, given the assumption that the face moves smoothly. Fig. 4 shows the tracked landmarks in the video.

3.2.3. Combination of face detection and landmark detection

Although participants are instructed to restrict their head movement during data capture, the faces of participants could still be out of the frontal view sometimes. Such cases will fail during face tracking as well as in the facial expression analysis. To address such a problem, face detection is combined with landmark tracking such that landmarks detected can be monitored. The frontal face detector will lose detection when the faces are out of the frontal view or are occluded. Then the AAM tracking will be stopped. The face detector will keep searching frames until the face is back to its frontal view, or the occlusion is over. Then the AAM tracker is re-initialized inside the detected face region. In our experiments, only about 1.4% of frames in all participants have shown non-frontal faces. The faces out of frontal view will be excluded from the subsequent facial expression analysis. Fig. 5 shows how face detection can find the face that is out of view and re-initialize the face tracking.

3.3. Facial expression feature extraction

Geometric features are extracted from landmarks to characterize facial expression changes. The first type of geometric



(a) (b) (c)

Fig. 5. Landmark tracking combined with face detection: (a) tracking when face is detected; (b) AAM tracking is stopped when face is out of view; (c) tracking is re-initialized when the face is back to frontal view.

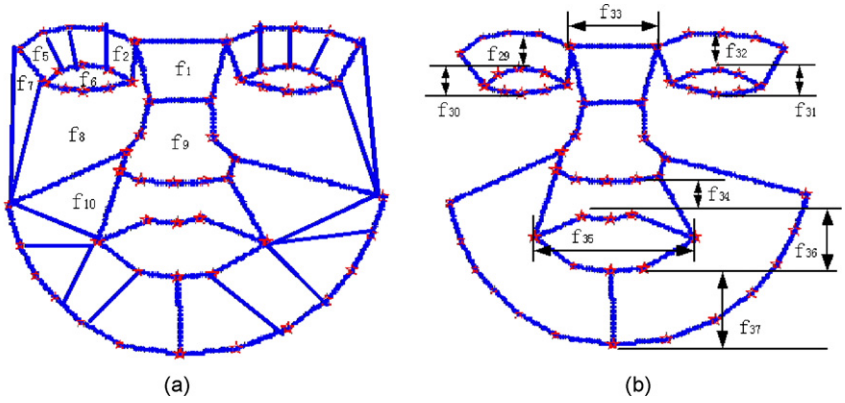


Fig. 6. Geometric features defined on landmarks for expression analysis: (a) 28 regions defined on landmarks; (b) distance features characterizing expression changes.

features are the area changes of 28 regions defined by 58 landmark points, as illustrated in Fig. 6(a). Such areas describe the global changes caused by facial expressions. There are also some facial actions that are closely related to expression changes. Such facial actions include eye opening, mouth opening, mouth corner movement and eyebrow movement. To specifically describe such actions, we define another type of geometric features, which measure distances between some landmark points, as illustrated in Fig. 6(b).

To eliminate effects of individual differences in facial expressions, extracted features are normalized in several ways. First, each face shape is normalized to the same scale. We use the face width to normalize faces, since it does not vary with facial expression changes. Second, geometric features are normalized by each subject's neutral faces. For example, each 2D geometric feature is divided by its corresponding value at the neutral expression of the same person. Thus, the geometric features only reflect the ratio changes of 2D face geometry, and individual topological differences are canceled. Finally, all the feature values are normalized to z-scores for subsequent analysis.

3.4. Facial expression classifiers

To quantify facial expressions, extracted features are used to train facial expression classifiers. We adopt support vector machines (SVM) (Cortes and Vapnik, 1995) as a pattern classification method to train classifiers. SVM is a binary classifier, which can separate two classes by projecting original data onto a high dimensional space through kernel functions. It provides good accuracy and generalization capability. At the training stage, SVM requires training samples to obtain class boundaries. At the classification stage, for a new data point, SVM returns a numeric result that represents the distance from the feature point to the class boundary. There are some efforts to interpret SVM outputs from the probabilistic point of view (Kwok, 2000; Platt, 2000). A direct method is to fit the output of SVM into parametric probability models (Platt, 2000). By assuming the distance output by SVM as a Gaussian likelihood model, the posterior probabilities can be directly fit with the sigmoid function as shown in Eq. (1):

$$p(Z|x=i) = N(\mu_i, \sigma_i)$$
$$p(x=i|Z) = \frac{1}{1+\exp(A_i z + B_i)} \quad (1)$$

where x is the class label, Z is the output of SVM. The parameters μ_i, σ_i, A_i, B_i are estimated from training data. We are mainly interested in the likelihood probability $p(Z|x)$, which will be used for the later Bayesian probability propagation.

The label x refers to the facial expression, and Z is the SVM output of extracted features. The class label x takes discrete values, i.e. $x=i$ indicates the existence of the ith facial expression. $p(Z|x=i)$ and $p(x=i|Z)$ are the likelihood and posterior probability of the ith facial expression respectively. However, SVM is essentially a binary classifier, while facial expression analysis is a multi-class problem as there are more than two facial expressions. There are usually two strategies, i.e., "one-against-another" and "one-against-all", to extend binary classifiers for a multi-class problem. In the "one-against-another" strategy, multiple binary classifiers are trained for each pair of classes. If there are k classes, there will be $k(k-1)/2$ binary classifiers. The final decision is made based on majority voting over all the binary classifiers. In another "one-against-all" strategy, k binary classifiers are trained for k classes, with each binary classifier trained to separate one facial expression from the other facial expressions. It is shown in Hsu (2002) that the "one-against-another" significantly increases the computational complexity, but improves the accuracy only slightly. Therefore, we apply the "one-against-all" strategy, i.e., training one SVM classifier for each expression using extracted features. For analysis of new data, the outputs from SVM classifiers, $p(Z|x=i)$, will be used for the probabilistic propagation in video sequences.

3.5. Probabilistic propagation in video: creation of probabilistic profile of facial expressions

The probabilistic outputs of facial expression classifiers, $p(z|x_i)$, model facial expressions at individual frames only, but have not fully utilized the temporal information of facial expressions in videos. We apply a sequential Bayesian estimation

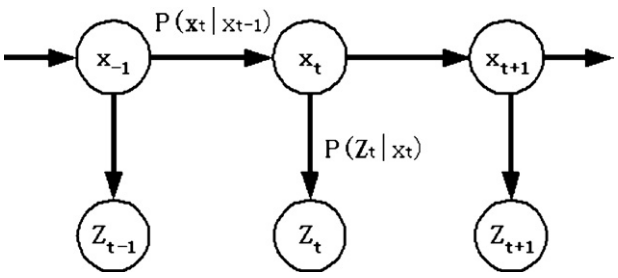


Fig. 7. A graphical model for facial expression inference in video.

scheme to propagate the posterior probabilities of facial expressions throughout the whole video. The sequential Bayesian estimation and its Monte Carlo derivations have been widely used in visual tracking (Yeasin et al., 2004), as they can handle sequential inference problems effectively and elegantly. Our method applies the sequential Bayesian estimation to infer the posterior probability $P(x_t|Z_{1:t})$ of facial expressions in video. In $P(x_t|Z_{1:t})$, x_t refers to the facial expression at the tth frame, and $Z_{1:t}$ represents the history of features extracted from frame 1 to frame t. To infer $P(x_t|Z_{1:t})$ from individual frames, a "dynamic model" is needed to describe the temporal relationship between facial expressions is needed. Such a dynamic model is denoted as $P(x_t|x_{1:t-1})$. Usually, there are two assumptions made in the sequential Bayesian estimation for purpose of simplicity: $P(x_t|x_{1:t-1}) = P(x_t|x_{t-1})$ and $P(Zt|x_{1:t}) = P(Zt|x_t)$. Such assumptions are called Markov properties, and have been widely adopted in the sequential inference. A graphical model that illustrates our sequential Bayesian estimation is shown in Fig. 7. With the assumptions of Markov property, posterior probabilities can be estimated from a measurement model $P(Zt|x_t)$ and a propagated prior $P(x_{t-1}|Z_{1:t-1})$, based on Bayes rule, as Eq. (2):

$$P(x_t|Z_{1:t}) = \frac{1}{C} P(Z_t|x_t) P(x_t|Z_{1:t-1}) \propto P(Z_t|x_t) \sum_{x_{t-1}} P(x_t|x_{t-1}) P(x_{t-1}|Z_{1:t-1}) \quad (2)$$

where $C = \sum_{x_t} P(Z_t|x_t) P(x_t|Z_{1:t-1})$ is a normalization constant that ensures that the summation of probability equals to 1. As shown in Eq. (2), the posterior probability $P(x_t|Z_{1:t})$ is sequentially estimated from the previous probability $P(x_{t-1}|Z_{1:t-1})$.

For the facial expression analysis of any participant using video, the likelihood measurement $P(Zt|x_t)$ is obtained by inputting features extracted from individual frames to the trained SVMs, which are described in Section 3.4. Then the posterior probability $P(x_t|Z_{1:t})$ is propagated throughout the video using sequential Bayesian inference, i.e., Eq. (2). The probabilities $P(x_t|Z_{1:t})$ therefore describe the temporal characteristics of facial expressions in videos, and provide the quantitative measurements that our method will use for subsequent analysis. These frame-wise probabilities help create a probabilistic profile for the expression, which can be visualized as a graph (see Fig. 10) with each curve corresponding to the response to the classifier from a particular expression. The five curves together form probabilistic profiles of facial expressions in videos.

3.6. Information extracted from probabilistic profiles: potential relevance to neuropsychiatric disorders

The probabilistic facial expression profiles provide rich information about the subtle and dynamic facial expression changes in video. Our method extracts several types of measurements from probabilistic profiles for facial expression analysis. The first measurement is the average of posterior probabilities of intended emotions, as a measurement of appropriate facial expressions. For the video segment of the ith intended emotion (e.g., one of happiness, sadness, anger, and fear), the averaged measurement is denoted as $\bar{P}_i = \frac{1}{n_i} \sum_{t=1}^{N_i} P(x_t = i|Z_{1:t})$, where n_i is the length of corresponding video for the ith intended emotion. The measurement $\bar{P}_i$ quantifies the correlation between participants' facial expressions and their intended emotions. A larger $\bar{P}_i$ refers to a greater expression of the intended emotion and a lower value corresponds to the amount of inappropriate affect. Therefore, by comparing the measurements of individuals from different groups, we can quantify the facial expression impairments.

Another measurement derived from a probabilistic facial expression profile is the probability of the neutral facial expressions in videos. For each video segment that contains one intended emotion, the posterior probability of the neutral expression indicates the lack of facial expression, and hence functions as a measure of flat affect, and can be correlated with flat affect ratings. Also, to eliminate the impact of different video lengths, we average the probability of neutral expression for each intended emotion, denoted as $\bar{N}_i = \frac{1}{n_i} \sum_{t=1}^{N_i} P(x_t = \text{Neutral}|Z_{1:t})$. Thus, the probabilistic profile and the measures of flat and inappropriate affect computed from the probabilistic profile of facial expressions, quantify the two major deficits associated with neuropsychiatric disorders.

Except for the average probabilities, two other measurements are the occurrence frequency of the appropriate and neutral expressions. Assuming that during a video, the number of frames where the maximal poster probability corresponds to the appropriate (when the expression picked by the probabilistic classifier is same as the intended) and neutral (when the classifier identifies the expression as neutral) expressions are l_a and l_n respectively, the occurrence frequency of appropriate and neutral expressions are defined as $f_a = l_a/n_i$, and $f_n = l_n/n_i$. Based on definitions, the two measurements describe appropriateness and flatness of facial expressions. These four measures indicate that the probabilistic profile has rich information for facial expression analysis. Developing more measurements from probabilistic profiles to better describe dynamics of facial expressions remains part of future research.

4. Results

In this section, we present results obtained by applying our framework to a few datasets that underline the generalization

Fig. 8. Emotional expressions from the professional actor database: (a) neutral; (b) happiness; (c) sadness; (d) anger; (e) fear.

capability, ease of applicability and automated nature of our method. We first train and validate the probabilistic facial expression classifiers that are to be applied at each frame, using a dataset of actors (Gur et al., 2002). The application of our framework is also validated by comparing the classification in video with human rating results. We then apply our computational framework to a collection of video segments from healthy people and present case studies on a patient with schizophrenia and a patient with Asperger's syndrome, which demonstrates the potential applicability of our framework.

4.1. Validation of probabilistic classifiers on actors

Although there are some existing facial expression databases (Lyons et al., 1999; Kanade et al., 2000; Bartlett et al., 2005), none of them are designed for clinical studies, especially the study of neuropsychiatric disorders such as schizophrenia. They mainly comprise posed expressions that actually do not follow the true trend of emotions, and usually contain expressions of only high intensity. In this study, we use a database of evoked facial expression images collected from professional actors, which have been acquired under experimental conditions that are similar to our patient/control data described below in Section 4.2. The actors database contains posed and evoked expressions of 32 professional actors (Gur et al., 2002). For each type of facial expression, the actors started with a neutral face, and then were guided by professional theatre directors through enactments of each of the four universal emotions of happiness, sadness, anger, and fear based on the evoked emotions procedure (Gur et al., 2002). Images were acquired while these actors were expressing emotion at three levels of intensity: mild, medium, and peak. Selected face examples are shown in Fig. 8.

The dataset is used as the training and validation data for facial expression classifiers. While posed databases have been used in the past for many expression studies, there is evidence that evoked emotions are more accurately perceived than posed expressions (Gur et al., 2002), and therefore we only use the evoked expressions in this study. The training images include four expressions (i.e., happiness, sadness, anger, fear) at all intensities of facial expression and neutral expression. We apply the method described in Section 3, except for the tracking part, on actors' images to create facial expression classifiers. The landmarks are detected on these facial images, and features are then extracted from these landmarks, as explained in Section 3. Using extracted features, total five SVM classifiers are trained, with one for each of the four expressions and the neutral expression. In order to test the accuracy of trained classifiers, they are further validated through a cross-validation procedure that is explained as follows. In each iteration of the cross-validation, face images from one subject are left out from the training data (neutral faces as well as faces with expression), and are tested on the classifiers trained on the remaining samples. The validation iterates until all the subjects are left out once and only once for testing. The testing accuracy averaged over all the data indicates the accuracy of trained classifiers. Table 1 summarizes the cross-validation accuracy of the facial expression classifiers. In this table, the rows show intended emotions, which are considered as ground truth in this validation, and the columns show classified expressions.

Our validation is further compared with human rating results. In a previous study (Gur et al., 2002), 41 students from undergraduate and graduate courses in psychology at the Drexel University were recruited as human raters. The raters were shown each face, and were asked to identify the emotional content of the face. The human raters were able to correctly identify 98% correct for happiness, and 67% correct for sadness, 77% correct for anger, 67% correct for fear. The overall accuracy of human raters is 77.8%, which is comparable with our cross validation accuracy. With more control/patient data being collected in our study, our ultimate goal is to use controls' data as the ground truth to train the facial expression classifiers.

4.2. Preliminary results on control/patient data

4.2.1. Data collection

In our preliminary study, facial expressions of individuals from different groups, including healthy controls, patients with schizophrenia, and patients with Asperger's syndrome, are acquired under the supervision of psychiatrists. The data was acquired under an approved IRB protocol of the University of Pennsylvania and permission has been obtained from subjects for the publication of pictures. All the participants are chosen in pairs matched for age, ethnicity, and gender. Each participant undergoes two conditions: posed and evoked. In the posed

Table 1
Confusion matrix from the cross-validation on actors' data

Classified intended	Happiness	Sadness	Anger	Fear	Neutral
Happiness	82.8%	11.2%	3.0%	3.0%	0.0%
Sadness	6.4%	73.3%	8.9%	10.2%	1.3%
Anger	3.9%	9.3%	76.4%	7.7%	2.8%
Fear	1.9%	7.6%	16.5%	74.1%	0.0%
Neutral	0.6%	0.0%	0.6%	0.0%	98.8%

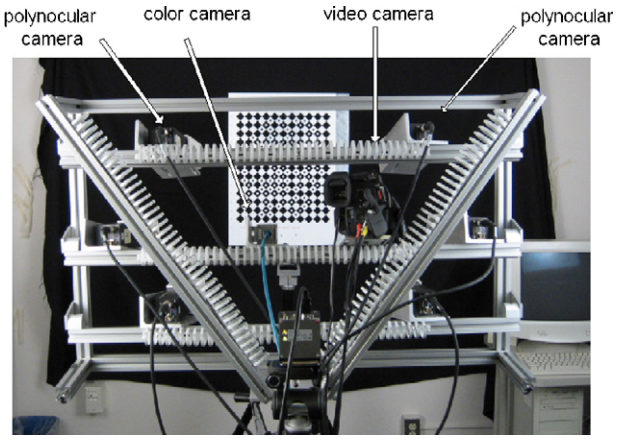


Fig. 9. The data capturing system.

session, participants are asked to express the emotions of happiness, anger, fear, sadness and disgust, at mild, medium, and peak levels. In the evoked session, participants are individually guided through vignettes, which are provided by participants themselves, and describe a situation in their life pertaining to each emotion. In order to elicit evoked expressions, the vignettes are recounted back to the participants by a psychiatrist, who guides them through all the three levels of expression intensity (mild, medium and peak) for each emotion. The videos and 2D images are acquired during the course of the expression using the setup illustrated in Fig. 9. There are six grayscale stereo camera, one color camera and a video camera (Verma et al., 2005). The color camera captures the 2D images, work on which has been described in Alvino et al. (2007). The six grayscale cameras and the color camera are calibrated to produce images that are used for 3D reconstruction of the faces. Our work using 3D surface data to analyze facial expressions is beyond the scope of this paper. Since evoked emotions are more accurately perceived than posed expressions, we only use the videos of evoked expressions for facial expression analysis, by applying the presented framework. We also exclude disgust from the analysis. Video recordings of facial emotional expression are segmented into 5 clips, 1 for each of the five emotions expressed. Each emotional segment begins from the mild intensity expressed, and ends at the extreme intensity, as identified during interview. The patient/control database is currently small and hence we use a few of the datasets to demonstrate the applicability of our framework. In future, as the dataset grows, we will be able to perform a group-based analysis, using the probabilistic profiles for the expressions obtained from our framework, via measures of flat and inappropriate affect computed from these.

4.2.2. Application of the video-based expression analysis framework

The method described in Section 3 is applied to several video clips, each of which contains one type of facial emotional expression of a participant, to obtain the probabilistic profile for facial expression. First, landmarks are detected and tracked in the video, and then facial expression features are extracted from tracked landmarks. The extracted features are input to facial expression classifiers that have been trained using actors' data, to obtain posterior probabilities of facial expression in videos. In order to validate our framework, we first compare our method with the human ratings using the Facial Expression Coding System (FACES) (Kring and Sloan, 2007) on a healthy control group. To further demonstrate its applicability to patients, we examine the method on healthy controls, a patient with schizophrenia, and a patient with Asperger's syndrome.

4.2.2.1. Validation on FACES. In order to validate the video-based framework, we compare our results with human ratings from Facial Expression Coding System (FACES) performed by human raters, on facial expressions from the healthy control group. In FACES, facial expressions in video segments are coded for frequency, duration, valance (positive or negative), and intensity (low, medium, high, very high). Two trained raters coded the frequency of facial expressions in each video segment. Expressions were coded if a neutral expression changed to an emotional expression and changed back to a neutral expression (1 expression coded) or to a different emotional expression (2 expressions coded). Facial changes independent of emotion expression (e.g. yawning, licking lips, talking, head nodding, head tilt, diverted eye gaze) were not counted as an emotion expression. For every expression, the emotion (happiness, sadness, anger, and fear), intensity (3-point scale of mild, moderate, and extreme), and duration (in seconds) were coded.

For each intended emotion, i.e., one of happiness, sadness, anger and fear, we have used 9 videos of healthy controls. For a rater to perform human FACES ratings, a video clip acquired from each participant is divided into separate segments, with each segment only corresponding to one type of intended emotion. All the segments are randomized such that raters were blind to the participants' intended emotion. With capturing speed at 15 frames per second, the length of segmented videos in the control group varies between 646 and 1431 frames for happiness, between 815 and 2620 frames for sadness, between 680 and 3252 frames for anger and between 1042 and 2578 frames for fear. The two raters are consistent with each other at most cases, even they may have small disagreements in the beginning and ending time of each segmented expression. FACES rating results from both raters are summarized in Table 2. In the table, rows show intended emotions of video segments, and columns show their FACES ratings.

Since there are possible inconsistency between human ratings and the intended emotion of participants, there are two types of percentages, shown in Table 2, to interpret the FACES rating results. Type I percentage refers to, among all the videos

Table 2
Confusion matrix of FACES ratings vs. intended emotions of controls

FACES intended	Happiness	Sadness	Anger	Fear	Percentage I
Happiness	41	2	4	3	82.0%
Sadness	22	30	2	1	54.5%
Anger	24	5	21	3	39.6%
Fear	9	18	3	25	45.5%
Percentage II	42.7%	54.6%	70.0%	78.1%	

captured in an intended emotion session, the percentage of expressions that are rated as the corresponding intended emotion. For example, among all video segments captured as part of the session for intended happiness, there are 82.0% expressions are rated as happiness by raters based on the FACES rule, and the remaining are rated as other expressions. The type II percentage illustrates, among all the expressions rated from FACES, the percentage of expressions that are actually from the correspond-

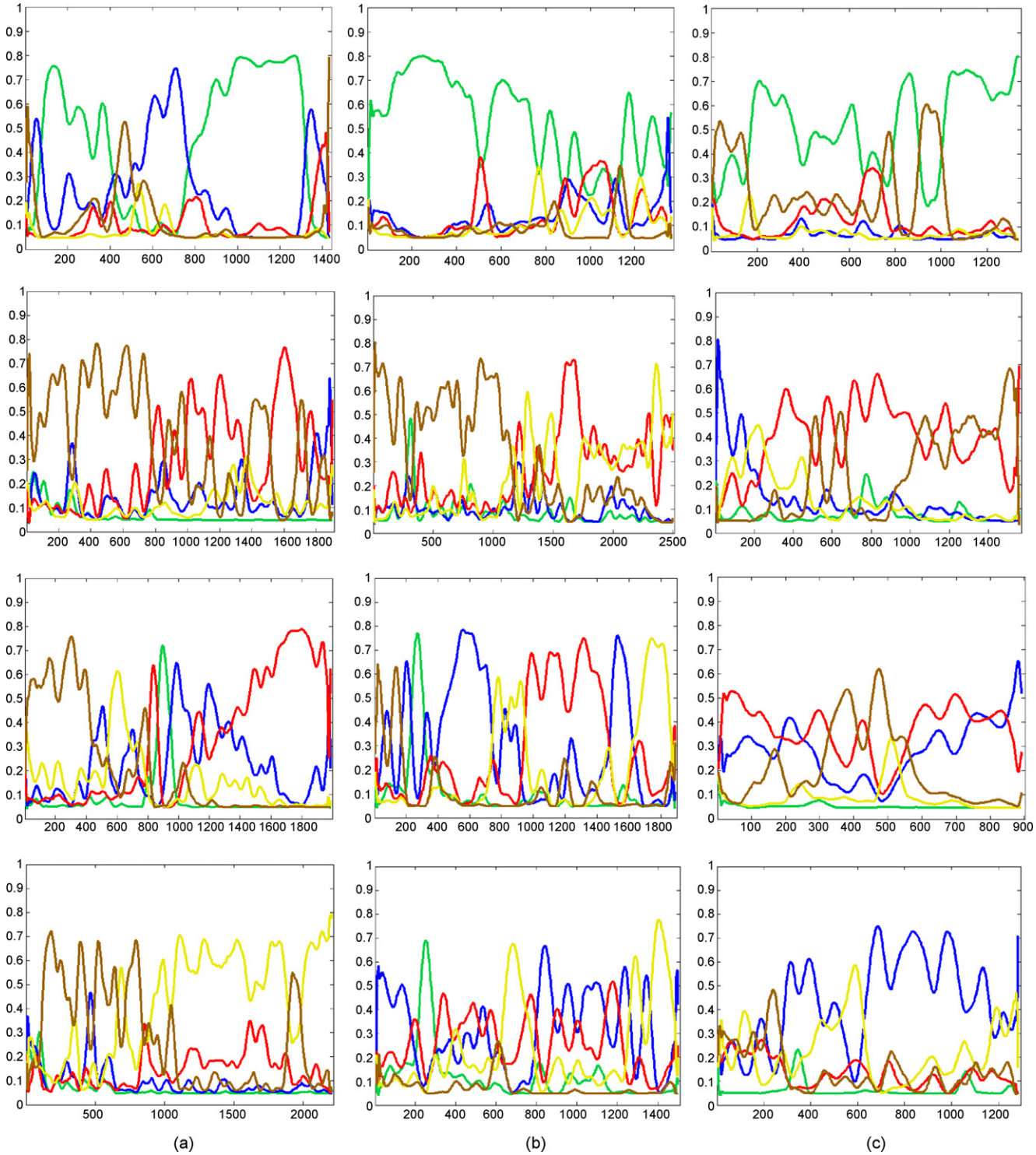


Fig. 10. Probabilistic profiles of facial expressions in video for (a) a healthy control; (b) a patient with schizophrenia; (c) a patient with Aspergers' syndrome. From top to bottom, graphs in each column show the probabilities obtained from an individual's intended happy, sad, anger, and fear emotions. In each figure, the horizontal axis is the frame number, and the vertical axis represents the posterior probability of facial expression. Profiles of different colors in the graphs represent different types of expressions: happy (green), sad (blue), anger (red), fear (yellow) and neutral (brown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 11. Emotional expressions of a healthy control: (a) happiness; (b) anger; (c) fear. The length of the bar is proportional to the probability associate with each expression. The original probability scales between 0 and 1.

ing intended emotion sessions. For example, among all the video segments rated as anger by FACES, 70.0% are from the intended anger sessions. The type I percentage is low for sadness, anger, and fear, demonstrating that videos may contain other expressions during one session of single intended emotion. The type II percentages show that the expression of happiness and sadness can appear in other emotion sessions, while anger and fear expressions appear more in the corresponding emotion sessions. Low percentages of both types demonstrate the uncertainty in expression and perception of emotions, and also highlight the difficulties of automatic analysis of evoked and subtle emotions.

We further compare our results from probabilistic profiles with human ratings from FACES, and show that our automatic method presents a reasonable accuracy. In this experiment, we validate only on those expressions in which the FACES ratings are consistent with intended emotions, to reduce the uncertainty factor in human emotion ratings. After generating probabilistic profiles, the mean posterior probabilities of each emotion in videos, i.e., $\bar{P}_i$, are used for facial expression recognition. The expression corresponding to the largest $\bar{P}_i$ is considered the intended emotion in the video. Table 3 summarizes the comparison results between FACES rating and our classification results. The rows show expressions rated from FACES, and the columns show automatically classified expressions based on the principle of maximal posterior probability. The recognition, except for the expression of sadness, provides reasonable results. Since all the classification results are based on the classifiers trained using actors' data, as we currently do not have enough controls for both training and validation, we expect that the accuracy would be increased when we have enough controls for training.

4.2.2.2. Case studies on individuals from different groups. The measurements extracted from probabilistic profiles can be used to examine different groups, such as healthy controls, and patients with deficits in facial expressions. Here we demonstrate the scalability of our method by applying it on three individuals, one from each group: healthy controls, patients with schizophrenia, and patients with Asperger's syndrome.

Fig. 10 shows the visualization of facial expressions probabilistic profiles as graphs of posterior probabilities of four facial expressions and neutral faces in each intended emotion of the three participants. In this figure, each color represents one of the four emotions: happiness (green), sadness (blue), anger (red), fear (yellow), and the neutral expression (brown). The horizontal

Table 3
Confusion matrix of classified expressions vs. FACES ratings of controls

Classified FACES	Happiness	Sadness	Anger	Fear	Accuracy I
Happiness	30	4	7	0	73.2%
Sadness	4	12	10	4	40.0%
Anger	0	5	16	0	76.2%
Fear	0	5	0	20	80.0%
Accuracy II	88.2%	46.2%	48.5%	83.3%	

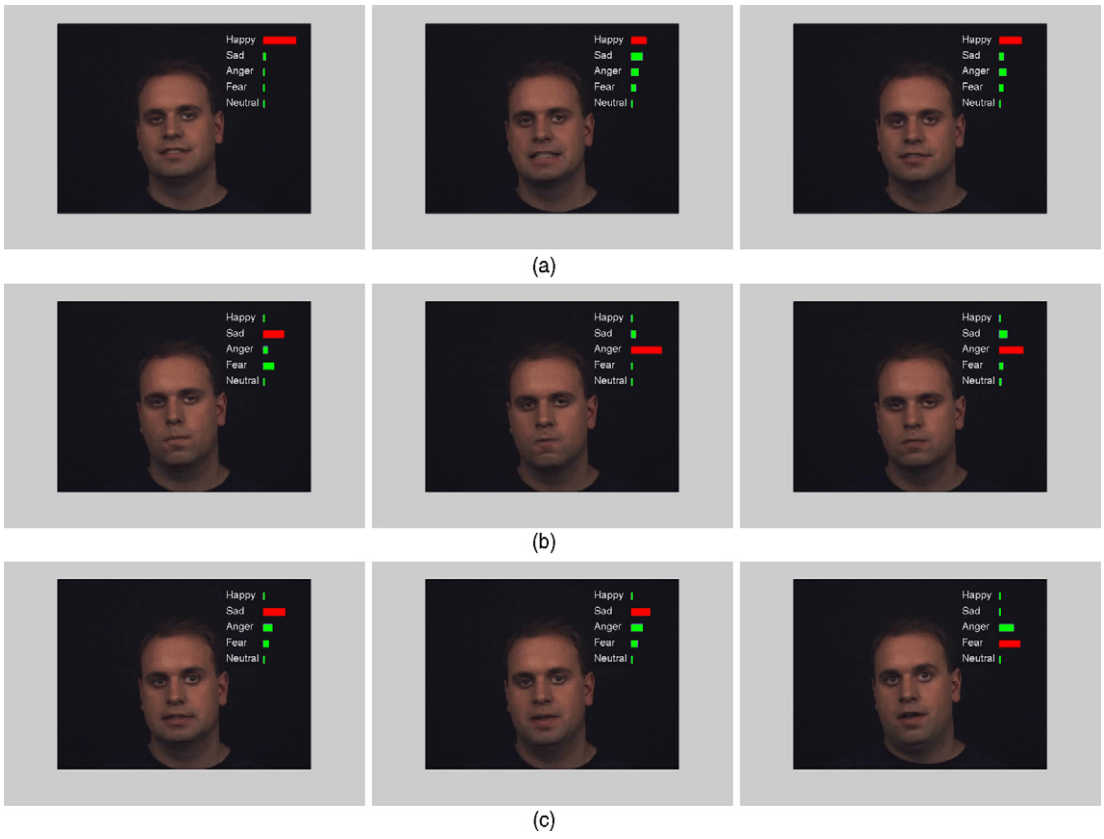


Fig. 12. Emotional expressions of a patient with schizophrenia: (a) happiness; (b) anger; (c) fear. The length of the bar is proportional to the probability associate with each expression.

axis represents the frame index, and the vertical axis represents the posterior probabilities $P(x_t|Z_{1:t})$ for one of four emotions and neutral expression, which is denoted as x_t, at the tth frames. Some frames from the videos corresponding to these profiles with the corresponding probabilities are shown in Figs. 11–13, where the probabilities are visualized as bars on the top right corner, with the bar of the longest length corresponding to the outcome of the frame on the application of the classifiers. As displayed in these figures, the posterior probabilities of expressions, $P(x_t|Z_{1:t})$, indicate the trends of facial expression changes of individuals in the video.

An inspection of Figs. 10–13 indicates that probabilistic profiles of facial expressions are able to capture subtle expressions and to identify expressions that are different from the intended emotion, hence determining the inappropriateness of emotion, as well as identify frames that have neutral expression thereby identifying the flatness of expression. The probability bars associated with the top right corner reveals that the classifier is able to correctly determine the type and intensity of emotion. In Fig. 11(c), neutral is picked up instead of fear. The classifier is able to identify the emotion correctly even when the expression deviated from the intended (Fig. 12(b), frame 1, sadness is identified instead of the intended anger and in Fig. 13(c) in which sadness is identified instead of intended fear). These expressions are rated to be correct by a human rater. Subtle expressions are also well identified (Fig. 13(a), frame 3, Fig. 12(a), frame 3).

Table 4
Average probability of intended emotion in videos of three participants

Group	Happiness	Sadness	Anger	Fear	Average
Healthy	0.3889	0.2193	0.3656	0.3789	0.3382
Schizophrenia	0.3706	0.1545	0.2540	0.2286	0.2519
Asperger's	0.3275	0.2227	0.3519	0.2780	0.2950

After obtaining the probabilistic profiles of facial expression for each intended emotion, we compute quantitative measurements to characterize facial expressions in video. As described in Section 3.6, four types of measurements are calculated, i.e., $\bar{P}_i$,$\bar{N}_i$, f_a, and f_n. Specifically, $\bar{P}_i$ and f_a quantifies the appropriate expression for the ith intended emotion (e.g., one of the four emotions: happiness, sadness, anger, fear), and $\bar{N}_i$ and f_n quantify the neutral expression in the ith intended emotion. These measurements will be used to correlate with clinical measurements of inappropriate and flat affect when we have collected enough samples for the group study. Tables 4 and 5 show two

Table 5
Average probability of neutral expression in videos of three participants

Group	Happiness	Sadness	Anger	Fear	Average
Healthy	0.1207	0.3830	0.2117	0.2712	0.2467
Schizophrenia	0.0845	0.3167	0.1154	0.0732	0.1475
Asperger's	0.2262	0.2488	0.2039	0.1499	0.2072

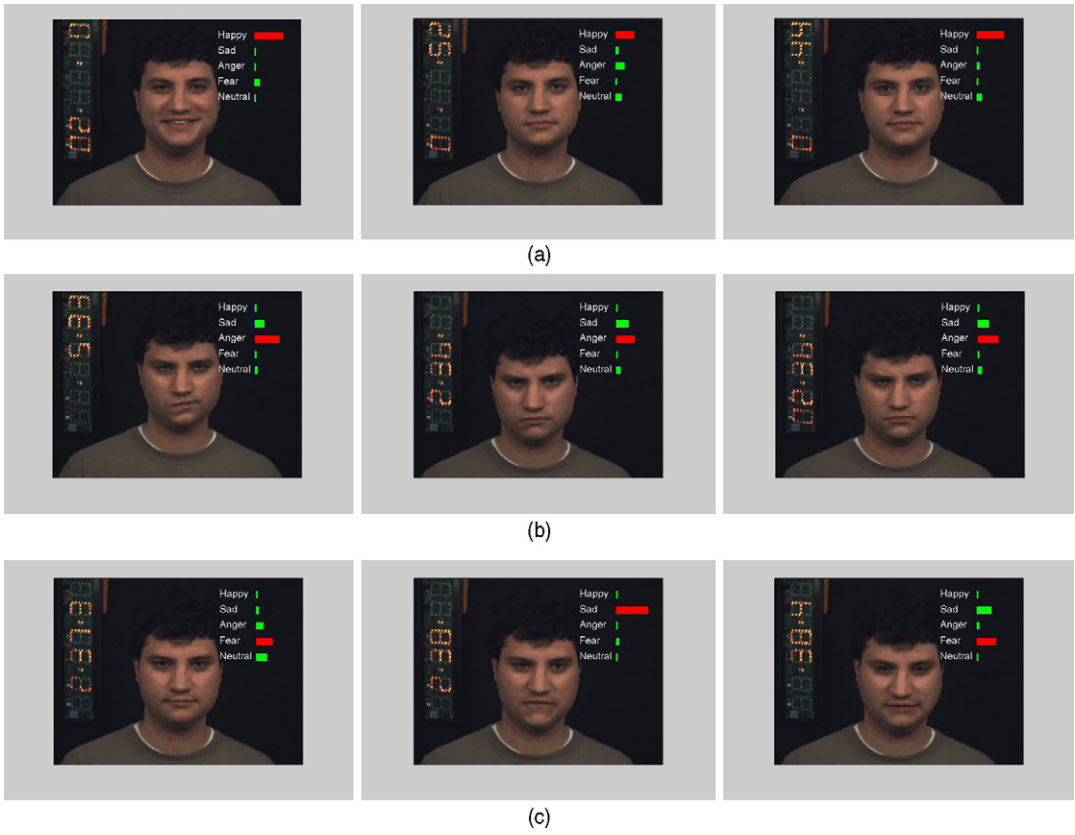


Fig. 13. Emotional expressions of a patient with Asperger's syndrome: (a) happiness; (b) anger; (c) fear. The length of the bar is proportional to the probability associate with each expression.

Table 6
Occurrence frequency of appropriate expressions in videos of three participants

Group	Happiness	Sadness	Anger	Fear	Average
Healthy	0.6431	0.1728	0.5172	0.8307	0.5410
Schizophrenia	0.7898	0.0833	0.3100	0.2459	0.3573
Asperger's	0.9545	0.1166	0.6136	0.2500	0.4837

types of average probabilities for each emotion, along with the averages over all the emotions, for the three participants. Tables 6 and 7 show two occurrence frequency measurements for each emotion.

Tables 4 and 6 demonstrate that overall, the healthy control expresses intended emotion better than the patient with Asperger's and schizophrenia (especially in the fear). The averages (column 6) in both the tables show that the levels of impairment of the Asperger's patient lie in between that of the controls and the schizophrenia patient. Tables 5 and 7 also show that the individuals demonstrate different levels of expressiveness. However, the control has more neutral expression than the two patients. As confirmed by clinical ratings (using SANS (Andreasen, 1984)) by two experts, the controls actually show almost the same level of flatness as the patients (the flatness index scores at 2 and 3 according to two raters). However, such an observation does not permit conclusions regarding group behavior of patients relative to controls. We expect to perform group difference studies when more patient data has been acquired.

Table 7
Occurrence frequency of neutral expression in videos of three participants

Group	Happiness	Sadness	Anger	Fear	Average
Healthy	0.0922	0.4884	0.2414	0.3043	0.2816
Schizophrenia	0.0235	0.3856	0.0930	0.0133	0.1289
Asperger's	0.2096	0.2744	0.2108	0.1402	0.2087

5. Discussion and future work

In this paper, we present an automated computational framework for analyzing facial expressions using video data, producing a probabilistic profile of expression change. The framework explores rich information contained in the video, by providing a probabilistic composition of each frame of the sequence, thereby highlighting subtle differences as well as the possibility of a mixture of emotions. The potential relevance for neuropsychiatric disorders stems from the propensity for impaired emotion expression including inappropriate or flat affect. Thus far diagnosis of impaired affect expression required trained clinical observers. The framework benefits from being automated, thereby helping in processing lengthy video sequences. It is also applicable to participants from groups with different pathologies or various stages of disease progression.

The preliminary results demonstrate the capability of our video-based expression analysis method in identifying characteristics of facial expressions through probabilistic expression profiles (Fig. 10). These expression profiles, in conjunction with the metrics of appropriateness and flatness computed from them, provide extensive information about the expression and capture the subtleties of expression change. Patients follow different trends of facial expression than healthy participants. The facial expressions of the healthy control are more consistent with the expected trend of intended emotion, that is the emotion gradually progresses from mild to moderate and finally to the peak level. Especially for the expressions of anger and fear, the facial expression trends of the healthy control better characterize the intended emotion than the patients. Another observation is that the intensity of expression of the healthy control is higher than the patients. The differences between the three subjects are mainly in the negative emotions of sadness and fear. Especially for fear, the healthy control is more expressive than the two patients. Also the measurements averaged over all expressions demonstrate the difference between individuals, although the differences in the happy and anger expressions are small. We believe that with additional enrollment of subjects, our framework will be able to identify significant group differences using the presented computational methods.

It is also observed that the facial expression recognition results of the expression of sadness (as seen in the graph of probabilities in Fig. 10) are not as good as other facial expressions. Sadness is somehow confused with anger expression perhaps owing to the following two reasons. First, the sad and anger expressions share some similar facial movements, such as eyebrow lower, and lip corner depressor (Kohler et al., 2004). Such facial movements may cause confusion between two expressions. Second, the participants (both patients and controls) usually show more subtle expressions than actors. Since our classifiers are built on actors' expressions, they may not recognize well the low intensity expression of sadness and concentrate more on salient facial expressions such as anger. Our solution to this problem is to retrain the facial expression classifiers using data from healthy controls when additional data from healthy controls is available. We believe that by training facial expression classifiers based on a healthy population, our method can better characterize the true trends of intended emotions. We expect that training with healthy controls will also help the separation between sadness and anger.

The experiments pave the way for establishing a video-based method for quantitative analysis of facial expressions in clinical research. The method can be applied to any disorder that causes affect deficits. The probabilistic profile of facial expressions provides a graphical visualization of affect deficits as well as measures to quantify flatness and inappropriateness of expression. In future, we will apply our framework to large population group-based studies, to quantify the group differences between healthy controls and patients, to correlate with clinical measurements, and to obtain a population profile of temporal change during the course of a facial expression. We expect that the knowledge obtained from such an analysis will help in diagnosis, prognosis, and studying treatment effects.

Acknowledgement

We would like to acknowledge support from NIH grants 1R01MH73174-01 (for method development) and R01-MH060722 (for data acquisition).

References

Alvino C, Kohler C, Barrett F, Gur RE, Gur RC, Verma R. Computerized measurement of facial expression of emotions in schizophrenia. J Neurosci Methods 2007;163(2):350–61.

Andreasen NC. Scale for the assessment of negative symptoms (SANS). Iowa City: University of Iowa; 1984.

Bartlett MS, Hager JC, Ekman P, Sejnowski TJ. Measuring facial expressions by computer image analysis. Psychophysiology 1999;36:253–63.

Bartlett MS, Littlewort G, Frank MG, Lainscsek C, Fasel I, Movellan J. Recognizing facial expression: machine learning and application to spontaneous behavior. CVPR 2005:568–73.

Berenbaum H, Oltmann TF. Emotional experience and expression in schizophrenia and depression. J Abnormal Psychiatry Res 1992;101(1):37–44.

Chang Y, Hu C, Turk M. Probabilistic expression analysis on manifolds. CVPR 2004.

Chang Y, Vieira M, Turk M, Velho L. Automatic 3D facial expression analysis in videos. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures; 2005.

Cohen I, Sebe N, Garg A, Chen LS, Huang TS. Facial expression recognition from video sequences: temporal and static modeling. Comput Vis Image Understand 2003a;91(1–2):160–87.

Cohen I, Seve N, Cozman GG, Cirelo MC, Huang TS. Learning bayesian network classifier for facial expression recognition using both labeled and unlabeled data. CVPR 2003b.

Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE Trans PAMI 2001;23(6):681–5.

Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.

Davatzikos C. Measuring biological shape using geometry-based shape transformations. Image Vis Comput 2001;19:63–74.

Ekman P, Friesen WV. Facial action coding system: a technique for the measurement of facial movement. Palo Alto, California: Consulting Psychologists Press; 1978.

Essa IA, Pentland AP. Facial expression recognition using a dynamic model and motion energy. ICCV 1995:360–7.

Essa IA, Pentland AP. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Trans PAMI 1997;19(7):757–63.

Fasel B, Luettin J. Recognition of asymmetric facial action unit activities and intensities. ICPR 2000:1100–3.

Fasel B, Luettin J. Automatic facial expression analysis: a survey. Pattern Recognit 2003;36(1):259–75.

Gaebel W, Wölwer WC. Facial expression and emotional face recognition in schizophrenia and depression. Eur Arch Psychiatry Clin Neurosci 1992;242:46–52.

Gur R, Sara R, Hagendoorn M, Marom O, Hughett P, Macy L, et al. A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. J Neurosci Methods 2002;115(2):137–43.

Gur RE, Kohler CG, Ragland JD, Siegel SJ, Lesko K, Bilker WB, et al. Flat affect in Schizophrenia: relation to emotion processing and neurocognitive measures. Schizophrenia Bull 2006;32(2):279–87.

Hellewell J, Connell J, Deakin JFW. Affect-judgment and facial recognition memory in schizophrenia. Psychopathology 1994;27:255–61.

Hsu C-W, Lin C-J. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks 2002;13(2).

Kanade T, Cohn JF, Tian Y. Comprehensive database for facial expression analysis. AFRG 2000:46–53.

Kimura S, Yachida M. Facial expression recognition and its degree estimation. In: IEEE Conference on Computer Vision and Pattern Recognition; 1997. p. 295–300.

Kohler CG, Turner T, Stolar NM, Bilker WB, Brensinger CM, Gur RE, Gur RC. Differences in facial expressions of four universal emotions. Psychiatry Res 2004;128(3):235–44.

Kring AM, Sloan DS. The facial expression coding system (FACES): development, validation, and utility. Psychol Assess 2007;19:210–24.

Kring AM, Neale MAJM, Harvey PD. A multichannel, multimethod assessment of affective flattening in schizophrenia. Psychiatry Res 1994;54: 211–22.

Kwok JT-Y. The evidence framework applied to support vector machines. IEEE Trans Neural Networks 2000;11(5):1162–73.

Lanitis A, Taylor CJ, Cootes TF. Automatic interpretation and coding of face images using flexible models. IEEE Trans PAMI 1997;19(7):743–56.

Li SZ, Zhang Z. FloatBoost learning and statistical face detection. IEEE Trans PAMI 2004;26(9):1112–23.

Lien JJ, Kanade T, Cohn JF, Li C-C. Subtly different facial expression recognition and expression intensity estimation. In: IEEE Conference on Computer Vision and Pattern Recognition; 1998. p. 853–9.

Lien JJ, Kanade T, Cohn JF, Li C-C. Detection, tracking, and classification of action units in facial expression. Robtics Autonomous Syst 2000;31:131–46.

Littlewort G, Bartlett M, Fasel I, Susskind J, Movellan J. Dynamics of facial expression extracted automatically from video. Image Vis Comput 2006;24(6):615–25.

Lyons MJ, Budynek J, Akamatsu S. Automatic classification of single facial images. IEEE Trans PAMI 1999;21(12):1357–62.

Mandal M, Pandey R, Prasad A. Facial expressions of emotions and schizophrenia: a review. Schizophrenia Bull 1998;24(3):399–412.

Morrison RL, Bellack AS, Mueser KT. Deficits in facial-affect recognition and schizophrenia. Schizophrenia Bull 1988;14(1):67–83.

Ohta H, Saji H, Nakatani H. Recognition of facial expressions using muscle-based feature models. In: IEEE International Conference on Pattern Recognition; 1998.

Pantic M, Rothkrantz LJM. Automatic analysis of facial expressions: the state of the art. IEEE Trans PAMI 2000;22(12):1424–45.

Platt J. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. Cambridge, MA: MIT Press; 2000.

Schneider FHH, Himer W, Huss D, Mattes R, Adam B. Computer-based analysis of facial action in schizophrenic and depressed patients. Eur Arch Psychiatry Clin Neurosci 1990;240(2):67–76.

Stegmann MB, Ersbol BK, Larsen R. FAME—a flexible appearance modeling environment. IEEE Trans Med Imaging 2003;22(10):1319–31.

Tian Y-L, Kanade T, Cohn JF. Recognizing action units for facial expression analysis. IEEE Trans PAMI 2001;23(2):97–115.

Tian Y-L, Kanade T, Cohn JF. Handbook of face recognition. Springer; 2005.

Verma R, Davatzikos C, Loughead J, Indersmitten T, Hu R, Kohler C, et al. Quantification of facial expressions using high dimensional shape transformations. J Neurosci Methods 2005;141:61–73.

Viola P, Jones M. Robust real-time object detection. Int J Comput Vis 2004;57(2):137–54.

Wang P, Ji Q. Learning discriminant features for multi-view face and eye detection. Comput Vis Image Understand 2007;105(2):99–111.

Wang Y, Ai H, Wu B, Huang C. Real time facial expression recognition with AdaBoost. ICPR 2004.

Wen Z, Huang TS. Capturing subtle facial motions in 3D face tracking. In: IEEE International Conference on Computer Vision and Pattern Recognition; 2003. p. 1343–50.

Wiskott L, Fellous J-M, Kruger N, v.d. Malsburg C. Face recognition by elastic bunch graph matching. IEEE Trans PAMI 1997;19(7):775–9.

Yacoob Y, Davis LS. Recognizing human facial expressions from long image sequeces using optical flow. IEEE Trans PAMI 1996;18(6):636–42.

Yang M-H, Kriegman DJ, Ahuja N. Detecting faces in images: a survey. IEEE Trans PAMI 2002;24(1):34–58.

Yeasin M, Bullot B, Sharma R. From facial expression to level of interest: a spatio-temporal approach. In: IEEE Conference on Computer Vision and Pattern Recognition; 2004.

Zhang Z, Lyons M, Schuster M, Akamatsu S. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. AFRG 1998:454–9.