Compound facial expressions of emotion

Shichuan Du, Yong Tao, and Aleix M. Martinez[1]

Department of Electrical and Computer Engineering, and Center for Cognitive and Brain Sciences, The Ohio State University, Columbus, OH 43210

Edited by David J. Heeger, New York University, New York, NY, and approved February 28, 2014 (received for review December 1, 2013)

Understanding the different categories of facial expressions of emotion regularly used by us is essential to gain insights into human cognition and affect as well as for the design of computational models and perceptual interfaces. Past research on facial expressions of emotion has focused on the study of six basic categories—happiness, surprise, anger, sadness, fear, and disgust. However, many more facial expressions of emotion exist and are used regularly by humans. This paper describes an important group of expressions, which we call compound emotion categories. Compound emotions are those that can be constructed by combining basic component categories to create new ones. For instance, happily surprised and angrily surprised are two distinct compound emotion categories. The present work defines 21 distinct emotion categories. Sample images of their facial expressions were collected from 230 human subjects. A Facial Action Coding System analysis shows the production of these 21 categories is different but consistent with the subordinate categories they represent (e.g., a happily surprised expression combines muscle movements observed in happiness and surprised). We show that these differences are sufficient to distinguish between the 21 defined categories. We then use a computational model of face perception to demonstrate that most of these categories are also visually discriminable from one another.

categorization | action units | face recognition

> Some men . . . have the same facial expressions. . . . For when one suffers anything, one becomes as if one has the kind of expression: when one is angry, the sign of the same class is angry.
>
> *Physiognomics*, unknown author (attributed to Aristotle), circa fourth-century B.C. (1)

As nicely illustrated in the quote above, for centuries it has been known that many emotional states are broadcasted to the world through facial expressions of emotion. Contemporaries of Aristotle studied how to read facial expressions and how to categorize them (2). In a majestic monograph, Duchenne (3) demonstrated which facial muscles are activated when producing commonly observed facial expressions of emotion, including happiness, surprise (attention), sadness, anger (aggression), fear, and disgust.

Surprisingly, although Plato, Aristotle, Descartes, and Hobbes (1, 4, 5), among others, mentioned other types of facial expressions, subsequent research has mainly focused on the study of the six facial expressions of emotion listed above (6–9). However, any successful theory and computational model of visual perception and emotion ought to explain how all possible facial expressions of emotion are recognized, not just the six listed above. For example, people regularly produce a happily surprised expression and observers do not have any problem distinguishing it from a facial expression of angrily surprised (Fig. 1 *H* and *Q*). To achieve this, the facial movements involved in the production stage should be different from those of other categories of emotion, but consistent with those of the subordinate categories being expressed, which means the muscle activations of happily surprised should be sufficiently different from those of angrily surprised, if they are to be unambiguously discriminated by observers. At the same time, we would expect that happily surprised will involve muscles typically used in the production of facial expressions of happiness and surprise such that both subordinate categories can be readily detected.

The emotion categories described above can be classified into two groups. We refer to the first group as basic emotions, which include happiness, surprise, anger, sadness, fear, and disgust (see sample images in Fig. 1 *B–G*). Herein, we use the term "basic" to refer to the fact that such emotion categories cannot be decomposed into smaller semantic labels. We could have used other terms, such as "component" or "cardinal" emotions, but we prefer basic because this terminology is already prevalent in the literature (10); this is not to mean that these categories are more basic than others, because this is an area of intense debate (11).

The second group corresponds to compound emotions. Here, compound means that the emotion category is constructed as a combination of two basic emotion categories. Obviously, not all combinations are meaningful for humans. Fig. 1 *H–S* shows the 12 compound emotions most typically expressed by humans. Another set of three typical emotion categories includes appall, hate, and awe (Fig. 1 *T–V*). These three additional categories are also defined as compound emotions. Appall is the act of feeling disgust and anger with the emphasis being on disgust; i.e., when appalled we feel more disgusted than angry. Hate also involves the feeling of disgust and anger but, this time, the emphasis is on anger. Awe is the feeling of fear and wonder (surprise) with the emphasis being placed on the latter.

In the present work, we demonstrate that the production and visual perception of these 22 emotion categories is consistent within categories and differential between them. These results suggest that the repertoire of facial expressions typically used by humans is better described using a rich set of basic and compound categories rather than a small set of basic elements.

Results

Database. If we are to build a database that can be successfully used in computer vision and machine learning experiments as well as cognitive science and neuroscience studies, data collection must adhere to strict protocols. Because little is known about compound emotions, our goal is to minimize effects due to lighting, pose, and subtleness of the expression. All other variables should, however, vary to guarantee proper analysis.

Significance

Though people regularly recognize many distinct emotions, for the most part, research studies have been limited to six basic categories—happiness, surprise, sadness, anger, fear, and disgust; the reason for this is grounded in the assumption that only these six categories are differentially represented by our cognitive and social systems. The results reported herein propound otherwise, suggesting that a larger number of categories is used by humans.

Author contributions: A.M.M. designed research; S.D. and Y.T. performed research; S.D. and A.M.M. analyzed data; and S.D. and A.M.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: martinez.158@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1322355111/-/DCSupplemental.

Fig. 1. Sample images of the 22 categories in the database: (*A*) neutral, (*B*) happy, (*C*) sad, (*D*) fearful, (*E*) angry, (*F*) surprised, (*G*) disgusted, (*H*) happily surprised, (*I*) happily disgusted, (*J*) sadly fearful, (*K*) sadly angry, (*L*) sadly surprised, (*M*) sadly disgusted, (*N*) fearfully angry, (*O*) fearfully surprised, (*P*) fearfully disgusted, (*Q*) angrily surprised, (*R*) angrily disgusted, (*S*) disgustedly surprised, (*T*) appalled, (*U*) hatred, and (*V*) awed.

Sample pictures for neutral and each of the six basic and 15 compound emotions are shown in Fig. 1. Images were only accepted when the experimenter obtained fully recognizable expressions. Nonetheless, all images were subsequently evaluated by the research team. Subjects who had one or more incorrectly expressed emotions were discarded. The images of 230 subjects passed this evaluation (see *Materials and Methods*).

Action Units Analysis. In their seminal work, Ekman and Friesen (12) defined a coding system that makes for a clear, compact representation of the muscle activation of a facial expression. Their Facial Action Coding System (FACS) is given by a set of action units (AUs). Each AU codes the fundamental actions of individual or groups of muscles typically seen while producing facial expressions of emotion. For example, AU 4 defines the contraction of two muscles resulting in the lowering of the eyebrows (with the emphasis being in the inner section). This AU is typically observed in expressions of sadness, fear, and anger (7).

We FACS coded all of the images in our database. The consistently active AUs, present in more than 70% of the subjects in each of the emotion categories, are shown in Table 1. Typical intersubject variabilities are given in brackets; these correspond to AUs seen in some but not all individuals, with the percentages next to them representing the proportion of subjects that use this AU when expressing this emotion.

As expected, the AU analysis of the six basic emotions in our database is consistent with that given in ref. 12. The only small difference is in some of the observed intersubject variability given in parentheses—i.e., AUs that some but not all subjects used when expressing one of the basic emotion categories; this is to be expected because our database incorporates a much larger set of subjects than the one in ref. 12. Also, all of the subjects we have FACS coded showed their teeth when expressing happiness (AU 25), and this was not the case in ref. 12. Moreover, only half of our subjects used AU 6 (cheek raiser) when expressing sadness, which suggests a small relevance of this AU as other studies have previously suggested (13–15). Similarly, most of our subjects did not include AU 27 (mouth stretch) in fear, which seems to be active only when this expression is exaggerated.

Table 1 also lists the AUs for each of the compound emotion categories. Note that the AUs of the subordinate categories are used to form the compound category unless there is a conflict. For example, lip presser (AU 24) may be used to express disgust while lips part (AU 25) is used in joy. When producing the facial expression of happily disgusted, it is impossible to keep both. In this case, AU 24 is dropped. Fig. 2 shows this and five other examples (further illustrated in Table 1). The underlined AUs of a compound emotion are present in both of their subordinate categories. An asterisk indicates the AU does not occur in either of the basic categories and is, hence, novel to the compound emotion. We did not find any such AU consistently used by most subjects; nevertheless, a few subjects did incorporate them, e.g., AU 25 (lips part) in sadly disgusted. Additional examples are given in Fig. S1, where we include a figure with the subordinate relations for the nine remaining compound facial expressions of emotion.

Table 1. Prototypical AUs observed in each basic and compound emotion category

Category	Prototypical (and variant AUs)
Happy	12, 25 [6 (51%)]
Sad	4, 15 [1 (60%), 6 (50%), 11 (26%), 17 (67%)]
Fearful	1, 4, 20, 25 [2 (57%), 5 (63%), 26 (33%)]
Angry	4, 7, 24 [10 (26%), 17 (52%), 23 (29%)]
Surprised	1, 2, 25, 26 [5 (66%)]
Disgusted	9, 10, 17 [4 (31%), 24 (26%)]
Happily surprised	1, 2, 12, 25 [5 (64%), 26 (67%)]
Happily disgusted	10, 12, 25 [4 (32%), 6 (61%), 9 (59%)]
Sadly fearful	1, 4, 20, 25 [2 (46%), 5 (24%), 6 (34%), 15 (30%)]
Sadly angry	4, 15 [6 (26%), 7 (48%), 11 (20%), 17 (50%)]
Sadly surprised	1, 4, 25, 26 [2 (27%), 6 (31%)]
Sadly disgusted	4, 10 [1 (49%), 6 (61%), 9 (20%), 11 (35%), 15 (54%), 17 (47%), 25 (43%)*]
Fearfully angry	4, 20, 25 [5 (40%), 7 (39%), 10 (30%), 11 (33%)*]
Fearfully surprised	1, 2, 5, 20, 25 [4 (47%), 10 (35%)*, 11 (22%)*, 26 (51%)]
Fearfully disgusted	1, 4, 10, 20, 25 [2 (64%), 5 (50%), 6 (26%)*, 9 (28%), 15 (33%)*]
Angrily surprised	4, 25, 26 [5 (35%), 7 (50%), 10 (34%)]
Angrily disgusted	4, 10, 17 [7 (60%), 9 (57%), 24 (36%)]
Disgustedly surprised	1, 2, 5, 10 [4 (45%), 9 (37%), 17 (66%), 24 (33%)]
Appalled	4, 10, [6 (25%)*, 9 (56%), 17 (67%), 24 (36%)]
Hatred	4, 10, [7 (57%), 9 (27%), 17 (63%), 24 (37%)]
Awed	1, 2, 5, 25, [4 (21%), 20 (62%), 26 (56%)]

AUs used by a subset of the subjects are shown in brackets with the percentage of the subjects using this less common AU in parentheses. The underlined AUs listed in the compound emotions are present in both their basic categories. An asterisk (*) indicates the AU does not appear in either of the two subordinate categories.

We note obvious and unexpected production similarities between some compound expressions. Not surprisingly, the prototypical AUs of hatred and appalled are the same, because they are both variations of angrily disgusted that can only be detected by the strength in the activation of their AUs. More interestingly, there is a noticeable difference in over half the subjects who use AU 7 (eyelid tightener) when expressing hate. Also interesting is the difference between the expression of these two categories and that of angrily disgusted, where AU 17 (chin raiser) is prototypical. These differences make the three facial expressions distinct from one another.

The facial expression of sadly angry does not include any prototypical AU unique to anger, although its image seems to

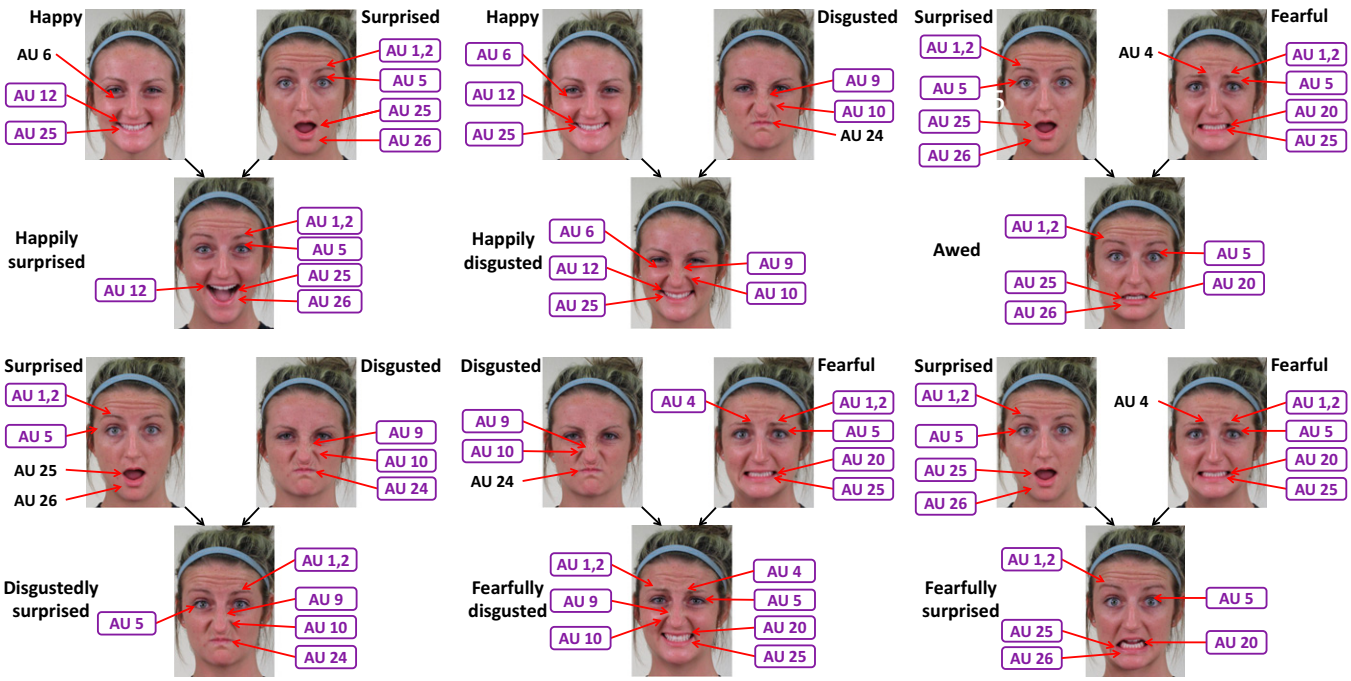

Fig. 2. Shown here are the AUs of six compound facial expressions of emotion. The AUs of the basic emotions are combined as shown to produce the compound category. The AUs of the basic expressions kept to produce the compound emotion are marked with a bounding box. These relationships define the subordinate classes of each category and their interrelatedness. In turn, these results define possible confusion of the compound emotion categories by their subordinates and vice versa.

express anger quite clearly (Fig. 1*K*). Similarly, sadly fearful does not include any prototypical AU unique to sadness, but its image is distinct from that of fear (Fig. 1 *D* and *J*).

Automatic Fiducial Detections. To properly detect facial landmarks, it is imperative we train the system using independent databases before we test it on all of the images of the dataset described in this work. To this end, we used 896 images from the AR face database (16), 600 images from the XM2VTS database (17), and 530 images from the facial expressions of American Sign Language presented in ref. 18, for a total of $2,026$ independent training images.

The problem with previous fiducial detection algorithms is that they assume the landmark points are visually salient. Many face areas are, however, homogeneous and provide only limited information about the shape of the face and the location of each fiducial. One solution to this problem is to add additional constraints (19). A logical constraint is to learn the relationship between landmarks—i.e., estimate the distributions between each pair of fiducials. This approach works as follows. The algorithm of ref. 18 is used to learn the local texture of the 94 facial landmarks seen in Fig. 3; this provides the distribution that permits the detection of each face landmark. We also compute the distribution defining the pairwise position of each two landmark points. These distributions provide additional constraints on the location of each fiducial pair. For example, the left corner of the mouth provides information on where to expect to see the right corner of the mouth. The joint probability of the 94 fiducials is

$$P(\mathbf{z}) = \prod_{i=1}^{93} \prod_{j=i+1}^{94} w_{ij} p_{ij}(\mathbf{z}_i, \mathbf{z}_j),$$

where $\mathbf{z}$ defines the location of each landmark point $\mathbf{z}_i \in \mathbb{R}^2$, $p_{ij}(.)$ is the learned probability density function (pdf) defining the distribution of landmark points $\mathbf{z}_i$ and $\mathbf{z}_j$ as observed in the training set (i.e., the set of $2,026$ training images defined above), and w_{ij} is a weight that determines the relevance of each pair. We assume the pdf of this model is Normal and that the weights are inverse proportional to the distance between fiducials $\mathbf{z}_i$ and $\mathbf{z}_j$. The solution is given by maximizing the above equation. Sample results of this algorithm are shown in Fig. 3. These fiducials define the external and internal shape of the face, because research has shown both external and internal features are important in face recognition (20).

Fig. 3. Shown here are two sample detection results on faces with different identities and expressions. Accurate results are obtained even under large face deformations. Ninety-four fiducial points to define the external and internal shape of the face are used.

Quantitative results of this pairwise optimization approach are given in Table 2, where we see that it yields more accurate results than other state of the art algorithms (21, 22). In fact, these results are quite close to the detection errors obtained with manual annotations, which are known to be between 4.1 and 5.1 pixels in images of this complexity (18). The errors in the table indicate the average pixel distance (in the image) from the automatic detection and manual annotations obtained by the authors.

Image Similarity. We derive a computational model of face perception based on what is currently known about the representation of facial expressions of emotion by humans (i.e., spatial frequencies and configural features) and modern computer vision and machine learning algorithms. Our goal is not to design an algorithm for the automatic detection of AUs, but rather determine whether the images of the 21 facial expressions of emotion (plus neutral) in Fig. 1 are visually discriminable.

In computer vision one typically defines reflectance, albedo, and shape of an image using a set of filter responses on pixel information (23–26). Experiments with human subjects demonstrate that reflectance, albedo, and shape play a role in the recognition of the emotion class from face images, with an emphasis on the latter (20, 27–29). Our face space will hence be given by shape features and Gabor filter responses. Before computing our feature space, all images are cropped around the face and downsized to 400×300 ($h \times w$) pixels.

The dimensions of our feature space defining the face shape are given by the subtraction of the pairwise image features. More formally, consider two fiducial points, $\mathbf{z}_i$ and $\mathbf{z}_j$, with $i \neq j$, i and $j = \{1, \ldots, n\}$, n the number of detected fiducials in an image, $\mathbf{z}_i = (z_{i1}, z_{i2})^T$, and z_{ik} the two components of the fiducial; their horizontal and vertical relative positions are $d_{ijk} = z_{ik} - z_{jk}$, $k = 1, 2$. Recall, in our case, $n = 94$. With 94 fiducials, we have $2 \cdot (94 \cdot 93)/2 = 8,742$ features (dimensions) defining the shape of the face. These interfiducial relative positions are known as configural (or second-order) features and are powerful categorizers of emotive faces (28).

A common way to model the primary visual system is by means of Gabor filters, because cells in the mammalian ventral pathway have responses similar to these (30). Gabor filters have also been successfully applied to the recognition of the six basic emotion categories (24, 31). Herein, we use a bank of 40 Gabor filters at five spatial scales (4:16 pixels per cycle at 0.5 octave steps) and eight orientations ($\theta = \{r\pi/8\}_{r=0}^{7}$). All filter (real and imaginary) components are applied to the 94 face landmarks, yielding $2 \cdot 40 \cdot 94 = 7,520$ features (dimensions). Borrowing terminology from computer vision, we call this resulting feature space the appearance representation.

Classification is carried out using the nearest-mean classifier in the subspace obtained with kernel discriminant analysis (see *Materials and Methods*). In general, discriminant analysis algorithms are based on the simultaneous maximization and minimization of two metrics (32). Two classical problems with the definition of these metrics are the selection of an appropriate pdf that can estimate the true underlying density of the data, and the homoscedasticity (i.e., same variance) assumption. For instance, if every class is defined by a single multimodal Normal distribution with common covariance matrix, then the nearest-mean classifier provides the Bayes optimal classification boundary in the subspace defined by linear discriminant analysis (LDA) (33).

Kernel subclass discriminant analysis (KSDA) (34) addresses the two problems listed above. The underlying distribution of each class is estimated using a mixture of Normal distributions, because this can approximate a large variety of densities. Each model in the mixture is referred to as a subclass. The kernel trick is then used to map the original class distributions to a space $\mathcal{F}$ where these can be approximated as a mixture of homoscedastic Normal distributions (*Materials and Methods*). In machine learning,

Table 2. Average detection error of three different algorithms for the detection of the 94 fiducials of Fig. 3

Method	Overall	Eyes	Eyebrows	Nose	Mouth	Face outline
AAM with RIK (21)	6.349	4.516	7.298	5.634	7.869	**6.541**
Manifold approach (22)	7.658	6.06	10.188	6.796	8.953	7.054
Pairwise optimization approach	**5.395**	**2.834**	**5.432**	**3.745**	**5.540**	9.523

The overall detection error was computed using the 94 face landmarks. Subsequent columns provide the errors for the landmarks delineating each of the internal facial components (i.e., eyes, brows, nose, and mouth) and the outline of the face (i.e., jaw line). Errors are given in image pixels (i.e., the average number of image pixels from the detection given by the algorithm and that obtained manually by humans). Boldface specifies the lowest detection errors.

the kernel trick is a method for mapping data from a Hilbert space to another of intrinsically much higher dimensionality without the need to compute this computationally costly mapping. Because the norm in a Hilbert space is given by an inner product, the trick is to apply a nonlinear function to each feature vector before computing the inner product (35).

For comparative results, we also report on the classification accuracies obtained with the multiclass support vector machine (mSVM) of ref. 36 (see *Materials and Methods*).

Basic Emotions. We use the entire database of $1,610$ images corresponding to the seven classes (i.e., six basic emotions plus neutral) of the 230 identities. Every image is represented in the shape, appearance, or the combination of shape and appearance feature spaces. Recall $d = 8,742$ when we use shape, $7,520$ when using appearance, and $16,262$ when using both.

We conducted a 10-fold cross-validation test. The successful classification rates were 89.71% (with SD 2.32%) when using shape features, 92% (3.71%) when using appearance features, and 96.86% (1.96%) when using both (shape and appearance). The confusion table obtained when using the shape plus appearance feature spaces is in Table 3. These results are highly correlated (0.935) with the confusion tables obtained in a seven-alternative forced-choice paradigm with human subjects (37). A leave-one-sample-out test yielded similar classification accuracies: 89.62% (12.70%) for shape, 91.81% (11.39%) for appearance, and 93.62% (9.73%) for shape plus appearance. In the leave-one-sample-out test, all sample images but one are used for training the classifier, and the left out sample is used for testing it. With n samples, there are n possible samples that can be left out. In leave-one-sample-out, the average of all these n options is reported.

For comparison, we also trained the mSVM of ref. 36. The 10-fold cross-validation results were 87.43% (2.72%) when using shape features, 85.71% (5.8%) when using appearance features, and 88.67% (3.98%) when using both.

We also provide comparative results against a local-based approach as in (38). Here, all faces are first wrapped to a normalized 250×200-pixel image by aligning the baseline of the eyes and mouth, midline of the nose, and left most, right most, upper and lower face limits. The resulting face images are divided in multiple local regions at various scales. In particular, we use partially overlapping patches of 50×50, 100×100, and 150×150 pixels. KSDA and the nearest-mean classifier are used as above, yielding an overall classification accuracy of 83.2% (4%), a value similar to that given by the mSVM and significantly lower than the one obtained by the proposed computational model.

Compound Emotions. We calculated the classification accuracies for the 5,060 images corresponding to the 22 categories of basic and compound emotions (plus neutral) for the 230 identities in our database. Again, we tested using 10-fold cross-validation and leave one out. Classification accuracies in the 10-fold cross-validation test were 73.61% (3.29%) when using shape features only, 70.03% (3.34%) when using appearance features, and 76.91% (3.77%) when shape and appearance are combined in a single feature space. Similar results were obtained using a leave-one-sample-out test: 72.09% (14.64%) for shape, 67.48% (14.81%) for appearance, and 75.09% (13.26%) for shape and appearance combined. From these results it is clear that when the number of classes grows, there is little classification gain when combining shape and appearance features, which suggests the discriminant information carried by the Gabor features is, for the most part, accounted for by the configural ones.

Table 4 shows the confusions made when using shape and appearance. Note how most classification errors are consistent with the similarity in AU activation presented earlier. A clear example of this is the confusion between fearfully surprised and awed (shown in magenta font in Table 4). Also consistent with the AU analysis of Table 1, fearfully surprised and fearfully disgusted are the other two emotions with lowest classification rates (also shown in magenta fonts). Importantly, although hate and appall represent similar compound emotion categories, their AUs are distinct and, hence, their recognition is good (shown in yellow fonts). The correlation between the production and recognition results (Tables 1 and 4) is 0.667 (see *Materials and Methods*).

The subordinate relationships defined in Table 1 and Fig. 2 also govern how we perceive these 22 categories. The clearest example is angrily surprised, which is confused 11% of the time for disgust; this is consistent with our AU analysis. Note that two of the three prototypical AUs in angrily disgusted are also used to express disgust.

The recognition rates of the mSVM of (36) for the same 22 categories using 10-fold cross-validation are 40.09% (5.19%) for shape, 35.27% (2.68%) for appearance, and 49.79% (3.64%) for the combination of the two feature spaces. These results suggest that discriminant analysis is a much better option than multiclass SVM when the number of emotion categories is large. The overall classification accuracy obtained with the local-approach of ref. 38 is 48.2% (2.13%), similar to that of the mSVM but much lower than that of the proposed approach.

It is also important to know which features are most useful to discriminate between the 22 categories defined in the present work; this can be obtained by plotting the most discriminant

Table 3. Confusion matrix for the categorization of the six basic emotion categories plus neutral when using shape and appearance features

	Neutral	Happiness	Sadness	Fear	Anger	Surprise	Disgust
Neutral	**0.967**	0	0.033	0	0	0	0
Happiness	0	**0.993**	0	0.007	0	0	0
Sadness	0.047	0.013	**0.940**	0	0	0	0
Fear	0.007	0	0	**0.980**	0	0.013	0
Anger	0	0	0.007	0	**0.953**	0	0.040
Surprise	0	0.007	0	0.020	0	**0.973**	0
Disgust	0.007	0	0.007	0	0.013	0	**0.973**

Rows, true category; columns, recognized category. Boldface specifies the best recognized categories.

Table 4. Confusion table for the 21 emotion categories plus neutral

0.1 1

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
a	.96	0	.03	0	0	0	0	0	0	0	0	0	.01	0	0	0	0	0	0	0	.01	0
b	0	.97	0	0	0	0	0	.01	.01	.01	0	0	0	0	0	0	0	0	0	0	0	0
c	.03	.01	.87	0	.02	0	0	0	0	0	.02	0	.01	0	0	0	0	0	.01	.01	.03	0
d	0	0	0	.87	.01	.01	.01	.01	0	.03	.01	0	0	.01	.03	0	0	0	.01	0	0	0
e	0	0	.02	0	.77	0	.06	0	0	0	.04	0	0	.01	0	0	0	0	0	.03	.07	0
f	0	0	0	.01	0	.92	0	0	0	0	0	.02	0	0	.01	0	.01	0	0	0	0	.03
g	0	.01	.01	.02	.03	0	.81	0	.01	.01	.04	0	0	.01	0	0	0	.01	.02	.03	0	0
h	0	.02	0	.02	0	0	0	.93	.01	0	0	0	0	0	0	.01	0	0	.01	0	0	.01
i	0	.03	0	.01	0	0	.01	.03	.91	.01	0	0	0	0	0	.01	0	.01	0	0	0	0
j	0	.01	0	0	0	0	0	0	0	.69	0	.01	.03	.11	.03	.03	.01	0	0	.01	0	.08
k	0	0	.01	0	.06	0	.03	0	.01	0	.85	0	.01	0	0	0	0	.01	.01	0	.01	0
l	0	0	0	0	0	0	0	0	0	.02	0	.77	0	.01	0	0	.19	0	0	0	.01	.01
m	0	0	.01	0	0	0	.01	0	0	.07	.01	0	.77	.08	0	.03	0	.01	0	0	0	.01
n	0	.01	.01	0	0	0	0	0	0	.16	0	.01	.05	.61	.02	.11	.01	0	.01	0	0	.01
o	0	0	.01	0	0	.01	0	0	0	.09	0	0	0	.05	.54	.01	.01	0	0	0	0	.29
p	0	0	0	0	0	0	0	0	.01	.17	0	0	.05	.15	.03	.58	0	0	0	0	0	.01
q	0	0	0	0	0	0	0	0	0	0	0	.27	0	.02	0	0	.71	0	0	0	0	0
r	0	0	0	0	.01	0	.11	0	0	0	.15	0	0	.01	0	.01	0	.70	0	0	.01	0
s	0	0	.05	.01	0	0	.01	0	0	0	.01	0	0	.01	0	.01	.01	0	.83	.02	.03	.01
t	0	0	0	0	.02	0	.01	0	0	0	.02	0	.02	.01	0	0	.01	0	0	.63	.29	0
u	.01	0	.01	0	.03	0	0	0	0	.01	.01	.01	.01	0	0	0	.01	0	0	.25	.65	0
v	0	0	0	0	0	0	0	0	0	.10	0	.01	0	.01	.28	.02	.01	0	0	0	0	.58

Cells with a higher detection rate than 0.1 (or 10%) have been colored blue, with darker colors indicating higher percentages. a, neutral; b, happy; c, sad; d, fearful; e, angry; f, surprised; g, disgusted; h, happily surprised; i, happily disgusted; j, sadly fearful; k, sadly angry; l, sadly surprised; m, sadly disgusted; n, fearfully angry; o, fearfully surprised; p, fearfully disgusted; q, angrily surprised; r, angrily disgusted; s, disgustedly surprised; t, appalled; u, hate; v, awed. Rows, true category; columns, recognized category.

features given by the eigenvector $\mathbf{v}_1$ of the LDA equation $\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{V}=\mathbf{V}\Lambda$, where $\mathbf{S}_W=\sum_{i=1}^{c}\sum_{k=1}^{m_i}(\mathbf{x}_{ik}-\mu_i)(\mathbf{x}_{ik}-\mu_i)^T$ is the within-class scatter matrix, $\mathbf{S}_B=m^{-1}\sum_{i=1}^{c}m_i(\mu_i-\mu)(\mu_i-\mu)^T$ is the between-class scatter matrix, $\mathbf{V}=(\mathbf{v}_1,\ldots,\mathbf{v}_p)$ is the matrix whose columns are the eigenvectors of the above equation, and $\Lambda=diag(\lambda_1,\ldots,\lambda_p)$ are the corresponding eigenvalues, with $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_p\geq 0$. Here, we used the eigenvector of LDA because that of KSDA cannot be used due to the preimage problem.

Fig. 4. Most discriminant configural features. The line color specifies the discriminability of the feature, with darker lines discriminating more, and lighter lines less. The first 22 results in these figures are for the pairwise classification: (*A*) neutral vs. other categories, (*B*) happy vs. other categories, (*C*) sad vs. other categories, (*D*) fearful vs. other categories, (*E*) angry vs. other categories, (*F*) surprised vs. other categories, (*G*) disgusted vs. other categories, (*H*) happily surprised vs. other categories, (*I*) happily disgusted vs. other categories, (*J*) sadly fearful vs. other categories, (*K*) sadly angry vs. other categories, (*L*) sadly surprised vs. other categories, (*M*) sadly disgusted vs. other categories, (*N*) fearfully angry vs. other categories, (*O*) fearfully surprised vs. other categories, (*P*) fearfully disgusted vs. other categories, (*Q*) angrily surprised vs. other categories, (*R*) angrily disgusted vs. other categories, (*S*) disgustedly surprised vs. other categories, (*T*) appalled vs. other categories, (*U*) hate vs. other categories, and (*V*) awe vs. other categories. In *W*, we show the most discriminant configural features for the classification of all 22 emotion categories combined.

Because the configural (shape) representation yielded the best results, we compute the eigenvector $\mathbf{v}_1^{shape}$ using its representation, i.e., $p = 8{,}742$. Similar results are obtained with the Gabor representation (which we called appearance). Recall that the entries of $\mathbf{v}_1$ correspond to the relevance of each of the p features, conveniently normalized to add up to 1. The most discriminant features are selected as those adding up to 0.7 or larger (i.e., $\geq 70\%$ of the discriminant information). Using this approach, we compute the most discriminant features in each category by letting $c = 2$, with one class including the samples of the category under study and the other class with the samples of all other categories. The results are plotted in Fig. 4 *A–V* for each of the categories. The lines superimposed on the image specify the discriminant configural features. The color (dark to light) of the line is proportional to the value in $\mathbf{v}_1^{shape}$. Thus, darker lines correspond to more discriminant features, lighter lines to less discriminant features. In Fig. 4*W* we plot the most discriminant features when considering all of the 22 separate categories of emotion, i.e., $c = 22$.

Discussion

The present work introduced an important type of emotion categories called compound emotions, which are formed by combining two or more basic emotion categories, e.g., happily surprised, sadly fearful, and angrily disgusted. We showed how some compound emotion categories may be given by a single word. For example, in English, hate, appalled, and awe define three of these compound emotion categories. In Chinese, there are compound words used to describe compound emotions such as hate, happily surprised, sadly angry, and fearfully surprised.

We defined 22 categories, including 6 basic and 15 compound facial expressions of emotion, and provided an in-depth analysis of its production. Our analysis includes a careful manual FACS coding (Table 1); this demonstrates that compound categories are clearly distinct from the basic categories forming them at the production level, and illustrates the similarities between some compound expressions. We then defined a computational model for the automatic detections of key fiducial points defining the shape of the external and internal features of the face (Fig. 3). Then, we reported on the automatic categorization of basic and compound emotions using shape and appearance features, Tables 3 and 4. For shape, we considered configural features. Appearance is defined by Gabor filters at multiple scales and orientations. These results show that configural features are slightly better categorizers of facial expressions of emotion and that the combination of shape and appearance does not result in a significant classification boost. Because the appearance representation is dependent on the shape but also the reflectance and albedo of the face, the above results suggest that configural (second-order) features are superior discriminant measurements of facial expressions of basic and compound emotions.

Finally, we showed that the most discriminant features are also consistent with our AU analysis. These studies are essential before we can tackle complex databases and spontaneous expressions, such as those of ref. 39. Without an understanding of which AUs represent each category of emotion, it is impossible to understand naturalistic expressions and address fundamental problems in neuroscience (40), study psychiatric disorders (41), or design complex perceptual interfaces (42).

Fig. 4 shows the most discriminant configural features. Once more, we see that the results are consistent with the FACS analysis reported above. One example is the facial expression of happiness; note how its AU activation correlates with the results shown in Fig. 4*B*. Thick lines define the upper movement of the cheeks (i.e., cheek raiser, AU 6), the outer pulling of the lip corners (AU 12), and the parting of the lips (AU 25). We also see discriminant configural features that specify the squinting of the subject's right eye, which is classical of the Duchenne smile (3); these are due to AU 6, which wrinkles the skin, diminishing the intradistance between horizontal eye features.

Note also that although the most discriminant features of the compound emotion categories code for similar AUs than those of the subordinate basic categories, the actual discriminant configural features are not the same. For instance, happily surprised (Fig. 4*H*) clearly code for AU 12, as does happiness (Fig. 4*B*), but using distinct configural features; this suggests that the expression of compound emotions differs slightly from the expression of subordinate categories, allowing us (and the computational algorithms defined herein) to distinguish between them. Another interesting case is that of sadly angry. Note the similarity of its most discriminant configural features with those of angrily disgusted, which explains the small confusion observed in Table 4.

The research on the production and perception of compound emotion categories opens a new area of research in face recognition that can take studies of human cognition, social communication, and the design of computer vision and human–computer interfaces to a new level of complexity. A particular area of interest is the perception of facial expressions of compound emotions in psychiatric disorders (e.g., schizophrenia), social and cognitive impairments (e.g., Autism spectrum disorder), and studies of pain. Also of interest is to study cultural influences in the production and perception of compound facial expressions of emotion. And a fundamental question that requires further investigation is whether the cognitive representation and cognitive processes involved in the recognition of facial expressions are the same or different for basic and compound emotion categories.

Materials and Methods

Database Collection. ***Subjects.*** A total of 230 human subjects (130 females; mean age 23; SD 6) were recruited from the university area, receiving a small monetary reward for participating. Most ethnicities and races were included, and Caucasian, Asian, African American, and Hispanic are represented in the database. Facial occlusions were minimized, with no eyeglasses or facial hair. Subjects who needed corrective lenses wore contacts. Male subjects were asked to shave their face as cleanly as possible. Subjects were also asked to uncover their forehead to fully show their eyebrows.

Procedure. Subjects were seated 4 ft away from a Canon IXUS 110 camera and faced it frontally. A mirror was placed to the left of the camera to allow subjects to practice their expressions before each acquisition. Two 500-W photography hot lights were located at 50° left and right from the midline passing through the center of the subject and the camera. The light was diffused with two inverted umbrellas, i.e., the lights pointed away from the subject toward the center of the photography umbrellas, resulting in a diffuse light environment.

The experimenter taking the subject's pictures suggested a possible situation that may cause each facial expression, e.g., disgust would be expressed when smelling a bad odor. This was crucial to correctly produce compound emotions. For example, happily surprised is produced when receiving wonderful, unexpected news, whereas angrily surprised is expressed when a person does something unexpectedly wrong to you. Subjects were also shown a few sample pictures. For the six basic emotions, these sample images were selected from refs. 7 and 43. For the compound emotions, the exemplars were pictures of the authors expressing them and synthetic constructs from images of refs. 7 and 43. Subjects were not instructed to try to look exactly the same as the exemplar photos. Rather, subjects were encouraged to express each emotion category as clearly as possible while expressing their meaning (i.e., in the example situation described by the experimenter). A verbal definition of each category accompanies the sample picture. Then the suggested situation was given. Finally, the subject produced the facial expression. The photos were taken at the apex of the expression. Pictures taken with the Canon IXUS are color images of $4{,}000 \times 3{,}000$ ($h \times w$) pixels.

KSDA Categorization. Formally, let m be the number of training samples, and c the number of classes. KSDA uses the kernel between-subclass scatter matrix and the kernel covariance matrix as metrics to be maximized and minimized, respectively. These two metrics are given by

$\Sigma_B^\Phi = \sum_{i=1}^{c-1}\sum_{j=1}^{h_i}\sum_{l=i+1}^{c}\sum_{q=1}^{h_l} p_{ij}p_{lq}(\mu_{ij}^\phi - \mu_{lq}^\phi)(\mu_{ij}^\phi - \mu_{lq}^\phi)^T$ and $\Sigma_X^\Phi = \sum_{i=1}^{c}\sum_{j=1}^{h_i}\Sigma_{ij}^\Phi = \sum_{i=1}^{c}\sum_{j=1}^{h_i} m_{ij}^{-1}\sum_{k=1}^{m_{ij}}(\phi(\mathbf{x}_{ijk}) - \mu^\phi)(\phi(\mathbf{x}_{ijk}) - \mu^\phi)^T$, where $\phi(\cdot) : \mathbb{R}^p \to \mathcal{F}$ defines the mapping from the original feature space of d dimensions to the kernel space $\mathcal{F}$, $\mathbf{x}_{ijk}$ denotes the k^{th} sample in the j^{th} subclass in class i, $p_{ij} = m_{ij}/m$ is the prior of the j^{th} subclass of class i, m_{ij} is the number of samples in the j^{th} subclass of class i, h_i is the number of subclasses in class i, $\mu_{ij}^\phi = m_{ij}^{-1}\sum_{k=1}^{m_{ij}}\phi(\mathbf{x}_{ijk})$ is the kernel sample mean of the j^{th} subclass in class i, and $\mu^\phi = m^{-1}\sum_{i=1}^{c}\sum_{j=1}^{h_i}\sum_{k=1}^{m_{ij}}\phi(\mathbf{x}_{ijk})$ is the global sample mean in the kernel space. Herein, we use the radial basis function to define our kernel mapping, i.e., $k(\mathbf{x}_{ijk},\mathbf{x}_{lpq}) = \exp\left(-\frac{\|\mathbf{x}_{ijk} - \mathbf{x}_{lpq}\|^2}{\sigma^2}\right)$.

KSDA maps the original feature spaces to a kernel space where the following homoscedastic criterion is maximized (34):

$$Q(\phi, h_1, \ldots, h_C) = \frac{1}{h}\sum_{i=1}^{C-1}\sum_{j=1}^{h_i}\sum_{l=i+1}^{C}\sum_{q=1}^{h_k}\frac{tr\left(\Sigma_{ij}^\Phi \Sigma_{lq}^\Phi\right)}{tr\left(\Sigma_{ij}^{\Phi^2}\right) + tr\left(\Sigma_{lq}^{\Phi^2}\right)},$$

where Σ_{ij}^Φ is the sample covariance matrix of the j^{th} subclass of class i (as defined above), and h is the number of summing terms. As a result, classification based on the nearest mean approximates that of the Bayes classifier.

The nearest-mean classifier assigns to a test sample $\mathbf{t}$ the class of the closest subclass mean, i.e., $\arg_i \min_{i,j}\left\|\mathbf{t}^\phi - \mu_{ij}^\phi\right\|_2$, where $\|.\|_2$ is the 2-norm of a vector; this is done in the space defined by the basis vectors of KSDA.

SVM Categorization. We compared the proposed classification approach with that given by the mSVMs of ref. 36. Though many SVM algorithms are defined for the two-class problem, this approach can deal with any number of classes. Formally, let the training data of the c^{th} class be

$$D_c = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathbb{R}^p, y_i = \{1, -1\}\}_{i=1}^m,$$

where $\mathbf{x}_i$ is the p-dimensional feature vector defining the i^{th} sample image (with $p = 8{,}742$ and 7.520 when using only shape or appearance, and 16,262 when both are considered simultaneously), m is the number of samples, $y_i = 1$ specifies the training feature, vector $\mathbf{x}_i$ belongs to category c , and $y_i = -1$ indicates it belongs to one of the other classes.

SVM seeks a discriminant function $f(\mathbf{x}_i) = h(\mathbf{x}_i) + b$, where $f(.) : \mathbb{R}^p \to \mathbb{R}$, $h \in \mathcal{H}$ is a function defined in a reproducing kernel Hilbert space (RKHS) and $b \in \mathbb{R}$ (44). Here, the goal is to minimize the following objective function:

$$\frac{1}{m}\sum_{i=1}^{m}(1 - y_i f(\mathbf{x}_i))_+ + \lambda\|h\|_{\mathcal{H}}^2,$$

where $(a)_+ = a$ if $a > 0$ and 0 otherwise ($a \in \mathbb{R}$), and $\|\cdot\|_{\mathcal{H}}$ is the norm defined in the RKHS. Note that in the objective function thus defined, the first term computes the misclassification cost of the training samples, whereas the second term measures the complexity of its solution.

It has been shown (45) that for some kernels (e.g., splines, high-order polynomials), the classification function of SVM asymptotically approximates the function given by the Bayes rule. This work was extended by ref. 36 to derive an mSVM. In a c class problem, we now define the i^{th} training sample as $(\mathbf{x}_i, \mathbf{y}_i)$, where $\mathbf{y}_i$ is a c-dimensional vector with a 1 in the l position and $-1/(c-1)$ elsewhere, l is the class label of $\mathbf{x}_i$ ($l \in 1, \ldots, c$). We also define the cost function $L(\mathbf{y}_i) : \mathbb{R}^c \to \mathbb{R}^c$, which maps the vector $\mathbf{y}_i$ to a vector with a zero in the l^{th} entry and ones everywhere else.

The goal of mSVM is to simultaneously learn a set of c functions $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_c(\mathbf{x}))^T$, with the constraint $\sum_{j=1}^{k} f_j(\mathbf{x}) = 0$; this corresponds to the following optimization problem (36):

$$\min_{\mathbf{f}} \frac{1}{m}\sum_{i=1}^{m} L(\mathbf{y}_i)\cdot(\mathbf{f}(\mathbf{x}_i) - \mathbf{y}_i)_+ + \frac{\lambda}{2}\sum_{j=1}^{c}\|h_j\|_{\mathcal{H}}^2$$

$$\text{subject to} \sum_{j=1}^{c} f_j(\mathbf{x}) = 0,$$

where $f_j(\mathbf{x}) = h_j(\mathbf{x}) + b$ and $h_j \in \mathcal{H}$. This approach approximates the Bayes solution when the number of samples m increases to infinity. This result is especially useful when there is no dominant class or the number of classes is large.

Correlation Analyses. The first correlation analysis was between the results of the derived computational model (shown in Table 3) and those reported in ref. 37. To compute this correlation, the entries of the matrix in Table 3 were written in vector form by concatenating consecutive rows together. The same procedure was done with the confusion table of ref. 37. These two vectors were then norm normalized. The inner product between the resulting vectors defines their correlation.

The correlation between the results of the computational model (Table 4) and the FACS analysis (Table 1) was estimated as follows. First, a table of the AU similarity between every emotion category pair (plus neutral) was obtained from Table 1; this resulted in a 22×22 matrix, whose i,j entry defines the AU similarity between emotion categories i and j ($i,j = 1, \ldots, 22$). The i,j^{th} entry is given by

$$\frac{1}{s}\sum_{k=1}^{s}\left[1 - \frac{|u_i(\text{AU } k) - u_j(\text{AU } k)|}{\max(u_i(\text{AU } k), u_j(\text{AU } k))}\right],$$

where $u_i(\text{AU } k)$ is the number of images in the database with AU k present in the facial expression of emotion category i, and s is the number of AUs used to express emotion categories i and j. The resulting matrix and Table 4 are written in vector form by concatenating consecutive rows, and the resulting vectors are norm normalized. The correlation between the AU activation of two distinct emotion categories and the recognition results of the computational model is given by the inner product of these normalized vectors. When computing the above equation, all AUs present in emotion categories i and j were included, which yielded a correlation of 0.667. When considering the major AUs only (i.e., when omitting those within the parentheses in Table 1), the correlation was 0.561.

ACKNOWLEDGMENTS. We thank the reviewers for constructive comments. This research was supported in part by National Institutes of Health Grants R01-EY-020834 and R21-DC-011081.

1. Aristotle, *Minor Works*, trans Hett WS (1936) (Harvard Univ Press, Cambridge, MA).
2. Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull* 115(1):102–141.
3. Duchenne CB (1862). *The Mechanism of Human Facial Expression* (Renard, Paris); reprinted (1990) (Cambridge Univ Press, London).
4. Borod JC, ed (2000) *The Neuropsychology of Emotion* (Oxford Univ Press, London).
5. Martinez AM, Du S (2012) A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *J Mach Learn Res* 13: 1589–1608.
6. Darwin C (1965) *The Expression of the Emotions in Man and Animals* (Univ of Chicago Press, Chicago).
7. Ekman P, Friesen WV (1976) *Pictures of Facial Affect* (Consulting Psychologists Press, Palo Alto, CA).
8. Russell JA (2003) Core affect and the psychological construction of emotion. *Psychol Rev* 110(1):145–172.
9. Izard CE (2009) Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annu Rev Psychol* 60:1–25.
10. Ekman P (1992) An argument for basic emotions. *Cogn Emotion* 6(3-4):169–200.
11. Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF (2012) The brain basis of emotion: A meta-analytic review. *Behav Brain Sci* 35(3):121–143.
12. Ekman P, Friesen WV (1978) *Facial Action Coding System: A Technique for the Measurement of Facial Movement* (Consulting Psychologists Press, Palo Alto, CA).
13. Kohler CG, et al. (2004) Differences in facial expressions of four universal emotions. *Psychiatry Res* 128(3):235–244.
14. Hamm J, Kohler CG, Gur RC, Verma R (2011) Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *J Neurosci Methods* 200(2):237–256.
15. Seider BH, Shiota MN, Whalen P, Levenson RW (2011) Greater sadness reactivity in late life. *Soc Cogn Affect Neurosci* 6(2):186–194.
16. Martinez AM, Benavente R (1998) The AR Face Database. CVC Technical Report no. 24 (Computer Vision Center, Univ of Alabama, Birmingham, AL).
17. Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999). XM2VTSDB: The Extended M2VTS Database. *Proceedings of the Second International Conference on Audio- and Video-Based Biometric Person Authentication* (Springer, Heidelberg), pp. 72–77.
18. Ding L, Martinez AM (2010) Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *IEEE Trans Pattern Anal Mach Intell* 32(11):2022–2038.
19. Benitez-Quiroz CF, Rivera S, Gotardo PF, Martinez AM (2014) Salient and non-salient fiducial detection using a probabilistic graph model. *Pattern Recognit* 47(1):208–215.
20. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc IEEE* 94(11):1948–1962.
21. Hamsici OC, Martinez AM (2009). Active appearance models with rotation invariant kernels. *IEEE International Conference on Computer Vision*, 10.1109/ICCV.2009.5459365.
22. Rivera S, Martinez AM (2012) Learning deformable shape manifolds. *Pattern Recognit* 45(4):1792–1801.
23. De la Torre F, Cohn JF (2011). Facial expression analysis. *Guide to Visual Analysis of Humans: Looking at People*, eds Moeslund TB, et al (Springer, New York), pp 377–410.

24. Bartlett MS, et al. (2005). Recognizing facial expression: Machine learning and application to spontaneous behavior. *IEEE Comp Vis Pattern Recog* 2:568–573.
25. Simon T, Nguyen MH, De La Torre F, Cohn JF (2010). Action unit detection with segment-based SVMS. *IEEE Comp Vis Pattern Recog*, 10.1109/CVPR.2010.5539998.
26. Martínez AM (2003) Matching expression variant faces. *Vision Res* 43(9):1047–1060.
27. Etcoff NL, Magee JJ (1992) Categorical perception of facial expressions. *Cognition* 44 (3):227–240.
28. Neth D, Martinez AM (2009) Emotion perception in emotionless face images suggests a norm-based representation. *J Vis* 9(1):1–11.
29. Pessoa L, Adolphs R (2010) Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nat Rev Neurosci* 11(11):773–783.
30. Daugman JG (1980) Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res* 20(10):847–856.
31. Lyons MJ, Budynek J, Akamatsu S (1999) Automatic classification of single facial images. *IEEE Trans Pattern Anal Mach Intell* 21(12):1357–1362.
32. Martinez AM, Zhu M (2005) Where are linear feature extraction methods applicable? *IEEE Trans Pattern Anal Mach Intell* 27(12):1934–1944.
33. Fisher RA (1938) The statistical utilization of multiple measurements. *Ann Hum Genet* 8(4):376–386.
34. You D, Hamsici OC, Martinez AM (2011) Kernel optimization in discriminant analysis. *IEEE Trans Pattern Anal Mach Intell* 33(3):631–638.
35. Wahba G (1990) *Spline Models for Observational Data* (Soc Industrial and Applied Mathematics, Philadelphia).
36. Lee Y, Lin Y, Wahba G (2004) Multicategory support vector machines. *J Am Stat Assoc* 99(465):67–81.
37. Du S, Martinez AM (2011) The resolution of facial expressions of emotion. *J Vis* 11(13):24.
38. Martinez AM (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans Pattern Anal Mach Intell* 24(6):748–763.
39. O'Toole AJ, et al. (2005) A video database of moving faces and people. *IEEE Trans Pattern Anal Mach Intell* 27(5):812–816.
40. Stanley DA, Adolphs R (2013) Toward a neural basis for social behavior. *Neuron* 80(3): 816–826.
41. Kennedy DP, Adolphs R (2012) The social brain in psychiatric and neurological disorders. *Trends Cogn Sci* 16(11):559–572.
42. Pentland A (2000) Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans Pattern Anal Mach Intell* 22(1):107–119.
43. Ebner NC, Riediger M, Lindenberger U (2010) FACES—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behav Res Methods* 42(1):351–362.
44. Vapnik V (1999) *The Nature of Statistical Learning Theory* (Springer, New York).
45. Lin Y (2002) Support vector machines and the Bayes rule in classification. *Data Min Knowl Discov* 6(3):259–275.