



# HHS Public Access

Author manuscript

*Curr Dir Psychol Sci.* Author manuscript; available in PMC 2018 June 01.

Published in final edited form as:

*Curr Dir Psychol Sci.* 2017 June ; 26(3): 263–269. doi:10.1177/0963721417698535.

## Computational Models of Face Perception

**Aleix M Martinez**

The Ohio State University

### Abstract

Faces are one of the most important means of communication in humans. For example, a short glance at a person's face provides information on identity and emotional state. What are the computations the brain uses to solve these problems so accurately and seemingly effortlessly? This article summarizes current research on computational modeling, a technique used to answer this question. Specifically, my research studies the hypothesis that this algorithm is tasked to solve the inverse problem of production. For example, to recognize identity, our brain needs to identify shape and shading image features that are invariant to facial expression, pose and illumination. Similarly, to recognize emotion, the brain needs to identify shape and shading features that are invariant to identity, pose and illumination. If one defines the physics equations that render an image under different identities, expressions, poses and illuminations, then gaining invariance to these factors is readily resolved by computing the inverse of this rendering function. I describe our current understanding of the algorithms used by our brains to resolve this inverse problem. I also discuss how these results are driving research in computer vision to design computer systems that are as accurate, robust and efficient as humans.

### Keywords

face recognition; face processing; affect; categorization; language evolution

## Computational Models of Face Perception

After finishing this sentence, look around, find a person you know, and then look briefly at his or her face. What can you tell about this person? Likely the person's name (identity) and emotional state come to mind. Most of us effortlessly extract this information from the smallest of glimpses.

This article reviews our current understanding of the computations that are performed by the brain to achieve these seemingly effortless tasks—visual recognition of identity and emotion. This means that we assume the brain is a type of computer running algorithms specifically dedicated to the interpretation of other people's faces. The goal is to decode and understand these algorithms. This is called computational modeling.

Specifically, this paper details how current progress in computational modeling is helping us understand how the brain recognizes faces.

My use of computational models is based on the hypothesis that the brain is tasked to solve the “inverse problem of image production.” That is, if  $f(\cdot)$  defines how a facial attribute maps

onto an image in the retina, then the brain's goal is to solve the inverse problem,  $f^{-1}(\cdot)$ —how the image on one's retina translates into understanding a facial attribute.

For example, imagine you are looking at Sally's face. Here, the brain's goal is to recover the name "Sally." More formally, the retinal image  $I=f(\text{Sally's face})$ . And the goal is to compute the inverse function,  $\text{Sally} = f^{-1}(I)$ .

The identity of someone's face is engraved in the person's three-dimensional face structure and the reflectance properties of the person's skin. These are examples of some of its diagnostic features. But this is not what we see. Rather, the two-dimensional shape of the face on your retinal image depends on the viewing angle and the person's facial expression. The brain's goal is to uncover the diagnostic features and filter out variations due to expression, viewing angle and illumination.

It is imperative to note that computational modeling is only useful if it identifies these diagnostic features, algorithms and mechanisms involved in the recognition of faces. I show that some machine learning approaches, such as deep learning, are not generally helpful to answer these questions.

## Recognition of Identity

Look at the left image in Figure 1. This is a two-dimensional image,  $I$ . Now, look at the image to its right. This image defines the shape  $s$  of the main facial components of that face.

Given many face images and their shapes ( $\{I_i, s_i\}, i = 1, \dots, n$ ), one can compute the mean shape as well as the major differences (i.e., largest standard deviations) between shapes. These variances are given by Principal Component Analysis (PCA), a statistical technique that allows us to find the shape features that produce maximum variability (Martinez & Kak, 2001). The resulting representation is called a norm-based face space because all faces are described based on their deviation from the mean (norm) sample shape (Leopold et al., 2006), Figure 2.

Recall, however, that our retinal image  $I_i$  is two-dimensional but that diagnostic features exist in three-dimensional space. Can we design an algorithm that estimates the three-dimensional shape of a face from a single image? Yes. In fact, everyday experience proves this. When you looked at a face at the beginning of this article, you probably just saw it from a single viewing angle. Yet, you were able to mentally imagine other views of that face as well.

My students and I have shown that the computations needed to solve this problem are quite simple (Zhao et al., 2016). This algorithm works as follows: Given a set of two-dimensional images and their corresponding three-dimensional shapes ( $\{I_i, S_i\}, i = 1, \dots, n$ ), we use a machine learning technique called regression (You et al., 2014) to learn the (inverse) functional mapping from retinal image to three-dimensional shape,  $S_i = f(I_i)$ , Figure 3. Once this function has been learned using the available training data, we can use it to estimate the three-dimensional shape  $\hat{S}$  of a previously unseen face image  $\hat{I}$ , i.e.,  $\hat{S} = f(\hat{I})$ .

The above modeling shows how to map an initial face image to a rotation-invariant representation. Physiological studies though, suggest the existence of an intermediate representation invariant to mirror images of profile views (Meyers et al., 2016). A simple analysis of projective geometry shows that the basis functions (e.g., PCs) obtained with the above formulation yields the same response for mirror-symmetric images (Leibo et al., 2016). As an intuitive proof, note that faces are symmetric about the vertical midline of the face. Thus, rotation a face  $90^\circ$  to the left and right yield basically the same image, up to a mirror projection.

## Deep Learning

If the function  $f(\cdot)$  presented above is defined by many parameters, the resulting regression approach is called deep learning. In deep learning, one must use a very large number of training samples to successfully estimate that same number of parameters, an increasingly popular approach called big data.

Deep learning has recently achieved good results in several computer vision problems, including face recognition (e.g., Kemelmacher-Shlizerman et al., 2016). Unfortunately, this technique does not generally help us uncover the underlying computations of the algorithm used by our visual system.

For one, we do not yet know how to apply deep learning to solve many problems in face recognition, e.g., recognition under varying illumination or the recognition of emotion.

Also, the reliance on big data makes deep learning an unlikely model of human vision. Humans generally learn from a single sample (Martinez, in press, 2002), not thousands as required by current deep learning algorithms.

And, crucially, deep learning does not generally provide information on the mechanisms used by the brain to decode facial attributes. That is, we might be able to design computer algorithms that identify people's faces very accurately, yet learn nothing about the brain.

To clarify this point, imagine a physicist trying to understand the behavior of a number of particles. Given enough observations of the behavior of these particles, deep learning could certainly be used to identify a function describing their behavior. This function would allow us to predict the state  $y$  of the particles  $x$  after an event  $f(\cdot)$ ,  $y=f(x)$ . But, this would not provide any insights on the mechanisms involved in that process, i.e., the laws of physics.

The same applies to the study of the visual system. It is not sufficient to demonstrate there exists a function that maps an image to a facial attribute. We also wish to uncover the specific computations used by the brain to accomplish this. I argue we need to refocused our research toward computational models that can solve this problem.

For example, Gilad et al. (2009) suggested that the local contrast polarity between a few regions of the face (especially those around the eyes) encode critical information about a person's identity and that the brain uses this information to recognize people's faces. Subsequently, Ohayon et al. (2012) identified cells in the macaque monkey brain that

selectively respond to such contrast variations. Computer vision algorithms based on this local contrast polarity successfully detect and recognize faces in images (Zhao et al., 2016; Turk, 2013). Furthermore, this model explains how one is able to recognize partially occluded and expression variant faces (Martinez & Jia, 2009); for example, using graph matching (Aflalo et al., 2015; Zhao & Martinez, 2016). These results do point toward an understanding of some of the computations of the perception of face images.

## Facial Expressions of Emotion

Another aspect of face perception is our remarkable ability to interpret facial expressions. Facial expressions convey a lot of information, such as a person's emotional state.

As with identity, the representation of facial expressions also uses a norm-based model (Neth & Martinez, 2009, 2010). However, the dimensions employed in the recognition of emotion are, for the most part, different (Sormaz et al., 2016; Richoz et al., 2015; Martinez & Du, 2012). We thus say that the form of the face space is the same for expression and identity but that the dimensions that define this space differ for the two judgments. What are the features that represent these dimensions, then?

Studying the physical reality of expression shows that they are produced by contracting and relaxing different muscles in the face (Duchenne, 2006). Thus, I hypothesize that the brain solves the inverse problem by attempting to decode which facial muscle actions  $h(I)$  are active during a particular expression (Martinez, in press). My research group has recently developed a computer vision system based on this model using an algorithm that accounts for shape and shading features, and incorporating it into machine learning algorithms that identify which of these features best discriminate the muscles involved in each expression (Benitez-Quiroz et al., 2016a; Du et al., 2014).

Take the example of a small cheek muscle that is used to pull the lips outwardly to create a smile. Unsurprisingly, our machine learning approach identified shape and shading changes in the corners of the mouth as the most discriminant feature for smiles. Likewise, contracting a set of three facial muscles located at the top of the face results in the lowering of the inner corners of the brows. Yet, the most discriminant shape and shading features to detect this facial action are associated with more distal parts of the face (i.e., the brow to mouth distance and the face's height/width ratio) because they change when one contracts these muscles (Du et al., 2014). Accordingly, the algorithm assumes these muscles are active when processing the faces of people who have unusually large distances between their brows and mouths plus really thin faces (Martinez, in press).

This effect is clearly visible in Figure 4. In the left image, we see the male character in Wood's painting *American Gothic*. Note that this person is not expressing any emotion. Yet, you are likely to perceive sadness in his expression. Using morphing software, we can decrease the intra brow-mouth distance and make the face wider (right image). Notice now how the face looks angry because we have incidentally altered it to display the image features associated with the facial muscle actions used to express anger.

If this algorithm is indeed implemented in our brains, then there should be an area of the brain dedicated to the detection of these facial muscle actions. In a recent paper, my research group has identified one such region just behind the right ear (Srinivasan et al., 2016).

## Compound Emotions

An ongoing debate in emotion theory is the number of facial expressions that we can visually recognize. Darwin (1965) argued that six facial expressions of emotion can be visually recognized across cultures. However, my group's computational modeling presented above suggests that the visual system does not attempt to categorize facial expressions but, rather, simply identifies the facial muscle actions involved in the production of expressions. Why should that be? It is obviously easier to visually identify six facial expressions than to try to decode individual facial muscle actions. My hypothesis is that by identifying facial muscle actions, the visual system can categorize many more than six facial expressions. Our current model suggests people might be able to recognize over a hundred categories (Benitez-Quiroz et al., 2016a). It is certainly easier to identify a few facial muscle actions than to derive an algorithm that can discern such a large number of categories.

So far, we have identified twenty-three facial expressions of emotion, including compound emotions (e.g., angrily surprised, happily disgusted; Du et al., 2014; Du & Martinez, 2015). We are currently studying an even larger number of facial expressions that correspond to about 400 affect concepts (e.g., anxiety, embarrassment and fatigue; Benitez-Quiroz et al., 2016a). And, although we do not yet know which are universally used and recognized, our preliminary analysis suggests the number of universally recognized expressions is much larger than current models propound.

Everyday experience seems to corroborate our ability to use the face to express many more than just a few emotion categories. People seem to use their faces to communicate a large number of concepts. But which ones?

## Grammatical Markers

Our ability to produce and visually recognize compound facial expressions of emotion allows people to communicate complex concepts non-verbally. Of note, I hypothesize that compound emotions have evolved into grammatical markers. For example, in a recent paper (Benitez-Quiroz et al., 2016b), my research group has shown that compounding the facial expressions of anger, disgust and contempt yields an expression that serves as a marker of negation. If this is part of human language, we call it a grammatical marker. Specifically, a grammatical marker of negation (i.e., negative polarity). This means you can use this expression to convert a positive sentence into a negative one.

We call it the “not face,” Figure 5. We have shown that this compound facial expression of emotion is used in a variety of cultures and languages, including English, Spanish, Mandarin Chinese and American sign language (ASL). Crucially, in ASL the “not face” is, sometimes, the sole marker of negation. That is, if you do not see the face of the signer, you may not know if the signed sentence is positive or negative.

A fundamental and unanswered question in the cognitive sciences is: where does language come from? While most of our human abilities can be traced back to similar or more primitive versions of the same ability in our closest living species, language cannot. The idea that the “not face” is a compound facial expression of emotion is significant because it provides a plausible evolutionary path for the emergence of language through the expression of emotion.

As significant as this result might be, more research is needed to test this hypothesis. Answering the question of the origins of language is one of the most exiting problems in science. But, although the results above show how computational models can aid in this search, additional studies will need to be completed to provide a clear picture of the emergence of grammatical markers through the expression of emotion.

## How many facial expressions?

It is still unclear how many facial expressions are commonly used to communicate affect. Although research in my lab provides strong evidence for the existence of many such categories, other researchers suggest that emotions are not represented categorically in the brain (Skerry & Saxe, 2015) and that this representation is not as localized in a small brain area as our results propound (Wager et al., 2015). Others argue for a hierarchical organization of emotions (Jack et al., 2014). Since facial expressions are dynamic, the hypothesis is that information conveyed earlier is more informative of a few important emotion categories, and later components of the expression are more social specific.

Future research will hopefully resolve the details of the computations performed by our brains to interpret faces and facial expressions. This is important because these results will play a major role in the definition, diagnosis and treatment of psychopathologies. At present, heterogeneity and reification of psychopathologies are major challenges of translational research. It has been argued that a successful definition of the brains’ functional impairments will require a detail understanding of the brain’s computational mechanisms (Insel, 2014; Insel et al., 2010). Computational models are ideally suited to address these problems.

## Acknowledgments

The author is supported in part by the National Institutes of Health, grant R01-DC-014498, and the Human Frontier Science Program, grant RGP0036/2016.

## References

- Aflalo Y, Bronstein A, Kimmel R. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*. 2015; 112(10):2942–2947.
- Benitez-Quiroz, CF., Srinivasan, R., Martinez, AM. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*; Las Vegas, NV, USA. 2016a.
- Benitez-Quiroz CF, Wilbur RB, Martinez AM. The not face: A grammaticalization of facial expressions of emotion. *Cognition*. 2016b; 150:77–84. [PubMed: 26872248]
- Darwin, C. *The expression of the emotions in man and animals*. University of Chicago press; 1965.

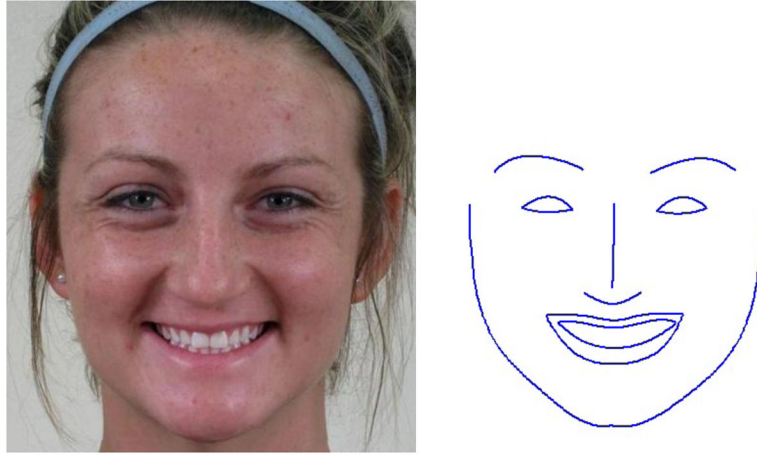
- Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*. 2014; 111(15):E1454–E1462.
- Du S, Martinez AM. Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in clinical neuroscience*. 2015; 17(4):443. [PubMed: 26869845]
- Duchenne, CB. *The Mechanism of Human Facial Expression*. Renard; Paris: Cambridge Univ Press; London: 1862. reprinted (1990)
- Gilad S, Meng M, Sinha P. Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences*. 2009; 106(13):5353–5358.
- Jack RE, Garrod OG, Schyns PG. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current biology*. 2014; 24(2):187–192. [PubMed: 24388852]
- Kemelmacher-Shlizerman, I., Seitz, S., Miller, D., Brossard, E. The megaface benchmark: 1 million faces for recognition at scale. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016.
- Leopold DA, Bondar IV, Giese MA. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*. 2006; 442(7102):572–575. [PubMed: 16862123]
- Loffler G, Yourganov G, Wilkinson F, Wilson HR. fMRI evidence for the neural representation of faces. *Nature neuroscience*. 2005; 8(10):1386–1391. [PubMed: 16136037]
- Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry*. 2014; 171(4):395–397. [PubMed: 24687194]
- Insel T, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*. 2010; 167(7):748–751. [PubMed: 20595427]
- Leibo JZ, Liao Q, Anselmi F, Freiwald WA, Poggio T. View-tolerant face recognition and Hebbian learning imply mirror-symmetric neural tuning to head orientation. *Current Biology*. 2016
- Martinez AM. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern analysis and machine intelligence*. 2002; 24(6):748–763.
- Martinez AM. *Visual Perception of Facial Expressions of Emotion*. *Current Opinions in Psychology*. (in press).
- Martinez A, Du S. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of Machine Learning Research*. 2012; 13:1589–1608. [PubMed: 23950695]
- Meyers EM, Borzello M, Freiwald WA, Tsao D. Intelligent information loss: The coding of facial identity, head pose, and non-face information in the macaque face patch system. *Journal of Neuroscience*. 2015; 35(18):7069–7081. [PubMed: 25948258]
- Jia, H., Martinez, AM. Support vector machines in face recognition with occlusions. *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*; IEEE; 2009. p. 136-141.
- Martinez AM, Kak AC. Pca versus lda. *IEEE Transactions on pattern analysis and machine intelligence*. 2001; 23(2):228–233.
- Neth D, Martinez AM. Emotion perception in emotionless face images suggests a norm-based representation. *Journal of vision*. 2009; 9(1):5–5.
- Neth D, Martinez AM. A computational shape-based model of anger and sadness justifies a configural representation of faces. *Vision research*. 2010; 50(17):1693–1711. [PubMed: 20510267]
- Ohayon S, Freiwald WA, Tsao DY. What makes a cell face selective? The importance of contrast. *Neuron*. 2012; 74(3):567–581. [PubMed: 22578507]
- Richoz AR, Jack RE, Garrod OG, Schyns PG, Caldara R. Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*. 2015; 65:50–64. [PubMed: 25638352]
- Schroff, F., Treibitz, T., Kriegman, D., Belongie, S. Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison. *2011 International Conference on Computer Vision*; IEEE; 2011. p. 2494-2501.
- Sinha P, Balas B, Ostrovsky Y, Russell R. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*. 2006; 94(11):1948–1962.

- Skerry AE, Saxe R. Neural representations of emotion are organized around abstract event features. *Current Biology*. 2015; 25(15):1945–1954. [PubMed: 26212878]
- Sormaz M, Young AW, Andrews TJ. Contributions of feature shapes and surface cues to the recognition of facial expressions. *Vision research*. 2016; 127:1–10. [PubMed: 27425385]
- Srinivasan R, Golomb JD, Martinez AM. A neural basis of facial action recognition in humans. *The Journal of Neuroscience*. 2016; 36(16):4434–4442. [PubMed: 27098688]
- Sterling, P., Laughlin, S. *Principles of neural design*. MIT Press; 2015.
- Turk M. Over twenty years of eigenfaces. *ACM Transactions on Multimedia Computing, Communications, and Applications*. 2013; 9(1s):45.
- Wager TD, Kang J, Johnson TD, Nichols TE, Satpute AB, Barrett LF. A Bayesian model of category-specific emotional brain responses. *PLoS Comput Biol*. 2015; 11(4):e1004066. [PubMed: 25853490]
- You D, Benitez-Quiroz CF, Martinez AM. Multiobjective optimization for model selection in kernel methods in regression. *IEEE transactions on neural networks and learning systems*. 2014; 25(10): 1879–1893. [PubMed: 25291740]
- Zhao, R., Wang, Y., Benitez-Quiroz, CF., Liu, Y., Martinez, AM. *European Conference on Computer Vision*. Springer International Publishing; 2016. Fast and Precise Face Alignment and 3D Shape Reconstruction from a Single 2D Image; p. 590-603.
- Zhao R, Martinez AM. Labeled graph kernel for behavior analysis. *IEEE transactions on pattern analysis and machine intelligence*. 2016; 38(8):1640–1650. [PubMed: 26415154]

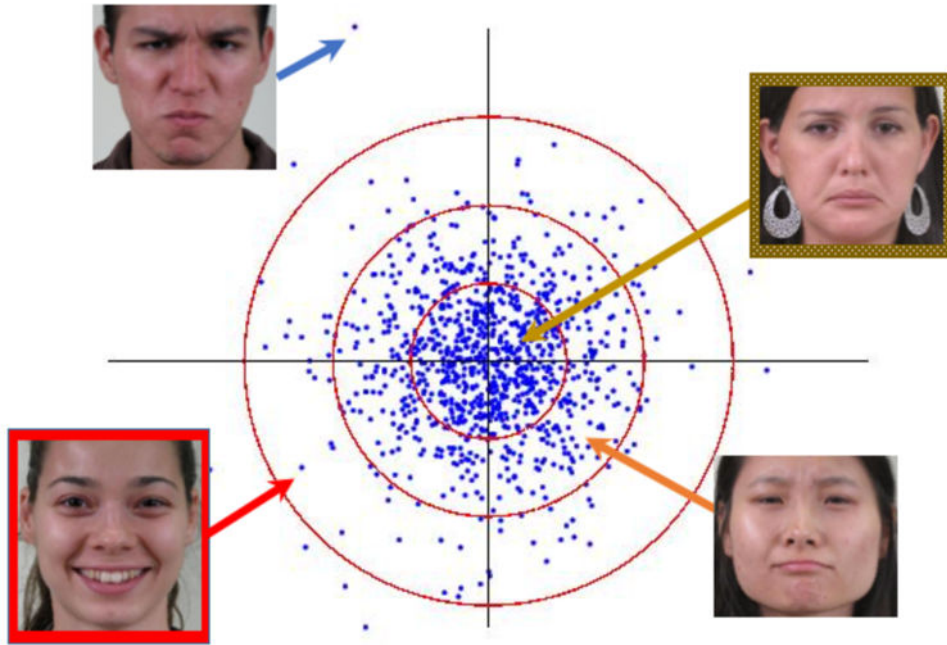
## Recommended Reading Page

1. Du S, Tao Y, Martinez AM. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*. 2014; 111(15):E1454–E1462. The computational model and analysis in this paper identifies a set of previously unknown facial expressions of emotion.
2. Srinivasan R, Golomb JD, Martinez AM. A neural basis of facial action recognition in humans. *The Journal of Neuroscience*. 2016; 36(16):4434–4442. This paper delineates the neural mechanisms that implement the computational model of the perception of facial expressions defined by the author's research group. [PubMed: 27098688]
3. Clark-Polner E, Johnson TD, Barrett LF. Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*. 2016:bhw028. An accessible argument against the categorical model of six (basic) emotion categories.
4. Sinha P, Balas B, Ostrovsky Y, Russell R. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*. 2006; 94(11):1948–1962. A well-articulated, comprehensive overview of image transformations that do not affect our perception of faces but greatly impact computer vision systems.
5. Phillips PJ, O'toole AJ. Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*. 2014; 32(1):74–85. An overview of recent face recognition competitions in computer vision, with a comparison to human performance.

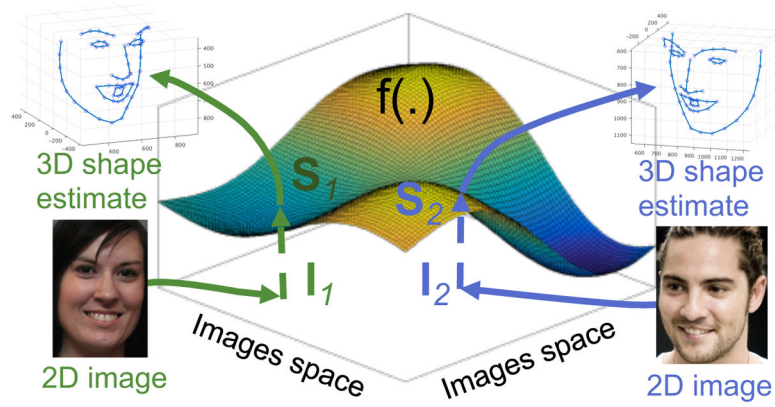




**Figure 1.**  
An image of a face (left) and its corresponding shape (right).



**Figure 2.** Norm-based shape space. Shown here are the two dimensions with largest variance in the shape space. Also shown are the images corresponding to four of these feature vectors. The farther away a face is from the origin of this space, the easier it is to recognize it. In the above, the red demarked face (bottom left) is easier to recognize than the one delineated with a brown-white pattern (top right).



**Figure 3.**

This plot illustrates the idea of a regressor. Here, two of the axes define the image space. The third (vertical) axis defines the 3D-shape. Of course, in reality the image space and the 3D-shape space are defined by more dimensions. The function  $f(\cdot)$  is the regressor. It defines a non-linear manifold. This manifold specifies the mapping between an image  $I_j$  and its 3D-shape  $S_j$ , i.e.,  $S_j = f(I_j)$ .



**Figure 4.** The *American Gothic* illusion. *Left image:* The male character in Wood’s famous *American Gothic* painting is typically described as having a sad expression. However, this man is not expressing any emotion; look closely, this person’s face is at rest – this is called a *neutral* expression. Research suggests that this is because Wood painted this person with an elongated face (i.e., a really thin face) and an exaggerated long distance between brows and mouth. *Right image:* If we morph the original face to be wider and have a significantly shorter brow to mouth distance, then we get a perception of anger. These results are consistent with the predictions of our computational model.



**Figure 5.**

The “not face.” This expression is used as a marker of negation in at least four different languages. That is, in some instances, when we create a negative expression (e.g., no, I didn’t go to the party), we also produce this expression. In American Sign Language (ASL) this expression may be the sole marker of negation. This is called a grammatical marker. Left to right: native speaker of Mandarin Chinese, Spanish, and ASL.