

A Real Time System for Model-based Interpretation of the Dynamics of Facial Expressions

Christoph Mayer, Matthias Wimmer, Freek Stulp, Zahid Riaz, Anton Roth, Martin Eggers, Bernd Radig
Technische Universität München
Boltzmannstr. 3, 85748 Garching
{mayerc,wimmerm,stulp,riaz,roth,egggers,radig}@in.tum.de

1. Motivation

Recent progress in the field of Computer Vision allows intuitive interaction via speech, gesture or facial expressions between humans and technical systems. Model-based techniques facilitate accurately interpreting images with faces by exploiting a priori knowledge, such as shape and texture information. This renders them an inevitable component to realize the paradigm of intuitive human-machine interaction.

Our demonstration shows model-based recognition of facial expressions in real-time via the state-of-the-art Candide-3 face model [1] as visible in Figure 1. This three-dimensional and deformable model is highly appropriate for real-world face interpretation applications. However, its complexity challenges the task of model fitting and we tackle this challenge with an algorithm that has been automatically learned from a large set of images. This solution provides both, high accuracy and runtime. Note, that our system is not limited to facial expression estimation. Gaze direction, gender and age are also estimated.

2. Face Model Fitting

Models reduce the large amount of image data to a small number of model parameters to describe the image content, which facilitates and accelerates the subsequent interpretation task. Cootes et al. [3] introduced modelling shapes with Active Contours. Further enhancements emerged the idea of expanding shape models with texture information [2]. Recent research considers modelling faces in 3D space [1, 10].

Fitting the face model is the computational challenge of finding the parameters that best describe the face within a given image. This task is often addressed by minimizing an objective function, such as the pixel error between the model's rendered surface and the underlying image content. This section describes the four main components of model-based techniques, see [9].

The face model contains a parameter vector \mathbf{p} that represents its configurations. We integrate the complex and deformable 3D wire frame Candide-3 face model [1]. The model consists of 116 anatomical landmarks and its parameter vector $\mathbf{p} = (r_x, r_y, r_z, s, t_x, t_y, \sigma, \alpha)^T$ describes the affine transformation $(r_x, r_y, r_z, s, t_x, t_y)$ and the deformation (σ, α) . The 79 deformation parameters indicate the shape of facial components such as the mouth, the eyes, or the eye brows, etc., see Figure 2.

The localization algorithm computes an initial estimate of the model parameters that is further refined by the subsequent fitting algorithm. Our system integrates the approach of [8], which detects the model's affine transformation in case the image shows a frontal view face.

The objective function yields a comparable value that specifies how accurately a parameterized model matches an image. Traditional approaches manually specify the objective function in a laborious and erroneous task. In contrast, we automatically learn the objective function from a large set of training data based on objective information theoretic measures [9]. This approach does not require expert knowledge and it is domain-independently applicable. As a result, this approach yields more robust and accurate objective functions, which greatly facilitate the task of the associated fitting algorithms. Accurately estimated model parameters in turn are required to infer correct high-level information, such as facial expression or gaze direction.



Figure 1. Interpreting expressions with the Candide-3 face model.

The fitting algorithm searches for the model parameters that best describe the visible face, which corresponds to the minimum of the objective function. Fitting algorithms have been the subject of intensive research and evaluation. We refer to [6] for a recent overview and categorization. Since we adapt the objective function rather than the fitting algorithm, we are able to use any standard fitting algorithm. Because facial expression recognition requires real-time capabilities, we chose a quick hill climbing algorithm. Note, that the reasonable specification of the objective function makes this local optimization method nearly as accurate as global optimization strategies, such as Genetic Algorithms.

3. Facial Expression Recognition

The correctly fitted face model is then successfully exploited to compute a large number of features that describe the facial expression. The model parameters represent structural information by providing information about face properties, such as the opening degree of the mouth or the raising of the eye brows. As described by Ekman [5] such properties indicate facial expressions and thus we infer the facial expression currently visible from them.

In addition, we take the motion of certain landmarks within the visible face into account. The relative location of these landmarks is connected to the structure of the face model. Note that we do not specify these locations manually, because this assumes a good experience of the designer in analyzing facial expressions. Furthermore, in contrast to the structural information these motions are estimated from several subsequent images and therefore considers temporal aspects of facial expressions.

This large amount of feature data provides a profound basis for the classification step, which therefore achieves great accuracy. A classifier infers the facial expression visible in the camera image. We learn a Binary Decision Tree, which is a robust and quick classifier. However, any other multi-class classifier that is able to derive the magnitude of

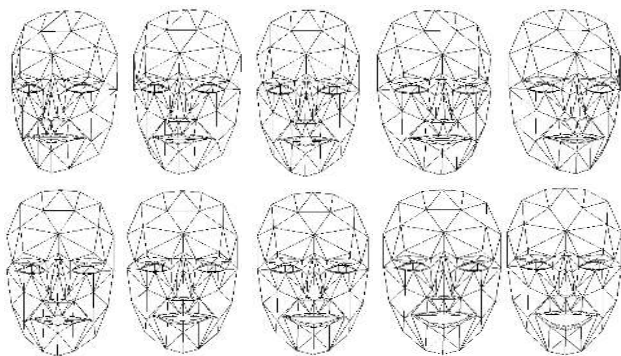


Figure 2. The Candide-3 face model is able to reflect various face shapes and facial expressions.

a facial expression from real valued features may be integrated as well, such as Relevance Vector Machines.

Our demonstrator recognizes the major human facial expressions using the Candide-3 model.

4. Discussion and Conclusion

Our system runs at 10 fps on a 2.0 GHz processor and an image resolution of 640×480 pixels. High quality objective functions that are learned from annotated example images ensure both an accurate and fast computation of the model parameters. Our demonstrator for facial expression estimation has been presented at several events with political audience and on TV. However, the approach of robust face models fitting, forms the basis of various more applications such as gaze detection or gender estimation.

The drawback of our approach is that the data base from which the objective function is learned needs to cover all aspects of face properties. If, for instance, the database did not contain images of bearded men the objective function will fail when confronted with such an image. Furthermore, the data base has to be manually annotated. Although no expert knowledge is required, this task requires a considerable amount of time. An online fitting demonstration is available¹.

References

- [1] J. Ahlberg. Candide-3 – an updated parameterized face. Technical Report LiTH-ISY-R-2326, Link 2001.
- [2] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *ECCV 1998*, Springer-Verlag.
- [3] T. F. Cootes and C. J. Taylor. Active shape models – smart snakes. *BMVC 1992*, pages 266 – 275. Springer Verlag.
- [4] P. Ekman. Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, 1971
- [5] P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for The Measurement of Facial Movement*. Consulting Psychologists Press, San Francisco, 1978.
- [6] R. Hanek. *Fitting Parametric Curve Models to Images Using Local Self-adapting Separation Criteria*. PhD thesis, Technische Universität München, 2004.
- [7] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *International Conference on Automatic Face and Gesture Recognition*, 2000.
- [8] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 2004.
- [9] M. Wimmer, F. Stulp, S. Pietzsch, and B. Radig. Learning local objective functions for robust face model fitting. *IEEE (PAMI)*, 30(8), 2008.
- [10] V. Blanz, K. Scherbaum, and H. Seidel. Fitting a morphable model to 3d scans of faces. *ICCV*, 2007.

¹<http://www9.cs.tum.edu/people/wimmer/MMTuploadImage.php>