# Analysis and Synthesis of Facial Expressions

**Peter Eisert**

Computer Vision & Graphics Group
Image Processing Department
Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute
Einsteinufer 37
D-10587 Berlin, Germany
phone: +49 30 31002 614
fax: +49 30 3927200
email: eisert@hhi.fhg.de
URL: http://bs.hhi.de/~eisert

# Analysis and Synthesis of Facial Expressions

**Peter Eisert**

*In this chapter, the state-of-the-art in facial animation and expression analysis is reviewed and new techniques for the estimation of 3-D human motion, deformation, and facial expressions from monocular video sequences are presented. Since illumination has a considerable influence on the appearance of objects in a scene, methods for the derivation of photometric scene properties from images are also addressed. For a particular implementation, the potential of these analysis techniques is illustrated for applications like character animation and model-based video coding. Experiments have shown that the usage of 3-D computer models allows video transmissions at bit-rates of a few kbit/s enabling a wide variety of new applications.*

**Keywords:** facial expression analysis, facial animation, lighting estimation, model-based video coding

## INTRODUCTION

Facial expression analysis and synthesis techniques have received increasing interest in recent years. Numerous new applications in the areas of low bit-rate communication, user-friendly computer interfaces, film industry, or medicine become available with today's computers. In this chapter, the state-of-the-art in facial animation and analysis is reviewed and new techniques for the estimation of 3-D human motion, deformation, and facial expressions from monocular video sequences are presented. The chapter starts with an overview over existing methods for representing human heads and facial expressions three-dimensionally in a computer. Algorithms for the determination of facial expressions from images and image sequences are reviewed focusing on feature-based and optical-flow based methods. For natural video capture conditions, scene lighting often varies over time. This illumination variability has a considerable influence not only on the visual appearance of the objects in the scene, but also on the performance of the estimation algorithms. Therefore, methods for determining lighting changes in the scene are discussed for the purpose of robust facial analysis under uncontrolled illumination settings. After this overview, an example of a hierarchical, gradient-based method for the robust estimation of MPEG-4 facial animation parameters is given illustrating the potential of model-based coding. This method is able to simultaneously determine both global and local motion in the face in a linear low-complexity framework. In order to improve the robustness against lighting changes in the scene, a new technique for the estimation of photometric properties based on *Eigen light maps* is added to the system. The performance of the presented methods is evaluated in some experiments given in the application section. First, the concept of model-based coding is described, where head-and-shoulder image sequences are represented by computer graphics models that are animated according to the facial motion and deformation extracted from real video sequences. Experiments validate that such sequences can be encoded at less than 1 kbit/s enabling a wide range of new applications. Given an object-based representation of the current scene, changes can easily be made by modifying the 3-D object models. In that context, we will show how facial expression analysis can be used to synthesize new video sequences of arbitrary people, who act exactly in the same way as the person in a reference sequence, which, e.g., enables applications in facial animation for film productions.

## REVIEW OF FACIAL ANALYSIS AND SYNTHESIS TECHNIQUES

### Facial Animation

Modeling the human face is a challenging task because of its familiarity. Already early in life, we are confronted with faces and learn how to interpret them. We are able to recognize individuals from a large number of similar faces and to detect very subtle changes in facial expressions. Therefore, the general acceptability of synthetic face images strongly depends on the 3-D head model used for rendering. As a result, significant effort has been spent on the accurate modeling of a person's appearance and his/her facial expressions (Parke et al., 96). Both problems are addressed in the following two sections.

## 3-D Head Models

In principle, most head models used for animation are based on triangle meshes (Rydfalk, 78, Parke, 82). Texture mapping is applied to obtain a photorealistic appearance of the person (Waters, 87, Terzopoulos et al., 93, Choi et al., 94, Aizawa et al., 95, Lee et al., 95). With extensive use of today's computer graphics techniques highly realistic head models can be realized (Pighin et al., 98).

Modeling the shape of a human head with polygonal meshes results in a representation consisting of a large number of triangles and vertices which have to be moved and deformed to show facial expressions. The face of a person, however, has a smooth surface and facial expressions result in smooth movements of surface points due to the anatomical properties of tissue and muscles. These restrictions on curvature and motion can be exploited by splines which satisfy certain continuity constraints. As a result, the surface can be represented by a set of spline control points that is much smaller than the original set of vertices in a triangle mesh. This has been exploited by Hoch et al. (1994) where B-splines with about 200 control points are used to model the shape of human heads. In (Ip et al., 96), non-uniform rational B-splines (NURBS) represent the facial surfaces. Both types of splines are defined on a rectangular topology and therefore do not allow a local patch refinement in areas that are highly curved. To overcome this restriction, hierarchical splines have been proposed for the head modeling (Forsey et al., 88) to allow a recursive subdivision of the rectangular patches in more complex areas.

Face, eyes, teeth, and the interior of the mouth can be modeled similarly with textured polygonal meshes, but a realistic representation of hair is still not available. A lot of work has been done in this field to model the fuzzy shape and reflection properties of the hair. For example, single hair strands have been modeled with polygonal meshes (Watanabe et al., 92) and the hair dynamics have been incorporated to model moving hair (Anjyo et al., 92). However, these algorithms are computationally expensive and are not feasible for real-time applications in the near future. Image-based rendering techniques (Gortler et al., 96, Levoy et al., 96) might provide new opportunities for solving this problem.

### Facial Expression Modeling

Once a 3-D head model is available, new views can be generated by rotating and translating the 3-D object. However, for the synthesis of facial expressions, the model can no longer be static. In general, two different classes of facial expression modeling can be distinguished in model-based coding applications: the clip-and-paste method and algorithms based on the deformation the 3-D surfaces.

For the *clip-and-paste method* (Aizawa et al., 89, Welsh et al., 90, Chao et al., 94), templates of facial features like eyes and the mouth are extracted from previous frames and mapped onto the 3-D shape model. The model is not deformed according to the facial expression but remains rigid and is used only to compensate for the global motion given by head rotation and translation. All local variations in the face must therefore be described by texture changes of the model. During encoding of a video sequence, a codebook containing templates for different facial expressions is built. A new expression can then be synthesized by combining several feature templates that are specified by their position on the model and their template index from the codebook. As a result, a discrete set of facial expressions can be synthesized. However, the transmission of the template codebook to the decoder consumes a large number of bits which makes the scheme unsuitable for coding purposes (Welsh et al., 90). Beyond that, the localization of the facial features in the frames is a difficult problem. Pasting of templates extracted at slightly inaccurate positions leads to an unpleasant jitter in the resulting synthetic sequence.

The *deformation method* avoids these problems by using the same 3-D model for all facial expressions. The texture remains basically constant and facial expressions are generated by deforming the 3-D surface (Noh et al., 01). In order to avoid the transmission of all vertex positions in the triangle mesh, the facial expressions are compactly represented using high-level expression parameters. Deformation rules associated with the 3-D head model describe how certain areas in the face are deformed if a parameter value changes. The superposition of many of these local deformations is then expected to lead to the desired facial expression. Due to the advantages of the deformation method over the clip-and-paste method (Welsh et al., 90), it is used in most current approaches for representing facial expressions. The algorithms proposed in this chapter are also based on this technique and therefore the following review of related work focuses on the deformation method for facial expression modeling.

One of the first systems of facial expression parameterization was proposed by Hjortsjö (1970) and later extended by the psychologists Ekman and Friesen (1978). Their *facial action coding system* (FACS) is widely used today for the description of facial expressions in combination with 3-D head models (Aizawa et al., 89, Li, 93, Choi et al., 94, Hoch et al., 94). According to that scheme, any facial expression results from the combined action of the 268 muscles in the face. Ekman and Friesen discovered that the human face performs only 46 possible basic actions. Each of these basic actions is affected by a set of muscles that cannot be controlled independently. To obtain the deformation of the facial skin that is caused by a change of an action unit, the motion of the muscles and their influence on the facial tissue can be simulated using soft tissue models (Terzopoulos et al., 93, Lee et al., 95). Due to the high computational complexity of muscle-based tissue simulation, many applications model the surface deformation directly (Aizawa et al., 89, Choi et al., 94) using heuristic transforms between action units and surface motion.

Very similar to the FACS is the parameterization in the *synthetic and natural hybrid coding* (SNHC) part of the MPEG-4 video coding standard (MPEG, 99). Rather than specifying groups of muscles that can be controlled independently and that sometimes lead to deformations in larger areas of the face, the single parameters in this system directly correspond to locally limited deformations of the facial surface. There are 66 different facial animation parameters (FAPs) that control both global and local motion.

Instead of using facial expression descriptions that are designed with a relation to particular muscles or facial areas, data-driven approaches are also used for the modeling. By linearly interpolating 3-D models in a database of people showing different facial expressions, new expressions can be created (Vetter et al., 98, Blanz et al., 99). Ortho-normalizing this *face-space* using a KLT leads to a compact description that allows the representation of facial expressions with a small set of parameters (Hölzer, 99, Kalberer et al., 01).

## Facial Expression Analysis

Synthesizing realistic head-and-shoulder sequences is only possible if the facial animation parameters are appropriately controlled. An accurate estimation of these parameters is therefore essential. In the following sections, different methods are reviewed for the estimation of 3-D motion and deformation from monoscopic image sequences. Two different groups of algorithms are distinguished: feature-based approaches which track distinct features in the images and optical flow based methods that exploit the entire image for estimation.

Feature-Based Estimation

One common way for determining the motion and deformation in the face between two frames of a video sequence is the use of feature points (Kaneko et al., 91, Terzopoulos et al., 93, Gee et al., 94, Huang et al., 94, Lopez et al., 95, Pei, 98). Highly discriminant areas with large spatial variations such as areas containing the eyes, nostrils, or mouth corners are identified and tracked from frame to frame. If corresponding features are found in two frames, the change in position determines the displacement.

How the features are searched depends on properties such as color, size, and shape. For facial features, extensive research has been performed especially in the area of face recognition (Chellappa et al., 95). Templates (Brunelli et al., 93), often used for finding facial features, are small reference images of typical features. They are compared at all positions in the frame to find a good match between the template and the current image content (Thomas et al., 87). The best match is said to be the corresponding feature in the second frame. Problems with templates arise from the wide variability of captured images due to illumination changes or different viewing positions. To compensate for these effects, eigen-features (Moghaddam et al., 97, Donato et al., 99) which span a space of possible feature variations or deformable templates (Yuille, 91) which reduce the features to parameterized contours can be utilized.

Instead of estimating single feature points, the whole contour of features can also be tracked (Huang et al., 91, Pearson, 95) using *snakes*. Snakes (Kas et al., 87) are parameterized active contour models that are composed of internal and external energy terms. Internal energy terms account for the shape of the feature and smoothness of the contour while the external energy attracts the snake towards feature contours in the image.

All feature-based algorithms have in common that single features like the eyes can be found quite robustly. Dependent on the image content, however, only a small number of feature correspondences can typically be determined. As a result, the estimation of 3-D motion and deformation parameters from the displacements lacks the desired accuracy if a feature is erroneously associated with a different feature in the second frame.

### Optical Flow Based Estimation

Approaches based on optical flow information utilize the entire image information for the parameter estimation, leading to a large number of point correspondences. The individual correspondences are not as reliable as the ones obtained with feature-based methods, but due to the large number of equations, some mismatches are not critical. In addition, possible outliers (Black et al., 96) can generously be removed without obtaining an underdetermined system of equations for the determination of 3-D motion.

One way of estimating 3-D motion is the explicit computation of an optical flow field (Horn et al., 81, Barron et al., 94, Dufaux et al., 95) which is followed by the derivation of motion parameters from the resulting dense displacement field (Netravali et al., 84, Essa et al., 94, Bartlett et al., 95). Since the computation of the flow field from the optical flow constraint equation (Horn et al., 81), which relates image gradient information (Simoncelli, 94) to 2-D image displacements, is an underdetermined problem, additional smoothness constraints have to be added (Horn, 86, Barron et al., 94). A non-linear cost function (Barron et al., 94) is obtained that is numerically minimized. The use of hierarchical frameworks (Enkelmann, 88, Singh, 90, Sezan et al., 93) can reduce the computational complexity of the optimization in this high-dimensional parameter space. However, even if the global minimum is found, the heuristical smoothness constraints may lead to deviations from the correct flow field, especially at object boundaries and depth discontinuities.

In model-based motion estimation, the heuristical smoothness constraints are therefore often replaced by explicit motion constraints derived from the 3-D object models. For rigid body motion estimation (Kappei, 88, Koch, 93), the 3-D motion model, specified by three rotational and three translational degrees of freedom, restricts the possible flow fields in the image plane. Under the assumption of perspective projection, known object shape, and small motion between two successive video frames, an explicit displacement field can be derived that is linear in the six unknown degrees of freedom (Longuet, 84, Netravali et al., 84, Waxman et al., 87). This displacement field can easily be combined with the optical flow constraint to obtain a robust estimator for the 6 motion parameters. Iterative estimation in an analysis-synthesis framework (Li et al., 93) removes remaining errors caused by the linearization of image intensity and the motion model.

For facial expression analysis, the rigid body assumption can no longer be maintained. Surface deformations due to facial expressions have to be considered additionally. Most approaches found in the literature (Ostermann, 94, Choi et al., 94, Black et al., 95, Pei, 98, Li et al., 98) separate this problem into two steps. First, global head motion is estimated under the assumption of rigid body motion. Local motion caused by facial expressions is regarded as noise (Li et al., 94b) and therefore the textured areas around the mouth and the eyes are often excluded from the estimation (Black et al., 95, Li et al., 94b). Given head position and orientation, the remaining residuals of the motion-compensated frame are used to estimate local deformations and facial expressions. In (Black et al., 95,Black et al., 97), several 2-D motion models with 6 (affine) or 8 parameters are used to model local facial deformations. By combining these models with the optical flow constraint, the unknown parameters are estimated in a similar way as in the rigid body case. High-level facial animation parameters can finally be derived from the estimated set of 2-D motion parameters. Even higher robustness can be expected by directly estimating the facial animation parameters using more sophisticated motion models. In (Choi et al., 94), a system is described that utilizes an explicit 3-D head model. This head model directly relates changes of facial animation parameters to surface deformations. Orthographic projection of the motion constraints and combination with optical flow information result in a linear estimator for the unknown parameters. The accuracy problem of separate global and local motion estimation is here relaxed by an iterative framework that alternately estimates the parameters for global and local motion.

The joint estimation of global head motion together with facial expressions is rarely addressed in the literature. In (Li et al., 93, Li et al., 94), a system for the combined estimation of global and local motion is presented that stimulated the approaches presented in the next section. A 3-D head model based on the Candide (Rydfalk, 78) model is used for image synthesis and provides explicit 3-D motion and deformation constraints. The affine motion model describes the image displacements as a linear function of the 6 global motion parameters and the facial action units from the FACS system which are simultaneously estimated in an analysis-synthesis framework. Another approach that allows a joint motion and deformation estimation has been proposed by DeCarlo et al. (1996, 1998). A deformable head model is employed that consists of 10 separate face components that are connected by spring-like forces incorporating anthropometric constraints (DeCarlo et al., 98b, Farkas, 95). Thus, the head shape can be adjusted similar to the estimation of local deformations. For the determination of motion and deformation, again a 3-D motion model is combined with the optical flow constraint. The 3-D model also includes a dynamic, Lagrangian description for the parameter changes similar to the work in (Essa et al., 94, Essa et al., 97). Since the head model lacks any color information, no synthetic frames can be rendered which makes it impossible to use an analysis-synthesis loop. Therefore, additional edge forces are added to avoid an error accumulation in the estimation.

### Illumination Analysis

In order to estimate the motion of objects between two images, most algorithms make use of the *brightness constancy assumption* (Horn, 86). This assumption, which is an inherent part of all optical flow-based and many template-based methods, implies that corresponding object points in two frames show the same brightness. However, if the lighting in the scene changes, the brightness of corresponding points might differ significantly. But also if the orientation of the object surface relative to a light source changes due to object motion, brightness is in general not constant (Verri et al., 89). On the contrary, intensity changes due to varying illumination conditions can dominate the effects caused by object motion (Pentland, 91, Horn, 86, Tarr, 98). For accurate and robust extraction of motion information, lighting effects must be taken into account.

In spite of the relevance of illumination effects, they are rarely addressed in the area of 3-D motion estimation. In order to allow the use of the optical flow constraint for varying brightness, higher order differentials (Treves et al., 94) or pre-filtering of the images (Moloney, 91) have been applied. Similarly, *lightness algorithms* (Land et al., 71, Ono et al., 93, Blohm, 97) make use of the different spectral distributions of texture and intensity changes due to shading, in order to separate irradiance from reflectance. If the influence of illumination cannot be suppressed sufficiently by filtering as, e.g., in image regions depicting highlights caused by specular reflections, the corresponding parts are often detected (Klinker et al., 90, Stauder, 94, Schluens et al. 95) and classified as outliers for the estimation.

Rather than removing the disturbing effects, explicit information about the illumination changes can be estimated. This not only improves the motion estimation but also allows the manipulation and visual enhancement of the illumination situation in an image afterwards (Blohm, 97). Under controlled conditions with, e.g., known object shape, light source position (Sato et al., 97, Sato et al., 96, Baribeau et al., 92), and homogeneous non-colored surface properties (Ikeuchi et al., 91, Tominaga et al., 00), parameters of sophisticated reflection models like the Torrance-Sparrow model (Torrance et al., 67, Nayar et al., 91, Schlick, 94) which also includes specular reflection can be estimated from camera views. Since the difficulty of parameter estimation increases significantly with model complexity, the analysis of global illumination scenarios

(Heckbert, 92) with, e.g., inter-reflections (Forsyth et al., 91) is only addressed for very restricted applications (Wada et al., 95).

In the context of motion estimation, where the exact position and shape of an object are often not available, mostly simpler models are used that account for the dominant lighting effects in the scene. The simplest scenario is the assumption of pure ambient illumination (Foley et al., 90). In (Gennert et al., 87, Moloney et al., 91, Negahdaripour et al., 93), the optical flow constraint is extended by a two parameter function to allow for global intensity scaling and global intensity shifts between the two frames. Local shading effects can be modeled using additional directional light sources (Foley et al., 90). For the estimation of the illuminant direction, surface normal information is required. If this information is not available as, e.g., for the large class of *shape from shading* algorithms (Horn et al., 89, Lee et al., 89), assumptions about the surface normal distribution are exploited to derive the direction of the incident light (Pentland, 82, Lee et al., 89, Zheng et al., 91, Bozdagi et al., 94).

If explicit 3-D models and with that surface normal information are available, more accurate estimates of the illumination parameters are obtainable (Stauder, 95, Deshpande et al., 96, Brunelli, 97, Eisert et al., 97). In these approaches, Lambertian reflection is assumed in combination with directional and ambient light. Given the surface normals, the illumination parameters are estimated using neural networks (Brunelli, 97), linear (Deshpande et al., 96, Eisert et al., 97), or non-linear (Stauder, 95) optimization.

Rather than using explicit light source and reflection models to describe illumination effects, also multiple images captured from the same viewing position but under varying illumination can be exploited. Hallinan et al. showed (Hallinan, 94, Epstein et al., 95) that five eigen images computed from a set of differently illuminated facial images are sufficient to approximate arbitrary lighting conditions by linearly blending between the eigen images. An analytic method for the derivation of the eigen components can be found in (Ramamoorthi, 02). This low-dimensional space of face appearances can be represented as an illumination cone as shown by Belhumeur et al. (1998). In (Ramamoorthi et al., 01), the reflection of light was theoretically described by convolution in a signal-processing framework. Illumination analysis or inverse rendering can then be considered as deconvolution. Beside the creation of arbitrarily illuminated face images, the use of multiple input images also allows the estimation of facial shape and thus a change of head pose in 2-D images (Georgiades et al., 99). Using eigen light maps of explicit 3-D models (Eisert et al., 02) instead of blending between eigen images, extends the applicability of the approach also to locally deforming objects like human faces in image sequences.

For the special application of 3-D model-based motion estimation, relatively few approaches have been proposed that incorporate photometric effects. In (Bozdagi et al., 94), the illuminant direction is estimated according to (Zheng et al., 91), first without exploiting the 3-D model. Given the illumination parameters, the optical flow constraint is extended to explicitly consider intensity changes caused by object motion. For that purpose, surface normals are required which are derived from the 3-D head model. The approach proposed in (Stauder, 95, Stauder, 98) makes explicit use of normal information for both illumination estimation and compensation. Rather than determining the illuminant direction from a single frame, the changes of surface shading between two successive frames are exploited to estimate the parameters. The intensity of both ambient and directional light, as well as the direction of the incident light are determined by minimizing a non-linear cost function. Experiments performed for both approaches show that the consideration of photometric effects can significantly improve the accuracy of estimated motion parameters and the reconstruction quality of the motion-compensated frames (Bozdagi et al., 94, Stauder, 95).

## HIERARCHICAL MODEL-BASED FACIAL EXPRESSION ANALYSIS

The most challenging part of facial expression analysis is the estimation of 3-D facial motion and deformation from two-dimensional images. Due to the loss of one dimension caused by the projection of the real world onto the image plane, this task can only be solved by exploiting additional knowledge of the objects in the scene. In particular, the way the objects move can often be restricted to a low number of degrees of freedom that can be described by a limited set of parameters. In this section, an example of a new 3-D model-based method for the estimation of facial expressions is presented that makes use of an explicit parameterized 3-D human head model describing shape, color, and motion constraints of an individual person (Eisert, 00). This model information is jointly exploited with spatial and temporal intensity gradients of the images. Thus, the entire area of the image showing the object of interest is used instead of dealing with discrete feature points, resulting in a robust and highly accurate system. A linear and computationally efficient algorithm is derived for different scenarios. The scheme is embedded in a hierarchical analysis-synthesis framework to avoid error accumulation in the long-term estimation.

### Optical-flow based Analysis

In contrast to feature-based methods, gradient-based algorithms utilize the optical flow constraint equation

$$\frac{\partial I(X,Y)}{\partial X} d_x + \frac{\partial I(X,Y)}{\partial Y} d_y = I(X,Y) - I'(X,Y), \qquad (1)$$

where $\frac{\partial I}{\partial X}$ and $\frac{\partial I}{\partial Y}$ are the spatial derivatives of the image intensity at pixel position [X Y]. I'-I denotes the temporal change of the intensity between two time instants Δt=t'-t corresponding to two successive frames in an image sequence. This equation, obtained by Taylor series expansion up to first order of the image intensity, can be set up anywhere in the image. It relates the unknown 2-D motion displacement $\mathbf{d}=[d_x \; d_y]$ with the spatial and temporal derivatives of the images.

The solution of this problem is under-determined since each equation has two new unknowns for the displacement coordinates. For the determination of the optical flow or motion field, additional constraints are required. Instead of using heuristical smoothness constraints, explicit knowledge about the shape and motion characteristics of the object is exploited. Any 2-D motion model can be used as an additional motion constraint in order to reduce the number of unknowns to the number of motion parameters of the corresponding model. In that case, it is assumed that the motion model is valid for the complete object. An over-determined system of equations is obtained that can be solved robustly for the unknown motion and deformation parameters in a least-squares sense.

In the case of facial expression analysis, the motion and deformation model can be taken from the shape and the motion characteristics of the head model description. In this context, a triangular B-spline model (Eisert et al., 98a) is used to represent the face of a person. For rendering purposes, the continuous spline surface is discretized and approximated by a triangle mesh as shown in Fig. 6. The surface can be deformed by moving the splines' control points and thus affecting the shape of the underlying mesh. A set of facial animation parameters (FAPs) according to the MPEG-4 standard (MPEG, 99) characterizes the current facial expression and has to be estimated from the image sequence. By concatenating all transformations in the head model deformation and using knowledge from the perspective camera model, a relation between image displacements and FAPs can be analytically derived

$$\mathbf{d} = f\left(FAP_0, FAP_1, \ldots, FAP_{N-1}\right). \tag{2}$$

Combining this motion constraint with the optical flow constraint (1) leads to a linear system of equations for the unknown FAPs. Solving this linear system in a least squares sense, results in a set of facial animation parameters that determines the current facial expression of the person in the image sequence.

### Hierarchical Framework

Since the optical flow constraint equation (1) is derived assuming the image intensity to be linear, it is only valid for small motion displacements between two successive frames. To overcome this limitation, a hierarchical framework can be used (Eisert et al., 98a). First, a rough estimate of the facial motion and deformation parameters is determined from sub-sampled and low-pass filtered images, where the linear intensity assumption is valid over a wider range. The 3-D model is motion compensated and the remaining motion parameter errors are reduced on frames having higher resolutions.

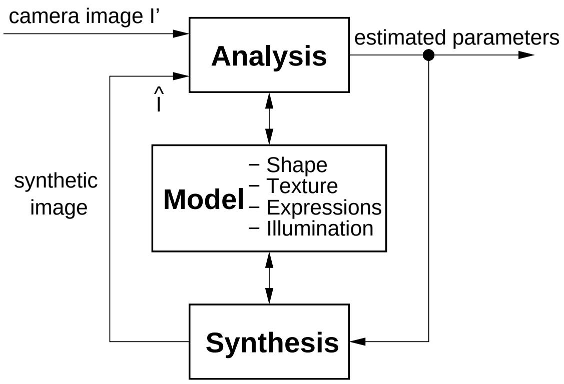

Figure 1: Analysis-synthesis loop of the model-based estimator.

The hierarchical estimation can be embedded into an analysis-synthesis loop as shown in Fig. 1. In the analysis part, the algorithm estimates the parameter changes between the previous synthetic frame Î and the current frame I' from the video sequence. The synthetic frame Î is obtained by rendering the 3-D model (synthesis part) with the previously determined parameters. This approximative solution is used to compensate the differences between the two frames by rendering the deformed 3-D model at the new position. The synthetic frame now approximates the camera frame much better. The remaining linearization errors are reduced by iterating through different levels of resolution. By estimating the parameter changes with a synthetic frame that corresponds to the 3-D model, an error accumulation over time is avoided.

## LINEAR ILLUMINATION ANALYSIS

For natural video capture conditions, scene lighting often varies over time. This illumination variability has a considerable influence not only on the visual appearance of the objects in the scene, but also on the performance of computer vision algorithms or video coding methods. The efficiency and robustness of these algorithms can be significantly improved by removing the undesired effects of changing illumination. In this section, we introduce a 3-D model-based technique for estimating and manipulating the lighting in an image sequence (Eisert et al., 02). The current scene lighting is estimated for each frame exploiting 3-D model information and by synthetic re-lighting of the original video frames. To provide the estimator with surface normal information, the objects in the scene are represented by 3-D shape models and their motion and deformation are tracked over time using a model-based estimation method. Given the normal information, the current lighting is estimated with a linear algorithm of low computational complexity using an orthogonal set of light maps.

### Light Maps

Instead of explicitly modeling light sources and surface reflection properties in the computer graphics scene and calculating shading effects during the rendering process as it is done in (Bozdagi et al., 94, Stauder, 95, Eisert et al., 98b), the shading and shadowing effects are here described by a linear superposition of several light maps which are attached to the object surface. Light maps are, similar to texture maps, two-dimensional images that are wrapped around the object containing shading instead of color information. During rendering, the unshaded texture map $I_{tex}^{C}(\mathbf{u})$ with $C \in \{R, G, B\}$ representing the three color components and the light map L(**u**) are multiplied according to

$$I^{C}(\mathbf{u}) = I_{tex}^{C}(\mathbf{u}) \cdot L(\mathbf{u}) \tag{3}$$

in order to obtain a shaded texture map $I^{C}(\mathbf{u})$. The two-dimensional coordinate **u** specifies the position in both texture map and light map that are assumed to have the same mapping to the surface. For a static scene and viewpoint independent surface reflections, the light map can be computed off-line which allows the use of more sophisticated shading methods as, e.g., radiosity algorithms (Goral et al., 84) without slowing down the final rendering. This approach, however, can only be used if both object and light sources do not move. To overcome this limitation, we use a linear combination of scaled light maps instead of a single one

$$I^{C}(\mathbf{u}) = I_{tex}^{C}(\mathbf{u}) \cdot \sum_{i=0}^{N-1} \alpha_i^C L_i(\mathbf{u}). \tag{4}$$

By varying the scaling parameter $\alpha_i^C$ and thus blending between different light maps L$_i$, different lighting scenarios can be created. Moreover, the light map approach can also model wrinkles and creases which are difficult to describe by 3-D geometry (Pighin et al., 98, Liu et al., 01). The N light maps L$_i$(**u**) can be computed off-line with the same surface normal information **n**(**u**) but with different light source configurations. In our experiments, we use one constant light map L$_0$ representing ambient illumination while the other light maps are calculated assuming Lambert reflection and point light sources located at infinity having illuminant direction **l**$_i$

$$\begin{aligned} L_0(\mathbf{u}) &= 1 \\ L_i(\mathbf{u}) &= \max\{-\mathbf{n}(\mathbf{u}) \cdot \mathbf{l}_i, 0\}, \quad 1 \le i \le N-1 \end{aligned} \tag{5}$$

This configuration can be interpreted as an array of point light sources whose intensities and colors can be individually controlled by the parameters $\alpha_i^C$. Fig. 2 shows an example of such an array with the illuminant direction varying between -60° and 60° in longitudinal and latitudinal direction, respectively.

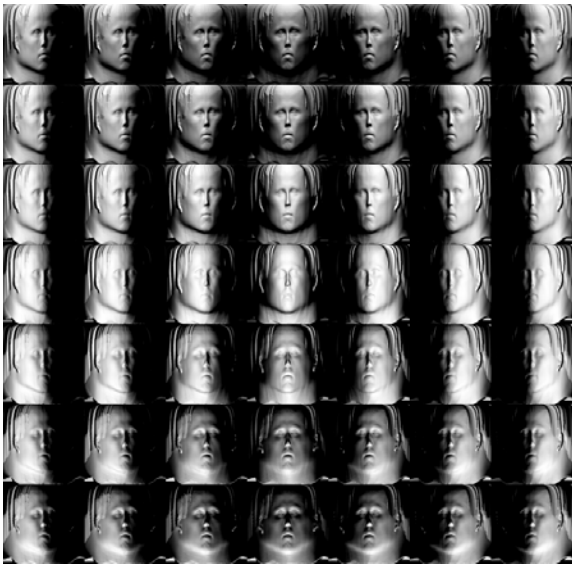

Figure 2: Array of light maps for a configuration with 7 by 7 light sources.

### Eigen Light Maps

In order to reduce the number of unknowns $\alpha_i^C$ that have to be estimated, a smaller orthogonal set of light maps is used rather than the original one. A Karhunen-Loève transformation (KLT) (Turk et al., 91) is applied to the set of light maps $L_i$ with $1 \leq i \leq N-1$ creating *eigen light maps* which concentrate most energy in the first representations. Hence, the number of degrees of freedom can be reduced without significantly increasing the mean squared error when reconstructing the original set. Fig. 3 shows the first four *eigen light maps* computed from a set of 50 different light maps. The mapping between the light maps and the 3-D head model is here defined by cylindrical projection onto the object surface.

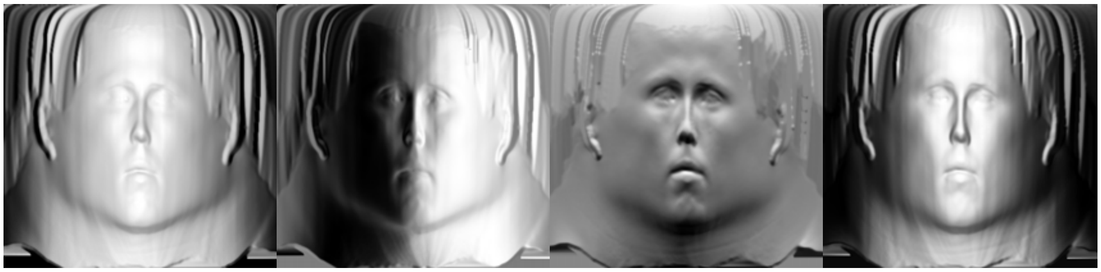

Figure 3: First four *eigen light maps* representing the dominant shading effects.

### Estimation of Lighting Properties

For the lighting analysis of an image sequence, the parameters $\alpha_i^C$ have to be estimated for each frame. This is achieved by tracking motion and deformation of the objects in the scene as described above and rendering a synthetic motion-compensated model frame using the unshaded texture map $I_{tex}^C$. From the pixel intensity differences between the camera frame $I_{shaded}^C(\mathbf{x})$ with **x** being the pixel position and the model frame $I_{unshaded}^C(\mathbf{x})$, the unknown parameters $\alpha_i^C$ are derived. For each pixel **x**, the corresponding texture coordinate **u** is determined and the linear equation

$$I_{shaded}^C(\mathbf{x}) = I_{unshaded}^C(\mathbf{x}) \cdot \sum_{i=0}^{N-1} \alpha_i^C L_i(\mathbf{u}(\mathbf{x})). \tag{6}$$

is set up. Since each pixel **x** being part of the object contributes one equation, a highly over-determined linear system of equations is obtained that is solved for the unknown $\alpha_i^C$'s in a least-squares sense. Rendering the 3-D object model with the shaded texture map using the estimated parameters $\alpha_i^C$ leads to a model frame which approximates the lighting of the original frame. In the same way, the inverse of (6) can be used to remove the lighting variations in real video sequences as it is shown in Fig. 4.

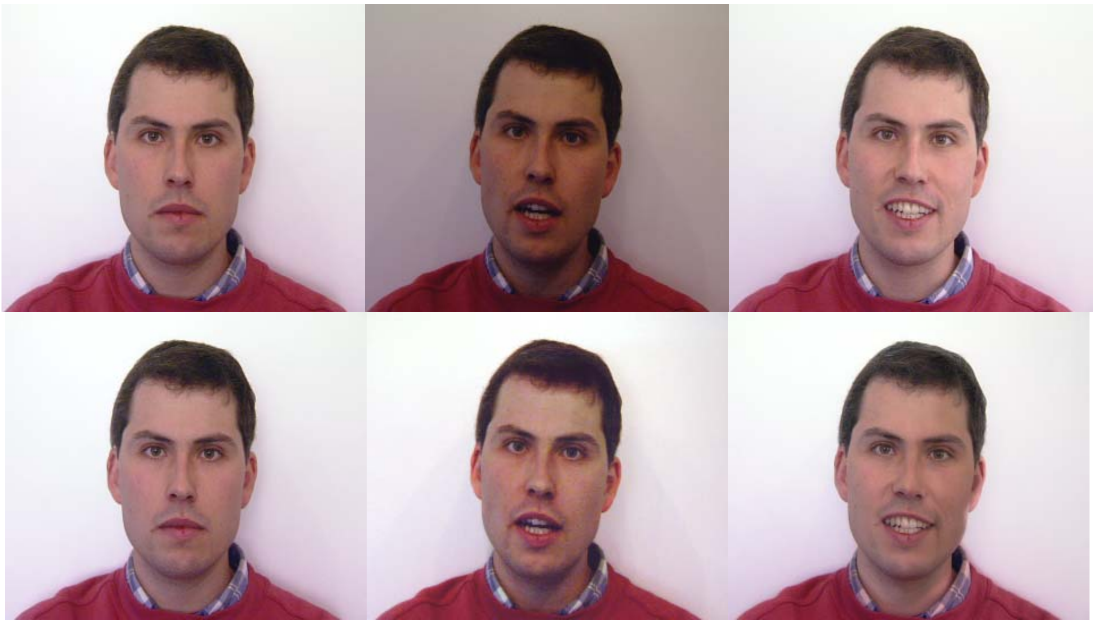

Figure 4: Upper row: original video frames, lower row: corresponding frames of illumination-compensated sequence with constant lighting.

## APPLICATIONS

In this section, two applications, model-based coding and facial animation, are addressed which make use of the aforementioned methods for facial expression analysis and synthesis. Experimental results from the approach in (Eisert, 00) are provided in order to exemplarily illustrate the applicability of model-based techniques to these applications.

### Model-based Coding

In recent years, several video coding standards, such as H.261/3 and MPEG-1/2/4 have been introduced to address the compression of digital video for storage and communication services. These standards describe a hybrid video coding scheme, which consists of block-based motion-compensated prediction (MCP) and DCT-based quantization of the prediction error. The recently determined H.264 standard also follows the same video coding approach. These waveform-based schemes utilize the statistics of the video signal without knowledge of the semantic content of the frames and achieve compression ratios of several hundreds to one at a reasonable quality.

If semantic information about a scene is suitably incorporated, higher coding efficiency can be achieved by employing more sophisticated source models. Model-based video codecs, e.g., use 3-D models for representing the scene content. Fig. 5 shows the structure of a model-based codec for the application of video telephony.

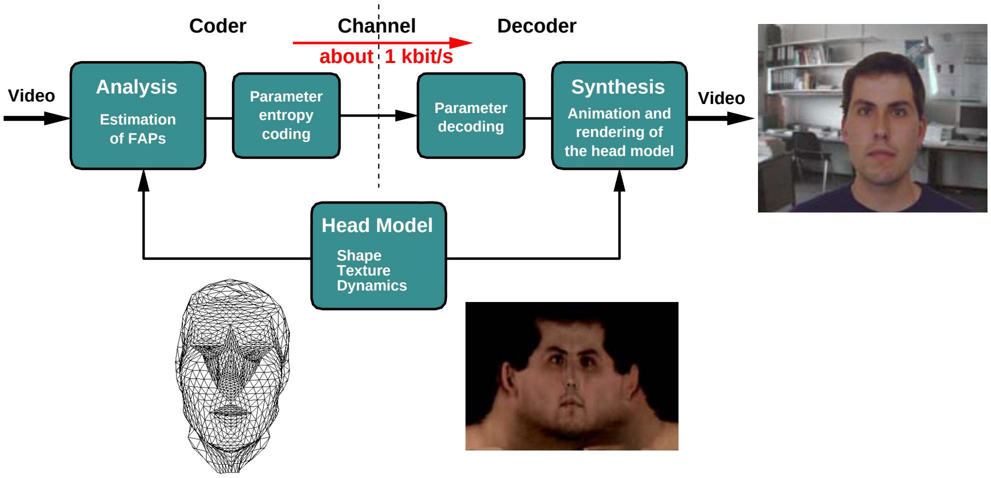

Figure 5: Structure of a model-based codec.

A video camera captures images of the head-and-shoulder part of a person. The encoder analyzes the frames and estimates 3-D motion and facial expressions of the person using a 3-D head model. A set of facial animation parameters (FAPs) is obtained that describes - together with the 3-D model - the current appearance of the person. Only a few parameters have to be encoded and transmitted, resulting in very low bit-rates. The head model has to be transmitted only once if it has not already been stored at the decoder in a previous session. At the decoder, the parameters are used to deform the head model according to the person's facial expressions. The original video frame is finally approximated by rendering the 3-D model at the new position.

The use of model-based coding techniques in communication scenarios leads to extremely low bit-rates of only a few kbit/s for the transmission of head-and-shoulder image sequences. This enables video streaming also over low-bandwidth channels for mobile devices like PDAs or smart phones. The rendering complexity is comparable to that of a hybrid video codec and in experiments, frame rates of 30 Hz have been achieved on an iPAQ PDA. On the other hand, the intensive exploitation of a-priori knowledge restricts the applicability to special scenes that can be described by 3-D models available at the decoder. In a video-phone scenario, e.g., other objects like a hand in front of the face simply do not show up unless explicitly modeled in the virtual scene. In order to come up with a codec that is able to encode arbitrary scenes, hybrid coding techniques can be incorporated increasing bit-rate but assuring generality to unknown objects. The model-aided codec is an example of such an approach (Eisert et al., 00). Model-based coding techniques, however, offer also additional features besides low bit-rates, enabling many new applications that cannot be achieved with traditional hybrid coding methods. In immersive video-conferencing (Kauff et al., 02), multiple participants who are located at different places can be seated at a joint virtual table. Due to the 3-D representation of the objects, pose modification for correct seating positions can easily be accomplished as well as view-point corrections according to the user's motion. By replacing the 3-D model of one person by a different one, other people can be animated with the expressions of an actor as shown in the next section. Similarly, avatars can be driven to create user-friendly man-machine-interfaces, where a human-like character interacts with the user. Analyzing the user with a web cam also gives the computer feedback about the user's emotions and intentions (Picard, 97) Other cues in the face can assist the computer-aided diagnosis and treatment of patients in medical applications. For example, asymmetry in facial expressions caused by facial palsy can be measured three-dimensionally (Frey et al., 99) or craniofacial syndromes can be detected by the 3-D analysis of facial feature positions (Hammond et al., 01) These examples indicate the wide variety of applications for model-based facial analysis and synthesis techniques.

### Model-based View Synthesis

In this section, experimental results of a model-based video coding scheme using facial expression analysis are presented.

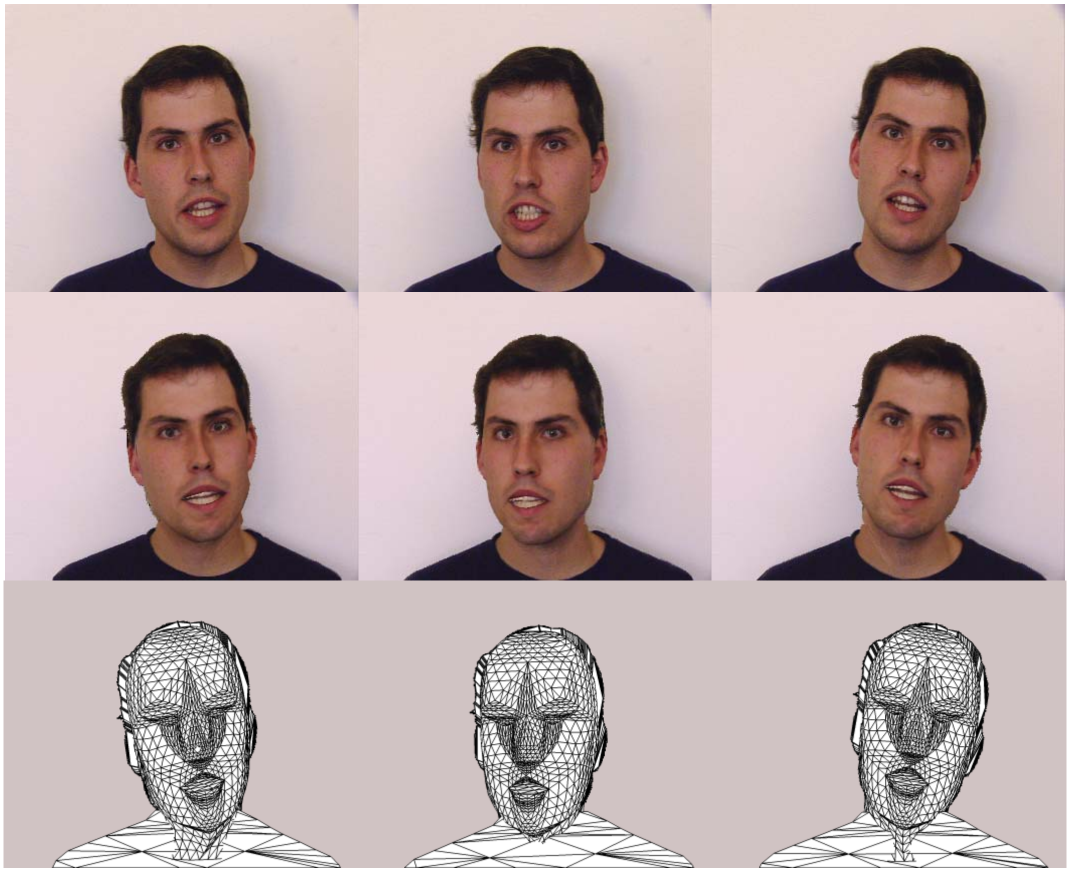

Figure 6: Upper row: Original video sequence. Middle row: Synthesized sequence. Lower row: hidden line representation.

Fig. 6 shows a head-and-shoulder video sequence recorded with a camera in CIF resolution and 25 Hz. A generic head model is roughly adjusted in shape to the person in the sequence and the first frame is projected onto the 3-D model. Non visible areas of the texture map are extrapolated. The model is encoded and transmitted to the decoder and neither changed nor updated during the video sequence. Only facial animation parameters and lighting changes are streamed over the channel. In this experiment, 18 facial animation parameters are estimated, quantized, encoded, and transmitted. The frames in the middle row of Fig. 6 are synthesized from the deformed 3-D model which is illustrated in the lower row of Fig. 6 by means of a wireframe. The bit-rate needed to encode these parameters is below 1 kbit/s at a quality of 34.6 dB PSNR. The PSNR between synthesized and original frames is here measured only in the facial area to exclude effects from the background that is not explicitly modeled. The trade-off between bit-rate, which can be controlled by changing the quantizer values for the FAPs, and reconstruction quality is shown in Fig. 7.

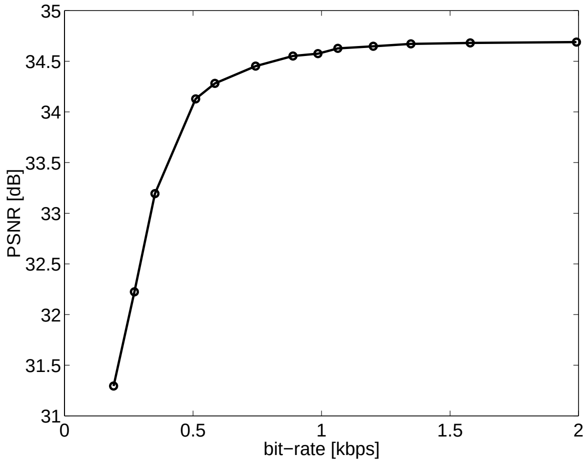

Figure 7: Reconstruction quality in PSNR over bit-rate needed for encoding the animation parameters.

### Facial Animation

The use of different head models for analysis and synthesis of head-and-shoulder sequences is also interesting in the field of character animation in film productions or web applications. The facial play of an actor sitting in front of a camera is analyzed and the resulting FAPs are used to control arbitrary 3-D models. This way, different people, animals, or fictitious creatures can be animated realistically. The exchange of the head model to animate other people is exemplarily shown in Fig. 8. The upper row depicts some frames of the original sequence used for facial expression analysis. Instead of rendering the sequence with the same 3-D head model used for the FAP estimation and thus reconstructing the original sequence, the head model is exchanged for image synthesis leading to new sequences with different people that move according to the original sequence. Examples of this character animation are shown in the lower two rows of Fig. 8. In these experiments, the 3-D head models for *Akiyo* and *Bush* are derived from a single image. A generic head model whose shape is controlled by a set of parameters is roughly adjusted to the outline of the face and the position of eyes and mouth. Then, the image is projected onto the 3-D model and used as a texture map. Since the topology of the mesh is identical for all models, the surface deformation description need not be changed and facial expressions can easily be applied to different people.

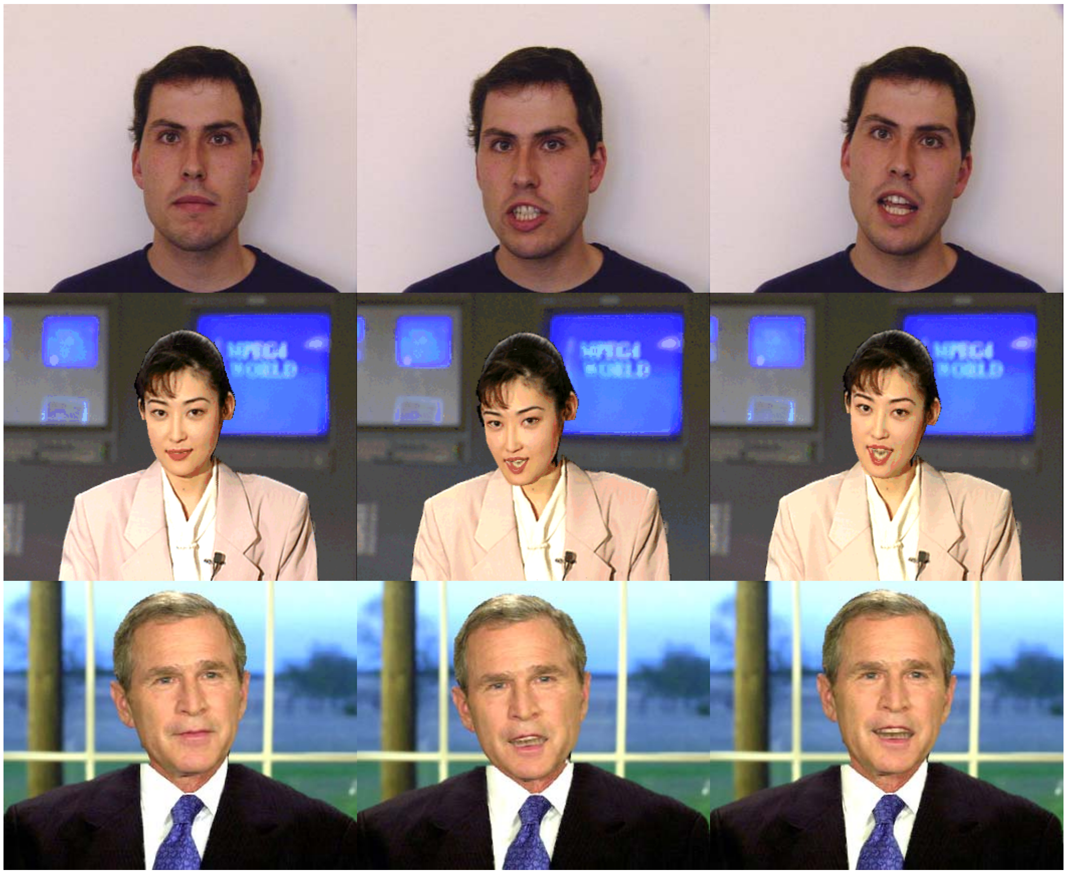

Figure 8: Animation of different people using facial expressions from a reference sequence. Upper row: reference sequence. Middle and lower row: Synthesized new sequences.

Since the same generic model is used for all people, point correspondences between surface points and texture coordinates are inherently established. This enables the morphing between different characters by linearly blending between the texture map and the position of the vertices. In contrast to 2-D approaches (Liu et al., 01), this might be done during a video sequence due to use of a 3-D model. Local deformations caused by facial expressions are not affected by this morphing. Fig. 9 shows an example of a view morphing process between two different people.

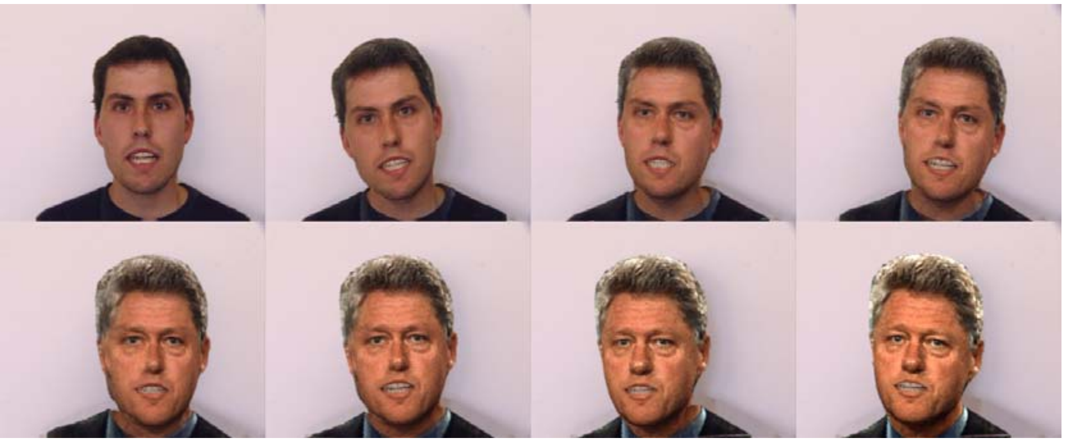

Figure 9: Motion-compensated 3-D morph between two people.

## CONCLUSIONS

Methods for facial expression analysis and synthesis have received increasing interest in recent years. The computational power of current computers and handheld devices like PDAs already allow a real-time rendering of 3-D facial models which is the basis for many new applications in the near future. Especially for handheld devices which are connected to the Internet via a wireless channel, bit-rates for streaming video is limited. Transmitting only facial expression parameters reduces the bandwidth requirements drastically to a few kbit/s. In the same way, face animations or new human-computer interfaces can be realized with low demands on storage capacities. On the high quality end, film productions may get new impacts for animation and realistic facial expression and motion capture without the use of numerous sensors that interfere with the actor. Last but not least, information about motion and symmetry of facial features can be exploited in medical diagnosis and therapy.

All these applications have in common that accurate information about 3-D motion deformation and facial expressions is required. In this chapter, the state-of-the-art in facial expression analysis and synthesis has been reviewed and a new method for determining FAPs from monocular images sequences has been presented. In a hierarchical framework, the parameters are robustly found using optical flow information together with explicit knowledge about shape and motion constraints of the objects. The robustness can further be increased by incorporating photometric properties in the estimation. For this purpose, a computationally efficient algorithm for the determination of lighting effects was given. Finally, experiments have shown that video transmission of head-and-shoulder scenes can be realized at data rates of a few kbit/s even with today's technologies enabling a wide variety of new applications.

## REFERENCES

Aizawa K. & Huang T. S. (1995). Model-based image coding: Advanced video coding techniques for very low bit-rate applications. Proc. IEEE, 83(2):259-271.

Aizawa K., Harashima H. & Saito T. (1989). Model-based analysis synthesis image coding (MBASIC) system for a person's face. Sig. Proc.: Image Comm., 1(2):139-152.

Anjyo K., Usami Y. & Kurihara T. (1992). A simple method for extracting the natural beauty of hair. SIGGRAPH, 26, 111-120.

Black M. J. & Anandan P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding, 63(1):75-104.

Barron J. L., Fleet D. J. & Beauchemin S. S. (1994). Systems and experiment - performance of optical flow techniques. Int. J. of Comp. Vision, 12(1):43-77.

Belhumeur P. N. & Kriegman D. J. (1998). What is the set of images of an object under all possible illumination conditions. Int. J. of Comp. Vision, 28(3):245-260.

Blohm W. (1997). Lightness determination at curved surfaces with apps to dynamic range compression and model-based coding of facial images. IEEE Tr. Image Proc., 6(8):1129-1138.

Brunelli R. & Poggio T. (1993). Face recognition: Features versus templates. IEEE Tr. PAMI, 15(10):1042-1052.

Baribeau R., Rioux M. & Godin G. (1992). Color reflectance modeling using a polychromatic laser range sensor. IEEE Tr. PAMI, 14(2):263-269.

Brunelli R. (1997). Estimation of pose and illuminant direction for face processing. Image-and-Vision-Computing, 15(10):741-748.

Bozdagi G., Tekalp A. M. & Onural L. (1994). 3-D motion estimation and wireframe adaption including photometric effects for model-based coding of facial image sequences. IEEE Tr. CSVT, 4(3):246-256.

Blanz V. & Vetter T. (1999). A morphable model for the synthesis of 3D faces. SIGGRAPH, 187-194.

Bartlett M., Viola P. A., Sejnowski T. J., Golomb B. A., Larsen J., Hager J. C. & Ekman P. (1995). Classifying facial action. Advances in Neural Inf. Proc. Systems 8, MIT Press, 823-829.

Black M. J. & Yacoob Y. (1995). Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. Proc. ICCV, 374-381.

Black M. J., Yacoob Y. & Ju S. X. (1997). Recognizing human motion using parameterized models of optical flow. In Motion-Based Recognition, 245-269. Kluwer Academic Publishers.

Choi C., Aizawa K., Harashima H. & Takebe T. (1994). Analysis and synthesis of facial image sequences in model-based image coding. IEEE Tr. CSVT, 4(3):257-275.

Chao S. & Robinson J. (1994). Model-based analysis/synthesis image coding with eye and mouth patch codebooks. Proc. of Vision Interface, 104-109.

Chellappa R., Wilson C. L. & Sirohey S. (1995). Human and machine recognition of faces: A survey. Proc. IEEE, 83(5):705-740.

Donato G., Bartlett M. S., Hager J. C., Ekman P. & Sejnowski T. (1999). Classifying facial actions. IEEE Tr. PAMI, 21(10):974-989.

Deshpande S. G. & Chaudhuri S. (1996). Recursive estimation of illuminant motion from flow field. Proc. ICIP, 3, 771-774.

Dufaux F. & Moscheni F. (1995). Motion estimation techniques for digital TV: A review and a new contribution. Proc. IEEE, 83(6):858-876.

DeCarlo D. & Metaxas D. (1996). The integration of optical flow and deformable models with applications to human face shape and motion estimation. Proc. CVPR, 231-238.

DeCarlo D. & Metaxas D. (1998). Deformable model-based shape and motion analysis from images using motion residual error. Proc. ICCV, 113-119.

DeCarlo D., Metaxas D. & Stone M. (1998). An anthropometric face model using variational techniques. SIGGRAPH, 67-74.

Ekman P. & Friesen W. V. (1978). Facial Action Coding System. Consulting Psychologists Press, Inc., Palo Alto.

Eisert P. & Girod B. (1997). Model-based 3D motion estimation with illumination compensation. In Proc. Int. Conf. on Image Proc. and its Applications, 1, 194-198.

Eisert P. & Girod B. (1998). Analyzing facial expressions for virtual conferencing. IEEE Computer Graphics and Applications, 18(5):70-78.

Eisert P. & Girod B. (1998). Model-based coding of facial image sequences at varying illumination conditions. Proc. 10th IMDSP. Workshop 98, 119-122.

Eisert P. & Girod B. (2002). Model-based enhancement of lighting conditions in image sequences. Proc. SPIE VCIP, VCIP-02.

R. Epstein, Hallinan P. & Yuille A. (1995). 5 plus or minus 2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In Proc. IEEE Work. on Physics-based Modeling in Comp. Vision.

Eisert P. (2000). Very Low Bit-Rate Video Coding Using 3-D Models. PhD thesis, University of Erlangen, Shaker Verlag, Aachen, Germany.

Enkelmann W. (1988). Investigations of multigrid algorithms for estimation of optical flow fields in image sequences. Comp. Vision, Graphics and Image Proc., 43(2):150-177.

Essa I. A. & Pentland A. P. (1994). A vision system for observing and extracting facial action parameters. Proc. CVPR, 76-83.

Essa I. A. & Pentland A. P. (1997). Coding, analysis, interpretation, and recognition of facial expressions. IEEE Tr. PAMI, 19(7):757-763.

Eisert P., Wiegand T. & Girod B. (2000). Model-aided coding: A new approach to incorporate facial animation into motion-compensated video coding. IEEE Tr. CSVT, 10(3):344-358.

Farkas L. G. (1995). Anthropometry of the Head and Face. Raven Press.

Forsey D. R. & Bartels R. H. (1988). Hierarchical B-spline refinement. SIGGRAPH, 22, 205-212.

Frey M., Giovanoli P., Gerber H., Slameczka M. & Stüssi E. (1999). Three-dimensional video analysis of facial movements: A new method to assess the quantity and quality of the smile. Plastic and Reconstructive Surgery, 104(7):2032-2039.

Foley J. D., van Dam A., Feiner S. K. & Hughes J. F. (1990). Computer Graphics, Principles and Practice. Addison-Wesley.

Forsyth D. & Zisserman A. (1991). Reflections on shading. IEEE Tr. PAMI, 13(7):671-679.

Georghiades A. S., Belhumeur P. N. & Kriegman D. J. (1999). Illumination-based image synthesis: Creating novel images of human faces under different pose and lighting. In Proc. IEEE Work. on Multi-View Modeling and Analysis of Visual Scenes.

Gee A. & Cipolla R. (1994). Determining the gaze of faces in images. Image and Vision Computing, 12(10):639-647.

Gortler S. J., Grzeszczuk R., Szeliski R. & Cohen M. F. (1996). The Lumigraph. SIGGRAPH, 43-54.

Gennert M. A. and Negahdaripour S. (1987). Relaxing the brightness constancy assumption in computing optical flow. Technical report, M.I.T. AI Lab Memo No. 975.

Goral C. M., Torrance K. E., Greenberg D. P. & Battaile B. (1984). Modeling the interaction of light between diffuse surfaces. SIGGRAPH, 18, 213-222.

Hallinan P. W. (1994). A low-dimensional representation of human faces for arbitrary lighting conditions. Proc. CVPR.

Horn B. K. P. & Brooks M. J. (1989). Shape from Shading. MIT Press, Cambridge, MA.

Heckbert P. S. (1992). Introduction to global illumination. Global Illumination Course, SIGGRAPH.

Hoch M., Fleischmann G. & Girod B. (1994). Modeling and animation of facial expressions based on B-splines. Visual Computer, 11:87-95.

Hammond, P. Hutton T. J., Patton M. A. & Allanson J. E. (2001). Delineation and visualisation of congenital abnormality using 3D facial images. Intell. Data Anal. in Medicine and Pharm.

Hjortsjö C.-H. (1970). Man's face and mimic language. Student literature, Lund, Sweden.

Huang T. S. & Netravali A. N. (1994). Motion and structure from feature correspondences: A review. Proc. IEEE, 82(2):252-268.

Hölzer A. (1999). Optimierung eines dreidimensionalen Modells menschlicher Gesichtsausdrücke für die Codierung von Videosequenzen. Diploma thesis, University of Erlangen-Nuremberg.

Horn B. K. P. (1986). Robot Vision. MIT Press, Cambridge.

Huang T. S., Reddy S. & Aizawa K. (1991). Human facial motion analysis and synthesis for video compression. Proc. SPIE VCIP, 234-241.

Horn B. K. P. & Schunck B. G. (1981). Determining optical flow. Artificial Intelligence, 17(1-3):185-203.

Ip H. H. S. & Chan C. S. (1996). Script-based facial gesture and speech animation using NURBS based face model. Computer and Graphics, 20(6):881-891.

Ikeuchi K. & Sato K. (1991). Determining reflectance properties of an object using range and brightness images. IEEE Tr. PAMI, 13(11):1139-1153.

Kappei F. (1988). Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge. PhD thesis, University Hannover.

Kalberer G. A. & Van Gool L. (2001). Lip animation based on observed 3D speech dynamics. Proc. SPIE VCIP, 16-25.

Kaneko M., Koike A. & Hatori Y. (1991). Coding of facial image sequence based on a 3-D model of the head and motion detection. J. of Visual Comm. and Image Rep., 2(1):39-54.

Koch R. (1993). Dynamic 3-D scene analysis through synthesis feedback control. IEEE Tr. PAMI, 15(6):556-568.

Kauff P. & Schreer O. (2002). Virtual team user environments - a step from tele-cubicles towards distributed tele-collaboration in mediated workspaces. Proc. ICME.

Klinker G. J., Shafer S. A. & Kanade T. (1990). A physical approach to color image understanding. Int. J. of Computer Vision, 4:7-38.

Kass M., Witkin A. & Terzopoulos D. (1987). Snakes: Active contour models. Int. J. of Computer Vision, 1(4):321-331.

Li Y. & Chen Y. (1998). A hybrid model-based image coding system for very low bit-rate coding. IEEE J. on Selected Areas in Communications, 16(1):28-41.

Li H. & Forchheimer R. (1994). Two-view facial movement estimation. IEEE Tr. CSVT, 4(3):276-287.

Longuet-Higgins H. C. (1984). The visual ambiguity of a moving plane. Proc. of the Royal Society of London, B 223:165-175.

Lopez R. & Huang T. S. (1995). 3D head pose computation from 2D images: Template versus features. Proc. ICIP, 599-602.

Levoy M. & Hanrahan P. (1996). Light field rendering. SIGGRAPH, 31-42.

Li H. (1993). Low Bitrate Image Sequence Coding. PhD thesis, Linköping University,. Linköping Studies in Science and Technology, No. 318.

Li H., Lundmark A. & Forchheimer R. (1994). Image sequence coding at very low bitrates: A review. IEEE Tr. Image Proc., 3(5):589-609.

Land E. H. & McCann J.J. (1971). Lightness and retinex theory. J. of the Opt. Soc. Am., 61:1-11.

Lee C. H. & Rosenfeld A. (1989). Shape from Shading, chapter Improved Methods of Estimating Shape from Shading using the Light Source Coordinate System, 323-347. MIT-Press, Cambridge.

Li H., Roivainen P. & Forchheimer R. (1993). 3-D motion estimation in model-based facial image coding. IEEE Tr. PAMI, 15(6):545-555.

Liu Z., Shan Y. & Zhang Z. (2001). Expressive expression mapping with ratio images. SIGGRAPH.

Lee Y., Terzopoulos D. & Waters K. (1995). Realistic modeling for facial animation. SIGGRAPH, 55-61.

Moloney C. R. & Dubois E. (1991). Estimation of motion fields from image sequences with illumination variation. Proc. ICASSP, 4, 2425-2428.

Moloney C. R. (1991). Methods for illumination-independent processing of digital images. In IEEE Pacific Rim Conf. on Comm., Computers and Sig. Proc., 2, 811-814.

Moghaddam B. & Pentland A. (1997). Probabilistic visual learning for object representation. IEEE Tr. PAMI, 19(7):696-710.

ISO/IEC FDIS 14496-2 (1999), Generic Coding of audio-visual objects: (MPEG-4 video), Final Draft International Standard, Document N2502.

Nayar S. K., Ikeuchi K. & Kanade T. (1991). Surface reflection: Physical and geometrical perspectives. IEEE Tr. PAMI, 13(7):611-634.

Noh J.-Y. & Neumann U. (2001). Expression cloning. SIGGRAPH.

Netravali A. N. & Salz J. (1985). Algorithms for estimation of three-dimensional motion. AT&T Technical J., 64(2):335-346.

Negahdaripour S. & Yu C. H. (1993). A generalized brightness change model for computing optical flow. Proc. ICCV, 2-11.

Ono E., Morishima S. & Harashima H. (1993). A model based shade estimation and reproduction schemes for rotational face. Proc. PCS, page 2.2.

Ostermann J. (1994). Object-based analysis-synthesis coding (OBASC) based on the source model of moving flexible 3D-objects. IEEE Tr. Image Proc., 705-711.

Parke F. I. (1982). Parameterized models for facial animation. IEEE Computer Graphics and Applications, 2(9):61-68.

Pearson D. E. (1995). Developments in model-based video coding. Proc. IEEE, 83(6):892-906.

Pentland A. (1982). Finding the illuminant direction. J. of the Opt. Soc. Am., 72(4):170-187.

Pentland A. (1991). Photometric motion. IEEE Tr. PAMI, 13(9):879-890.

Pighin F., Hecker J., Lischinski D., Szeliski R. & H. Salesin D. (1998). Synthesizing realistic facial expressions from photographs. SIGGRAPH, 75-84.

Picard R. W. (1997). Affective Computing. MIT Press, Cambridge, USA.

Pei S., Ko C. & Su M. (1998). Global motion estimation in model-based image coding by tracking three-dimensional contour feature points. IEEE Tr. CSVT, 8(2):181-190.

Parke F. I. & Waters K. (1996). Computer Facial Animation. A K Peters, Massachusetts.

Ramamoorthi R. (2002). Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object. IEEE Tr. PAMI.

Ramamoorthi R. & Hanrahan P. (2001). A signal-processing framework for inverse rendering. SIGGRAPH.

Rydfalk M. (1978). CANDIDE: A Parameterized Face. PhD thesis, Linköping University, LiTH-ISY-I-0866.

Schlick C. (1994). A survey of shading and reflectance models. Computer Graphics Forum, 13(2):121-132.

Sato Y. & Ikeuchi K. (1996). Reflectance analysis for 3D computer graphics model generation. Graphical Models and Image Processing, 58(5):437-451.

Simoncelli E. P. (1994). Design of multi-dimensional derivative filters. Proc. ICIP, 790-794.

Singh A. (1990). An estimation theoretic framework for image-flow computation. Proc. ICCV, 168-177.

Sezan M. I. & Lagendijk R. L. (1993). Motion Analysis and Image Sequence Processing, chapter Hierarchical Model-Based Motion Estimation, 1-22. Kluwer Academic Publishers.

Schluens K. & Teschner M. (1995). Analysis of 2D color spaces for highlight elimination in 3D shape reconstruction. In Proc. ACCV, 2, 801-805.

Stauder J. (1994). Detection of highlights for object-based analysis-synthesis coding. Proc. PCS, 300-303.

Stauder J. (1995). Estimation of point light source parameters for object-based coding. Sig. Proc.: Image Comm., 7(4-6):355-379.

Stauder J. (1998). Illumination analysis for synthetic/natural hybrid image sequence generation. In Comp. Graphics Intern. (CGI 98), 506-511.

Sato Y., Wheeler M. D. & Ikeuchi K. (1997). Object shape and reflectance modeling from observation. SIGGRAPH, 379-387.

Tarr M. J. (1998). Why the visual recognition system might encode the effects of illumination. Vision Research, 38:2259-2275.

Thomas G. A. & Hons B. A. (1987). Television motion measurement for DATV and other applications. BBC Research Department Report, 1-20.

Treves P. & Konrad J. (1994). Motion estimation and compensation under varying illumination. Proc. ICIP, I, 373-377.

Turk M. & Pentland A. (1991). Eigenfaces for recognition. J. for Cogn. Neurosc., 3(1):71-86.

Torrance K. E. & Sparrow E. M. (1967). Theory of off-specular reflection from roughened surfaces. J. of the Opt. Soc. Am., 1105-1114.

Tominaga S. & Tanaka N. (2000). Estimating reflection parameters from a single color image. IEEE Computer Graphics and Applications, 20(5):58-66.

Terzopoulos D. & Waters K. (1993). Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Tr. PAMI, 15(6):569-579.

Vetter T. & Blanz V. (1998). Estimating coloured 3D face models from single images: An example based approach. Proc. ECCV, 2, 499-513.

Verri A. & Poggio T. (1989). Motion field and optical flow: Qualitative properties. IEEE Tr. PAMI, 11(5):490-498.

Waters K. (1987). A muscle model for animating three-dimensional facial expressions. SIGGRAPH, 21, 17-24.

Waxman A. M., Kamgar-Parsi B. & Subbarao M. (1987). Closed-form solutions to image flow equations for 3D structure and motion. Int. J. of Comp. Vision, 1:239-258.

Watanabe Y. & Suenaga Y. (1992). A trigonal prism-based method for hair image generation. IEEE Computer Graphics and Applications, 12(1):47-53.

Welsh W. J., Searsby S. & Waite J. B. (1990). Model-based image coding. British Telecom Technology J., 8(3):94-106.

Wada T., Ukida H. & Matsuyama T. (1995). Shape from shading with interreflections under proximal light source. Proc. ICCV, 66-71.

Yuille A. L. (1991). Deformable templates for face recognition. J. of Cognitive Neuroscience, 3(1):59-70.

Zheng Q. & Chellappa R. (1991). Estimation of illuminant direction, albedo, and shape from shading. IEEE Tr. PAMI, 13(7):680-702.