Model-based Enhancement of Lighting Conditions in Image Sequences

Peter Eisert and Bernd Girod

Information Systems Laboratory
Stanford University
{eisert,bgirod}@stanford.edu
http://www.stanford.edu/~eisert

ABSTRACT

Illumination variability has a considerable influence 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 paper, we introduce a 3-D model-based technique for estimating and manipulating the lighting in an image sequence. 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. This results in a small set of parameters which efficiently represent the scene lighting. In our experiments, we demonstrate how this representation can be used to create video sequences with arbitrary new illumination. When encoding a video sequence for transmission over a network, significant coding gains can be achieved when removing the time varying lighting effects prior to encoding. Improvements of up to 3 dB in PSNR are observed for an illumination-compensated sequence.

Keywords: lighting enhancement, model-based illumination compensation, image sequence manipulation

1. INTRODUCTION

The performance of computer vision algorithms is often significantly influenced by the wide variability of an object's appearance caused by changes in the scene lighting. In face and object recognition, for example, a new face has to be matched with a database of facial images recorded under different lighting conditions which complicates the comparison severely [1–3]. Similarly, 3-D motion estimation algorithms and 3-D geometry reconstruction techniques rely on the establishment of image point correspondences which are difficult to determine if the illumination in the scene changes. Since the lighting in the scene cannot be explicitly controlled in many applications, the algorithms have to cope with this large range of different image data.

Rather than finding features in the image that are insensitive to variations in the lighting, some approaches explicitly model the lighting effects. In [4–6], photometric properties of the scene are estimated to increase the robustness of 3-D motion estimation from image sequences. In order to keep the computational complexity for the estimation low, these methods rely on simple local lighting scenarios to account for the dominant lighting effects in the scene. Typically, ambient illumination plus one additional point light source located at infinity are used.

In this paper, we present a new model-based illumination estimation technique that represents shading effects by a linear combination of orthogonal light maps. This representation allows to deal with multiple light sources and can describe to some extend also global illumination effects like inter-reflections or self-shadowing. The estimation, however, is still linear in the unknowns requiring only low computational complexity. The underlying concept is similar to the approaches presented in [7–9] where a linear superposition of still images showing a scene under different lighting conditions is used to create a wide range of new shading effects. We have extended this approach to the case of image sequences with moving objects by using a superposition of light maps attached to the surface of 3-D object models. These models are moved and deformed according to the objects in the scene using the model-based motion estimation technique presented in [10]. Since the light maps are connected to the object surface, they

are warped similarly. Given the light map representation, a set of light map weighting parameters is estimated for each frame, compactly describing the current lighting scenario for each object.

Once the time varying parameter set is determined, modifications to the lighting in the image sequence can be made by altering the estimated parameters. This is investigated for the special case of head-and-shoulder sequences typically found in video-telephony or video-conferencing applications. For that scenario, we often have to deal with poorly illuminated scenes and low data-rates of the connected network which requires efficient compression of the image sequences. Both problems can be relaxed if suitable information about the lighting can be determined. Sequences recorded under poor illumination conditions can be enhanced by properly adjusting the illumination in the frames. Experiments show that the visual quality of encoded images can be significantly improved by estimating and compensating the lighting variations in the video sequence prior to the encoding and transmission over a network.

2. ESTIMATION AND COMPENSATION OF SCENE LIGHTING

For the robust estimation of illumination parameters from head-and-shoulder image sequences, a model-based analysis-by-synthesis technique is employed. A 3-D head model which can be moved and deformed to represent facial expressions is used to provide shape and texture information of the person in the video. This model is illuminated by synthetic light sources and model frames are generated by rendering the virtual scene using computer graphics techniques. By varying the set of parameters defining position and shape of the object as well as the properties of the light sources in the scene, the appearance of the rendered model frame can be changed. Given a particular camera frame, the model-based estimator determines the optimal parameter setting which results in a minimum squared frame difference between original and model frame. The motion and deformation estimation part is described in [10]. In this paper, the estimation of illumination parameters is addressed.

Assuming that the 3-D model is already compensated for motion and deformation, the remaining differences between the original camera frame and the corresponding rendered model frame are caused by lighting differences between the frames plus some additional noise. A parameter set representing the photometric properties in the scene is estimated in order to minimize the frame difference leading to a model frame that optimally matches the original camera frame within the bounds of the scene model accuracies. In order to allow the use of linear lighting models, the non-linear γ-predistortion applied in the camera [11] is inverted before estimating the photometric properties. Hence, all image and texture intensities I used throughout the paper represent linear intensity values. After the estimation and compensation of the lighting, the images are again γ-predistorted.

Instead of explicitly modeling light sources and surface reflection properties and calculating shading effects during the rendering process as it is done in [4–6], the shading and shadowing effects in this work are 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 $I^C_{tex}(\mathbf{u})$ with $C \in \{R, G, B\}$ representing the three color components and the light map $L(\mathbf{u})$ are multiplied according to

$$I^C(\mathbf{u}) = I^C_{tex}(\mathbf{u}) \cdot L(\mathbf{u}) \tag{1}$$

in order to obtain a shaded texture map $I^C(\mathbf{u})$. The two-dimensional coordinate $\mathbf{u}$ 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 [12] 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

$$I^C(\mathbf{u}) = I^C_{tex}(\mathbf{u}) \cdot \sum_{i=0}^{N-1} \alpha^C_i L_i(\mathbf{u}). \tag{2}$$

By varying the scaling parameter α^C_i and thus blending between different light maps L_i, different lighting scenarios can be created. The N light maps $L_i(\mathbf{u})$ are again computed off-line with the same surface normal information $\mathbf{n}(\mathbf{u})$ 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 $\mathbf{l}_i$

$$\begin{aligned} L_0(\mathbf{u}) &= 1 \\ L_i(\mathbf{u}) &= \max\{-\mathbf{n}(\mathbf{u}) \cdot \mathbf{l}_i, 0\}, \quad 1 \leq i \leq N-1. \end{aligned} \tag{3}$$

This configuration can be interpreted as an array of point light sources whose intensities and colors can be individually controlled by the parameters α_i^C. Fig. 1 shows an example of such an array with the illuminant direction varying between -60^o and 60^o in longitudinal and latitudinal direction, respectively.

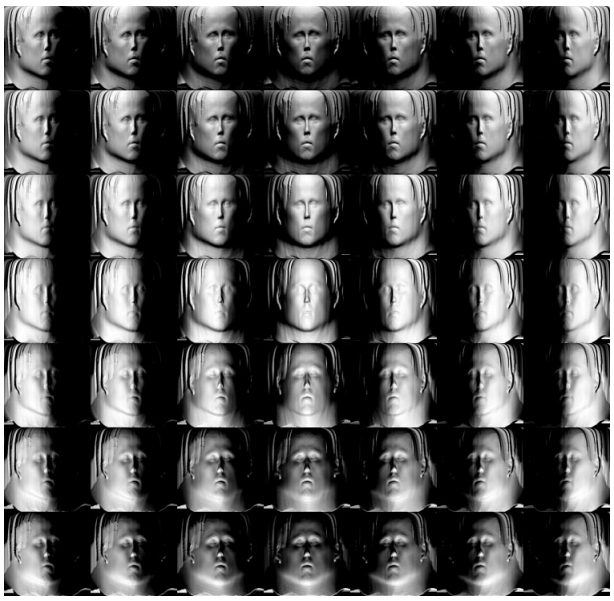

Figure 1. Array of lightmaps for a configuration with 7 by 7 light sources.

In order to reduce the number of unknowns α_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) [13] 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. 2 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.

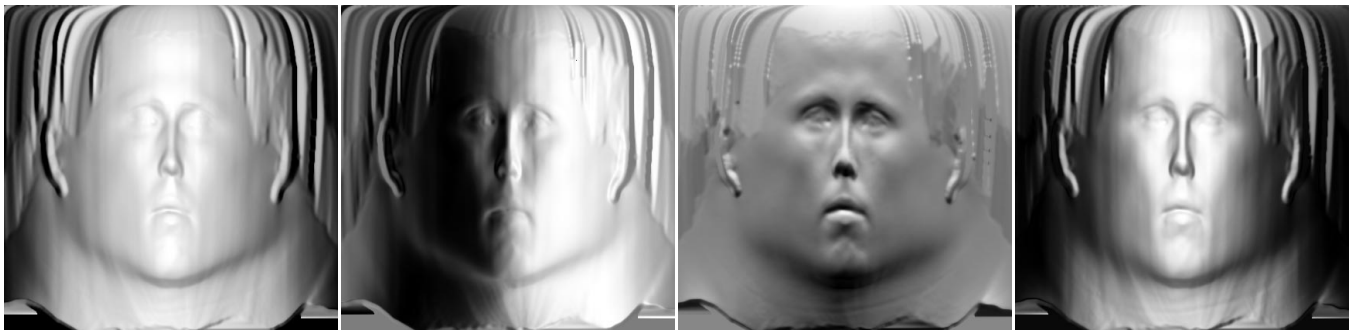

Figure 2. First four *eigen light maps* representing the dominant shading effects.

For the lighting analysis of an image sequence, the parameters α_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 in [10] and rendering a

synthetic motion-compensated model frame using the unshaded texture map I^C_{tex}. From the pixel intensity differences between the camera frame $I^C_{shaded}(\mathbf{x})$ with $\mathbf{x}$ being the pixel position and the model frame $I^C_{unshaded}(\mathbf{x})$, the unknown parameters α^C_i are derived. For each pixel $\mathbf{x}$, the corresponding texture coordinate $\mathbf{u}$ is determined and the linear equation

$$I^C_{shaded}(\mathbf{x}) = I^C_{unshaded}(\mathbf{x}) \cdot \sum_{i=0}^{N-1} \alpha^C_i L_i(\mathbf{u}(\mathbf{x})) \tag{4}$$

is set up. Since each pixel $\mathbf{x}$ being part of the object contributes one equation, a highly over-determined linear system of equations is obtained that is solved for the unknown α^C_i's by least-squares minimization. Rendering the 3-D object model with the shaded texture map using the estimated parameters α^C_i leads to a model frame which approximates the lighting of the original frame.

3. EXPERIMENTAL RESULTS

Figure 3. Original camera image (upper left) and three images with artificial illumination for different α configurations.

In order to illustrate the performance of the proposed method, several experiments with real image data are conducted. First, the light map approach is used to create different lighting scenarios from a single image. For the video frame shown in the upper left of Fig. 3, a head model is created from a 3-D laser scan. Position, orientation, and facial expressions are estimated and the head model is moved and deformed appropriately to match the original frame [10]. Using surface normal information from the 3-D model, 50 light maps as shown in Fig. 1 are computed according to (3) using an array of 7 by 7 different illuminant direction vectors $\mathbf{l}_i$. A KLT is applied to the light maps and the first four eigen light maps shown in Fig. 2 are used for the experiment. By varying the lighting parameters α_i^C, the original image $I^C_{unshaded}(\mathbf{x})$ is scaled according to (4) leading to differently illuminated images $I^C_{shaded}(\mathbf{x})$. Fig. 3 shows three examples of image re-lighting for different parameter sets α_i.

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

In a second experiment, the inverse problem is addressed where lighting variations over time are removed from an image sequence. A head-and-shoulder sequence with 134 frames is recorded in CIF resolution. During the acquisition of the video sequence, one natural light source was moved to change the illumination in the scene. The upper row of Fig. 4 shows three frames of the sequence illustrating the variations in scene lighting. For that sequence, object motion and deformation caused by facial expressions are tracked over time using the model-based approach in [10] and synthetic frames are created by rendering the appropriately moved and deformed 3-D head model. Since the texture of this model is extracted from the first frame of the sequence and not updated afterwards, the lighting in the model frames remains constant. Therefore, these model frames can be used to determine the changes in scene lighting of the original sequence by solving the system of equations given by (4). $I^C_{unshaded}$ corresponds to the pixel intensities of a model frame while I^C_{shaded} refers to the pixels of an original frame that is illuminated differently at each time instant. The system of equations is solved in a least-squares sense providing for each frame a set of parameters α_i^C that specifies changes in the illumination relative to the first frame of the sequence. For this experiment, the first most important eigen light maps computed from the above mentioned set of 50 light maps are used. By applying the inverse scaling

$$I^C_{constant}(\mathbf{x}) = \frac{I^C_{shaded}(\mathbf{x})}{\sum_{i=0}^{N-1} \alpha_i^C L_i(\mathbf{u}(\mathbf{x}))} \tag{5}$$

to the original frames I^C_{shaded}, the variations in the lighting can be removed and a new image sequence $I^C_{constant}$ with constant illumination is obtained. This is illustrated on the lower row of Fig. 4 that shows three frames of the illumination-compensated sequence.

The accuracy of the illumination compensation is directly related to the particular light map configuration. To investigate these effects, we perform scene lighting estimation and compensation with a different number of eigen light maps. Like in the previous experiment, motion, deformation, and lighting parameters are estimated for the sequence shown in Fig. 4. Then, motion- and illumination-compensated frames are generated by rendering the deformed 3-D head model and applying (4) to approximate the original shading. The approximation accuracy is determined by comparing the synthetic model frames with the original sequence. By varying the number of eigen light maps used for the description of lighting effects, different reconstruction qualities can be obtained.

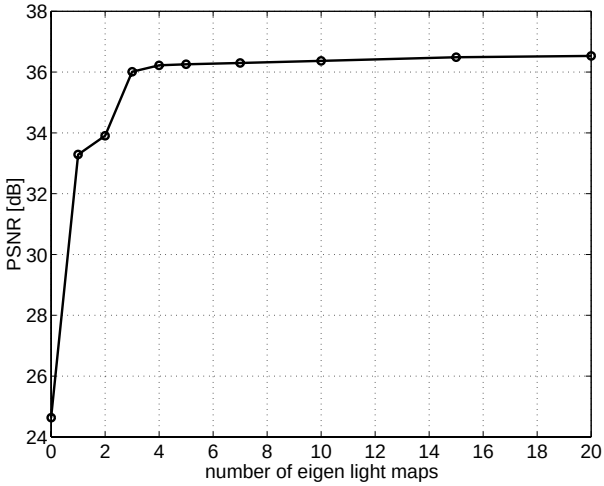

Figure 5. Reconstruction quality of motion- and illumination-compensated frames for a varying number of eigen light maps.

Fig. 5 shows the average reconstruction quality of the model frames measured in peak-signal-to-noise-ratio (PSNR) of the facial area. In this plot, the measurement using zero light maps corresponds to the experiment with no illumination compensation. One eigen light map refers to the case of simply scaling the three color channels appropriately. With an increasing number of light maps and thus lighting parameters, the accuracy of the model frames improves. However, the gain of adding a particular eigen light map depends on the current lighting in the scene, since each light source configuration is unequally well represented by the light map. Therefore, the curve in Fig. 5 need not be convex even if each additional light map improves the illumination compensation. For a large number of light maps, gains of up to 3.3 dB in PSNR are obtained compared to case when the image is simply scaled which requires no 3-D knowledge. Compared to the explicit 8 parameter light source model estimation described in [6], a gain of 0.95 dB in PSNR is observed for the proposed light map approach which requires 3 degrees of freedom for each additional light map. However, as indicated by the saturation effect of the curve in Fig. 5, a small number of light maps is sufficient to describe the dominant lighting effects.

Lighting compensation for an image sequence not only improves the visual quality in poorly illuminated scenes, but is also beneficial in video coding and transmission. If the illumination in the video varies the bit-rate required to encode the sequence at a certain quality increases drastically, since motion-compensated prediction becomes less efficient. When removing variations in the lighting from the sequence prior to its encoding, significant bit-rate savings can be obtained at the same reconstruction quality. To investigate this effect, the CIF video sequence in the upper row of Fig. 4 is encoded at 8.3 fps with an H.26L codec (test model TML 5.9 [14]). Additionally, the variations in scene lighting are removed resulting in the new sequence shown in the lower row of Fig. 4 which is encoded using the same parameter settings. Rate-distortion curves are measured for both sequences by varying the quantizer parameter over values 12, 14, 16, 20, 24, and 28. As shown in Fig. 6, the reconstruction quality measured in PSNR is higher for the illumination-compensated sequence. At the low bit-rate end, a gain of 2.9 dB in PSNR is achieved at the same average bit-rate. This corresponds to a bit-rate reduction of 42 % at the same average quality.

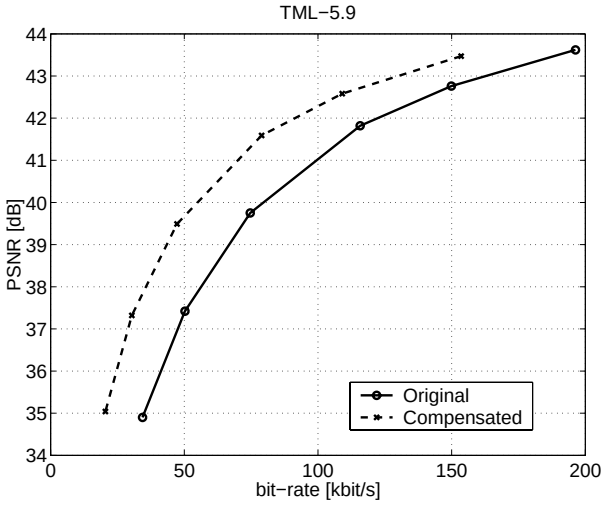

Figure 6. Rate-distortion plot illustrating the coding gain achieved by illumination compensation.

Fig. 7 finally shows decoded frames for both sequences encoded at the same average bit-rate of about 34 kbit/s. The left hand side of Fig. 7 depicts two decoded frames of the original video sequence. The corresponding frames of the right hand side are decoded from the modified sequence with less variations in the lighting and show less coding artifacts compared to the original sequence. Thus, model-based illumination compensation can be used as a preprocessing step to improve the visual quality of coded image sequences.

4. CONCLUSIONS

We present a 3-D model-based approach for estimating lighting variations in head-and-shoulder image sequences. Head pose and facial expressions of the person are tracked with a 3-D head model that provides surface normal information for the analysis of lighting variations. The shading of the model is achieved by linear superposition of multiple eigen light maps which are attached to the model surface. Experiments show that a small number of light maps is already sufficient to describe a wide variety of lighting effects. By comparing the pixel intensities of the original and the motion-compensated model frame, a system of linear equations is set up which is solved with low computational complexity for the parameters α_i^C representing the shading information. After having estimated the current lighting, variations over time can be eliminated by inverting the estimated scaling function. New image sequences are obtained that show constant lighting for all frames. When encoding these illumination-compensated sequences, significant improvements in coding efficiency can be observed. Experiments report an increase in PSNR of up to 2.9 dB at the same average bit-rate if the proposed illumination compensation is applied prior to encoding.

REFERENCES

1. M. Bichsel, "Illumination invariant object recognition", *Proc. International Conference on Image Processing (ICIP)*, vol. III, pp. 620–623, Oct. 1995.
2. Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, Jul. 1997.
3. R. Brunelli, "Estimation of pose and illuminant direction for face processing", Tech. Rep., MIT AI Lab, Technical Report No. 1499, 1994.
4. G. Bozdagi, A. M. Tekalp, and L. Onural, "3-D motion estimation and wireframe adaption including photometric effects for model-based coding of facial image sequences", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 3, pp. 246–256, June 1994.
5. J. Stauder, "Estimation of point light source parameters for object-based coding", *Signal Processing: Image Communication*, vol. 7, no. 4-6, pp. 355–379, Nov. 1995.
6. P. Eisert and B. Girod, "Model-based coding of facial image sequences at varying illumination conditions", in *Proc. 10th Image and Multidimensional Digital Signal Processing Workshop IMDSP '98*, Alpbach, Austria, Jul. 1998, pp. 119–122.
7. J. S. Nimeroff, E. Simoncelli, and J. Dorsey, "Efficient re-rendering of naturally illuminated environments", in *Proc Fifth Annual Eurographics Symposium on Rendering*, Darmstadt Germany, June 1994, pp. 359–373.

Figure 7. Decoded frames from two sequences encoded at the same average bit-rate. Left: original sequence with illumination variations. Right: image sequence encoded after illumination compensation.

8. S. R. Marschner, "Inverse lighting for photography", in *IS&T/SID Fifth Color Imaging Conference*, Scottsdale, AZ, Nov. 1997, pp. 262–265.
9. P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions", *International Journal of Computer Vision*, vol. 28, no. 3, pp. 245–260, Jul. 1998.
10. P. Eisert and B. Girod, "Analyzing facial expressions for virtual conferencing", *IEEE Computer Graphics and Applications*, vol. 18, no. 5, pp. 70–78, Sep. 1998.
11. C. A. Poynton, "Gamma and its disguises: The nonlinear mappings of intensity in perception, CRTs, film and video", *SMPTE Journal*, pp. 1099–1108, Dec. 1993.
12. C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces", in *Proc. Computer Graphics (SIGGRAPH)*, Jul. 1984, vol. 18, pp. 213–222.
13. M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal for Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
14. ITU-T Video Coding Experts Group SG16/Q6, "H.26L Test Model Long Term Number 6", http://standard.pictel.com/ftp/video-site/h26L/tml6d0.doc, Mar. 2001.