Relighting of Facial Video

Abstract

We present a novel method to relight video sequences given known surface shape and illumination. The method preserves fine visual details. It requires single view video frames, approximate 3D shape and standard studio illumination only, making it applicable in studio production. The technique is demonstrated for relighting video sequences of faces.

1. Introduction

Photo realistic relighting of video sequences is important in studio and film production. Beside the spectacular special effects of high budget movies, putting presenters into virtual environments during live studio broadcasts is becoming an often seen phenomenon in television. Our aim is to create a method which fits into the normal work flow of a television studio and is capable of altering the effect of lighting on the people in the studio. In the case of dynamic scenes this problem is difficult due to movement, requiring relighting at a single time instant. Approaches to re-illumination have generally only considered static scenes with known illumination and/or geometry. For static scenes it is possible to capture the scene under multiple illumination conditions. However, this is not possible for dynamic scenes, such as people and faces, where the subject moves and changes shape. In this paper, a method is presented for relighting frames of dynamic scenes.

There are several techniques which attempt to recover the full bidirectional reflection distribution function (BRDF) of a surface or a point. Image based techniques [1] yield the specular free image of the scene without recovering the BRDF. In the case of uncoloured, diffuse Lambertian objects, it is possible to recover the shape and the direction of illumination at the same time [2]. Shape based techniques require the 3D shape of the object in the scene to be known, however sometimes only diffuse reflection is considered [3] or a reference object of the same material is required to be relit and to be present in the scene [4]. In the case of a complex illumination, the problem of recovering the reflection properties of the objects in the scene is more complicated, hence many approaches restrict the illumination to a single point light source [5] or apply multiple point light sources to illuminate the scene [6].

The proposed technique combines the image based and shape based methods capable of handling high frequency textures like human skin. It handles diffuse and specular surfaces alike and accounts for self shadows as well. However, this approach cannot recover the full BRDF, but only the diffuse reflection parameters. Since we aim to apply the proposed method in a standard studio environment we reject the usage of special lighting or high speed cameras [7]. The approach allows us to relight dynamic scenes.

2. Estimating reflection parameters

In this paper we use the dichromatic [8]reflection model for our research, because it explicitly separates the specular and the diffuse reflection components:

$$I(\lambda)=\rho(\lambda)\int_L I_d(\lambda)\bar{n}\cdot\bar{l}_d+k_s\int_L I_s(\lambda,\sigma,\bar{n},\bar{l},\bar{v}) \quad (1)$$

The first part of expression 1 represents the diffuse part of the reflected light, where $\rho(\lambda)$ and $\bar{n}$ are the spectral reflectance (colour) and the surface normal at a given point of the surface of the object, respectively. L is the set of *effective light sources* - the light sources illuminating the surface point. $I_d(\lambda)$ and $\bar{l}_d$ are the spectral intensity and the direction of an elementary illuminant, respectively. The second part of expression 1 represents the specular reflection, where k_s is a scalar scaling factor for non-metallic materials and $I_s(\lambda,\sigma,\bar{n},\bar{l},\bar{v})$ represents the dependence of the specularity on the surface roughness (σ), the surface normal $(\bar{n})$, the illumination $(\bar{l})$ and the view direction $(\bar{v})$.

The diffuse component is independent of the view point of the observer, therefore it can always be substituted by a virtual point light source:

$$I(\lambda)=\rho(\lambda)I_D(\lambda)\bar{n}\cdot\bar{l}_D+k_s\int_L I_s(\lambda,\sigma,\bar{n},\bar{l},\bar{v}) \quad (2)$$

The colour of the diffuse reflection depends on the spectral reflectance of the surface point and the colour of the illumination. For non-metallic surfaces, k_s can be assumed to be independent of λ, therefore the specular reflection component is independent of the surface colour and only depends on the illumination colour.

Our aim in the further sections is to remove the effect of specular reflection from the facial images and to introduce a framework to estimate the diffuse spectral reflection of each point on the surface of an object. In this paper, we introduce a three steps relighting process; first the specular highlights are removed from the image using an image based technique resulting in a *natural image;* second we estimate diffuse saturation for each colour region of the facial image; third we estimate colour albedo at each point of the scanned face.

2.1.Monochrome light assumption

In the case of illumination with multiple coloured light sources and constant view point, the colour of specularly reflecting light changes across the surface with the change of surface normal. The perceived colour of a surface point is the combination of the diffuse and the specular colour.

In the case of a monochromatic illumination the colour of the specular reflection is constant and equivalent to the colour of the illumination. The colour of the diffuse reflection is the combination of the colour of the object surface and the colour of the illumination. If the colour of the illumination is known, the effect of the coloured illumination can be removed and turned into a white light illumination.

$$I_w(\lambda)=\frac{I(\lambda)}{r(\lambda)}=\rho(\lambda)I_D\bar{n}\cdot\bar{l}_D+k_s\int_L I_s(\sigma,\bar{n},\bar{l},\bar{v}) \quad (3)$$

where $r(\lambda)$ is the spectral distribution of the illumination and I_D and I_s are the intensity values.

The specular reflection of the spectrally normalised illumination, in expression 3, is always white, whilst the diffuse reflection preserves the original material colour. This means that the stronger the specular reflection the whiter the reflected light. The effect of white light specularity can be better illustrated using the HSV colour model. The hue component of the reflected light remains unchanged by white specularity, however the saturation of $I_w(\lambda)$ decreases as the intensity of the specular term increases.

The intensity of the reflected light in a combined diffuse-specular case:

$$I=k_d I_D+k_s I_S \quad (4)$$

where k_d is the diffuse reflection coefficient. We define saturation as :

$$S=1-3M/I \quad (5)$$

where $M=min(R,G,B)$ and *R,G,B* are the red, green blue pixel values, respectively. By substituting expression 4 into expression 5 the relationship between intensity and saturation can be derived:

$$SI=k_d I_D(1-M) \quad (6)$$

Equation 6 shows that, under the assumption of monochromatic illumination, the *SI* product is independent of the specularity. Hence it must have the same form under diffuse only illumination

$$S_D I_{diffuse}=k_d I_D(1-M) \quad (7)$$

where $I_{diffuse}=k_d I_D$ and in the case of a white light illumination S_D is equivalent to the saturation of the surface material. Substituting expression 6 into expression 7 we get:

$$SI=S_D I_{diffuse}=S_D k_d I_D=S_D k_d I_d \bar{n}\cdot\bar{l}_d \quad (8)$$

Since the right hand side of expression 8 depends only the diffuse parameters, the *SI* product is not dependent on the viewpoint of the observer.

2.2.The natural image

The *SI* product depends on the distribution of the illuminant, the shape and the diffuse reflectivity of the material making up the surface. The *SI* product for each object point can be calculated from a colour image of the object. If the illumination field and 3D object shape are known, the diffuse reflectance of the object can be recovered form a single image.

$$S_D k_d=SI/(I_D\bar{n}\cdot\bar{l}_D) \quad (9)$$

where *S* and *I* are the saturation and intensity of a point calculated as a result of the HSV conversion of the colour image. The $S_D k_d$ product is a illumination independent material property. The image of the object replaced by the $S_D k_d$ product is an illumination independent representation of the material of the object, therefore we call it the *natural image*.

For capturing 3D shape of dynamic objects, such as the human face, we developed an active light multiple new stereo capture system, that uses infra red patter projection to simultaneously capture shape and colour at video rate and produces detailed triangular mesh representation [9]. The illumination environment is captured by taking a series of images of a mirror ball with different shutter speed and reconstructing the *light source image* as a high dynamic range image. Each pixel in the light source image acts as a light source.

2.3.The colour albedo

The natural image represents properties related to the reflectivity of the material, however it does not contain any information about the spectral distribution of the reflected light. The object surface does not reflect light uniformly in all frequencies. Using the RGB colour model the spectral reflectivity of a material is $\rho(\lambda)=k_d\cdot(r_{\lambda_R},r_{\lambda_G},r_{\lambda_B})$ and $\|\rho(\lambda)\|=k_d$, where $r_{\lambda_{R,G,B}}$ are the *normalised spectral reflectance values* in the frequency range of the red, green and blue sensors, respectively. Due to the white light assumption the hue component of the colour of the reflected light is equivalent to the hue of the surface point, whilst the saturation of the reflected light is dependent on the amount of specularity. In the case of diffuse only reflection, the saturation of the reflected light, S_D is equivalent to the saturation of the colour of the surface point. Surface points move on a hyperbola in the saturation-intensity space. A colour pixel cannot be turned back to its diffuse state as long as its diffuse saturation value is not known. According to expression 9, S_D cannot be estimated directly, since it cannot be decoupled from $k_{d,}$ without extra information. This information is acquired by a region based approach.

If the object surface is made of a small number of uniformly saturated patches and the spatial distribution of the effective light sources is sparse, then it can be assumed that the reflected light from most of the surface area of a patch is diffuse, hence the saturation of the reflected light is equivalent to S_D. The peak of the histogram (S_{peak}) of all saturation values of a uniformly saturated patch provides a good approximation $S_D \approx S_{peak}$ of the saturation of the diffuse reflection for that surface patch.

Points with lower saturation value than the S_{peak} must be partly specular, however points with higher saturation value than S_{peak} cannot be specular. Hence the diffuse saturation for each point of a near uniformly saturated surface patch is

$$S_D = \begin{cases} S_{peak} & if \quad S < S_{peak} \\ S & if \quad S \geq S_{peak} \end{cases} \qquad (10)$$

By substituting S_D into expression 9, the value of the diffuse reflectance, k_d, is obtained. The estimated value of the diffuse saturation and the point based hue value of the surface colour make up the normalised spectral reflectance values of each surface point. The colour triplet of the spectral reflectances act as colour albedo.

If the object surface is detailed, the assumption of a small number of uniformly saturated surface patches does not hold, then the histogram technique cannot be applied directly. However, many materials can be approximated as near uniformly saturated with high frequency disturbances, such as the human skin. Skin has a dominant colour riddled with small anomalies caused by wrinkles, hair or discolouration. In the case of such material a modified technique is applied. If the total area of all disturbances are small compared to the area of the surface patch, the histogram technique still gives a good approximation of the diffuse saturation of the patch, however S_D cannot be applied to the disturbed points. To eliminate this problem, first the anomalies are removed by applying a median filter to the colour images. This removes small deviations from the material colour whilst minimising the effect of blurring. Second, the diffuse reflectance is estimated by finding the peak of the saturation histogram, approximating S_D:

$$S_D = \begin{cases} S_{peak} & if \quad S < S_{peak} \\ S_{median} & if \quad S \geq S_{peak} \end{cases} \qquad (11)$$

This sets the diffuse saturation near uniform. This value does not contain the small variation caused by the disturbances in the material colour. In this case, the material is a good approximation of the uniformly saturated patch assumption. Finally, to preserve surface details features removed by median filtering are reintroduced by modifying the value of the diffuse reflection coefficient:

$$k_d = k_d^{median} I / I_{median} \qquad (12)$$

Hence the original disturbances are represented as intensity variation, however their colour is lost.

2.4. Temporal consistency

The quality of a processed video sequence can be reduced by noise in the consecutive video frames. In this section, a method is presented to enforce temporal consistency between individual video frames while preserving the details of the surface.

The two major sources of noise are the image and the shape noise. Due to the thresholding in equation 11, image noise can result in small variations in the consecutive video frames. Shape noise causes the object surface to wave as randomly appearing bumps. Cast shadows and the noise in the estimated surface normals cause the surface intensity to vary.

The temporal filtering algorithm is executed on the albedo images. The 3D meshes of the face in two consecutive frames are rigidly aligned by establishing a transformation between them [9]. A correspondence via shape algorithm is applied to link pixels in consecutive video frames. The linked pixels are median filtered in a predefined neighbourhood to remove noise while preserving changes in time. This technique transforms shape noise into a correspondence noise. To reduce this, the pixel wise depth-map, used in the correspondence via shape algorithm, is smoothed.

To preserve details and avoid blurring, even in the case of significant correspondence error, temporal consistency is only enforced on the albedo image before the reintroduction of the surface details. The surface details are reintroduced after the temporal filtering, therefore equation 12 changes to

$$k_d(t) = F_w^{temporal}(k_d^{median}(t)) I(t) / I_{median}(t) \qquad (13)$$

where $F_w^{temporal}$ is the filtering in a w wide temporal window and t is the index of the video frame.

3. Results

Figure 2 depicts the original colour image. The middle and right images in figure 2 show the natural image and the recovered colour albedo. The colour albedo image is free from the effects of specularity and depicts the diffuse colour and reflectivity of each pixel. The colour albedo serves as a basis for further

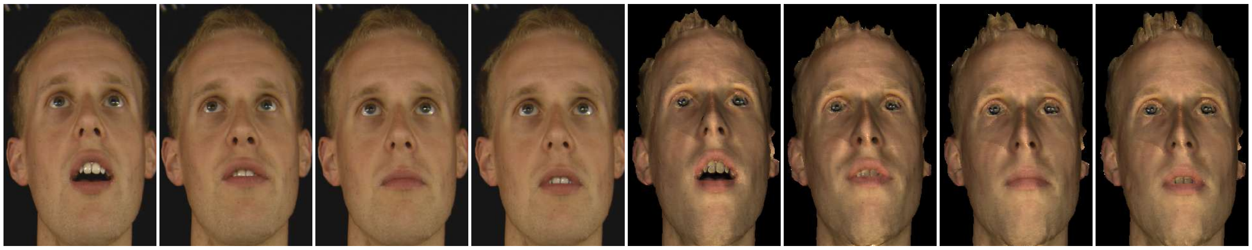

Figure 1 left, four frames of a captured sequence; right the same frames relit by a combination of three point light sources, one in the middle, one offset to the right and a lesser intensity source to the left.

relighting as shown in figure 3. The images in figure 3 were generated by applying the uniformly saturated surface assumption to the whole of the image including the eyes, lips and the hair. The results verify that the assumption is valid for most parts of the face.

Figure 4 illustrates relighting with complex light fields. In the beach image, the main illuminant is the sun low above the horizon, whilst in the forest image the light of the sun is filtered by the forest resulting in a irregular illumination.

Figure 1 shows four frames of a dynamic sequence and the same frames relit using a combination of three point light sources simulating candle light. Self shadowing is visible around the nose. The face deforms during speech resulting in changes of the relit model. This effect is visible in the left of the images over the lips and also below the nose.

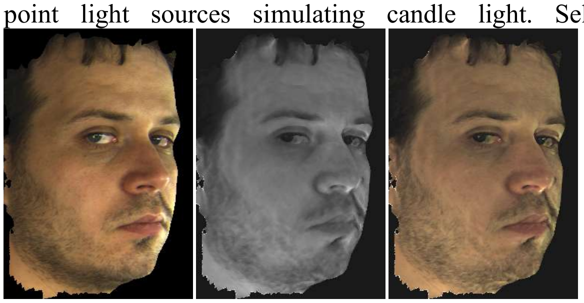

Figure 2 left, original colour image; middle, natural image, right, colour albedo image

Figure 3 left, diffuse rendering using a point light source illumination placed in the right of the image. Middle, diffuse rendering using a point light source in the centre of the image. Right, same as middle with added specularity visible mainly on the forehead and the cheek.

Figure 4 left, the original image; middle and right the face rendered in a beach and in a forest environment.

4. Conclusion and future work

The results demonstrate realistic relighting with preservation of fine visual detail. The approach developed is suitable for relighting moving objects such as faces from the captured image sequence. We have shown that relighting of scenes from a single colour image is possible if the illumination field and the 3D object shape are known. The presented technique fits into the standard work flow of studio production, however further enhancements are needed to achieve photorealism.

5. References

[1] R.T. Tan and K. Ikeuchi, "Separating reflection Components of textured Surfaces using a Single Image", IEEE T. Pattern Anal., 2005, pp. 178-193

[2] D. Samaras and D. Metaxas, "Incorporating Illumination Constraints in Deformable Models for Shape from Shading and Light Direction Estimation", IEEE T. Pattern Anal., 2003, pp. 247-263

[3] C.B. Madsen and R. Laursen, "A Model-Based Approach to Image Relighting with a Potential for Real-time Implementation", Int. Conf. VVG, 2005, pp. 237-243

[4] A. Hertzmann and S.M. Seitz, "Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs", IEEE T. Pattern Anal., 2005, pp. 1254-1264

[5] Y. Sato, M.D. Wheeler and K. Ikeuchi, "Object shape and reflectance modelling from observation", Proc. SIGGRAPH97, 1997, pp. 379 - 387

[6] S. Tominaga and B.A. Wandell, "Standard surface-reflection model and illumination estimation", J.Opt.Soc.Am., 1989, pp. 576-584

[7] A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins and Paul Debevec, "Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination", Proc. SIGGRAPH05, 2005, pp. 756-764

[8] Y. Yu and J. Malik, "Recovering Photometric Properties Of Architectural Scenes From Photographs", Proc. SIGGRAPH98, 1998, pp. 207-217

[9] I.A. Ypsilos, A. Hilton, and S. Rowe, "Video-rate Capture of Dynamic Face Shape and Appearance", IEEE Int. Conf. FG, 2004, pp. 117- 122