Research White Paper

WHP 143

11 January 2007

Multi-camera radiometric surface modelling for image-based re-lighting

Oliver Grau

BRITISH BROADCASTING CORPORATION

Multi-camera radiometric surface modelling for image-based re-lighting

Oliver Grau

Abstract

This contribution describes an automatic method to retrieve the diffuse radiometric surface model of moving persons or other objects along with the object geometry using a multi-camera system. The multi-camera equipped studio allows synchronised capture of the foreground action and a visual hull computation is then used to compute a 3D model of that scene. The diffuse surface reflection parameters are computed using the 3D model from that process together with an illumination map of the studio. The illumination map is a high dynamic range image generated from a series of images of the studio using a camera equipped with a spherical (fish-eye) lens. With this setup our method is able to capture any action in the studio under normal lighting.

This document was originally published in Proc. of 28th DAGM Symposium, September 12-14, 2006, Berlin, Germany.

Technical Note number: R&D 143

White Papers are distributed freely on request.

Authorisation of the Head of Research is required for publication.

Multi-camera radiometric surface modelling for image-based re-lighting

Oliver Grau

BBC Research, Kingswood Warren, Tadworth, UK,
oliver.grau@rd.bbc.co.uk,
WWW home page: http://www.bbc.co.uk/rd

Abstract. This contribution describes an automatic method to retrieve the diffuse radiometric surface model of moving persons or other objects along with the object geometry using a multi-camera system. The multi-camera equipped studio allows synchronised capture of the foreground action and a visual hull computation is then used to compute a 3D model of that scene. The diffuse surface reflection parameters are computed using the 3D model from that process together with an illumination map of the studio. The illumination map is a high dynamic range image generated from a series of images of the studio using a camera equipped with a spherical (fish-eye) lens. With this setup our method is able to capture any action in the studio under normal lighting.

1 Introduction

The integration of virtual, synthetic objects into real scenes has many applications in film and TV productions, for product and architectural visualisation. For highly realistic results it is important to match the lighting of the real with the synthetic components. Methods to capture the scene lighting of environments using high dynamic range images (HDRI) were pioneered by Paul Devebec [1] and are now widely used in production. These methods involve building up a panoramic representation, by either mapping the environment onto a sphere or a cube. The inserted virtual object is then lit by this HDR illumination map and pasted into the background plate, which is an image of the real scene.

Any object to be inserted into the HDR environment needs to have a surface description that gives the surface properties along with the object shape (or geometry) for the rendering system. This might be the BRDF (bidirectional reflectance distribution function) in the general case or more simplified colour, diffuse and specular reflection parameters. For synthetic objects these parameters are usually assigned manually in the rendering system.

The accurate measurement of the BRDF requires a very defined environment. Approaches usually assume a calibrated environment where camera parameters and object shape are precisley known, e.g. by using a laser scanner [2–4]. Further a fixed point light source with known position is used for the computation. Debevec et al. [5] described a method to capture an image based representation of a face by taking a series of images under controlled variation of a light source.

The setups for the mentioned approaches are designed to acquire the surface properties offline and are not able to capture a live action.

The aim for the approach described in this paper is to integrate a sequence of images taken from an actor in a studio into a different lighting situation. This task is also called re-lighting. Since the actor is moving it is not possible to use offline modelling tools like a laser scanner to create highly accurate 3D geometry. Furthermore the dynamic range of the images is limited (8-bit). The approach described here uses a multi-camera system that can capture images of a moving actor synchronously in a chroma-keying environment. A 3D model of the actor is generated using a visual hull computation. A brief outline of this step is given together with an overview of our approach in the next section. A more detailed description of the studio setup can be found in [6]. In addition to the 3D geometry an illumination map is created using spherical (fisheye) HDR images of the studio. The use of the illumination map makes our approach applicable to any lighting situation found in realistic production scenarios. For the computation of the surface properties we currently focus on the diffuse component.

The remainder of this paper describes a method for calibrating the radiometry of the camera system in section 3. Section 4 introduces a new method for the robust computation of radiometric surface properties from multi-camera images. The paper finishes with some results and concluding remarks.

2 Overview

The computation of radiometric surface properties from multi-camera images consists of the following processing steps:

1. Geometrical calibration of the multi-camera system
2. Foreground/background segmentation
3. 3D reconstruction using a visual hull computation
4. Radiometric camera calibration
5. Capture of an illumination map of the studio lighting environment
6. Computation of radiometric surface properties

The steps 1-3 are common practice to generate a 3D polygonal surface model O of the actor or object: The *geometrical calibration* uses a 1m x 1m planar chart. The cameras used for our experiments were Sony DXC-9100P cameras operating in 25 fps progressive mode. From a set of 10-20 multi-camera still frames the calibration method computes a set of camera parameters simultaneously considering centre-point shift and radial distortions. The *segmentation* is based on a chroma-keying facility that is available in the experimental studio, but other methods are applicable (e.g. difference keying). The *visual hull computation* is based on a hierarchical octree approach similar to [7, 8] with refinement using super-sampling and (moderate) Gaussian smoothing [9] followed by a marching cubes iso-surface generation.

The *radiometric camera calibration* consists of two steps: The set-up phase is basically a colour balance that defines the operating mode of the camera.

Fig. 1. Flow diagram of the transfer functions s and f that transform the scene radiance into image brightness (adapted from [10]).

The radiometric calibration estimates the transfer function of the cameras that relates the scene radiance L to the image brightness I measured by the cameras. Fig. 1 explains this concept (see [10] for more details); The image brightness I caused by a scene radiance L is here defined by the optical transfer function s and the camera response function f:

$$I = f(E_i) \tag{1}$$

$$E_i = s(L) \tag{2}$$

With the image irradiance E_i.

The optical transfer function s considers optical effects, like vignetting, lens aberrations, depth of focus and effects like fixed pattern noise. In this work these effects are not considered, with exception of vignetting for the capture of illumination maps. The camera response function f however is very important, because the cameras usually have non-linear response. In the case of the broadcast-style Sony video cameras used, this is a deliberate feature (gamma characteristic) to compress the irradiance range. The inverted camera response f^{-1} gives the relation of image brightness to the scene radiance (ignoring optical effects here).

A *map of the illumination* B is generated using images from the studio taken with a spherical lens (fisheye lens) and standard HDRI techniques as outlined in the next section.

With the 3D model O of the foreground object and the illumination map B we can now define a model that describes the expected observation of the illumination of the foreground object from the observation cameras. Fig. 2 illustrates this set-up.

The last step is the *computation of radiometric surface properties* that computes the dominant diffuse surface reflectance parameters. This approach is detailed in section 4.

3 Multi-camera radiometric calibration and capture of illumination map

The goal of the radiometric camera calibration is to colour balance the multi-camera system and to determine the response functions of the individual cameras. The colour balance sets for each camera the red, green and blue channel

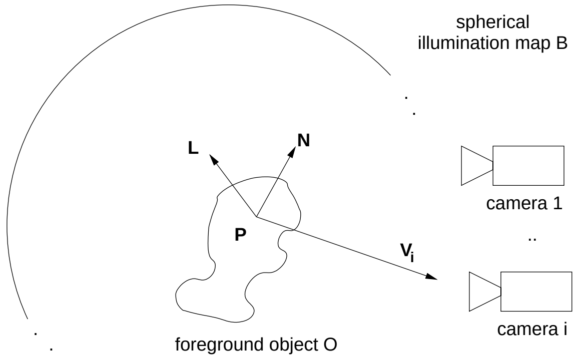

Fig. 2. Relation of a point P on the foreground object O and the spherical background model B. **N** represents the surface normal of the surface point, **L** represents the direction of an incoming light ray and $\mathbf{V}_i$ is a vector pointing to camera i.

to a reference white and black object in the scene that are visible to all cameras simultaneously. As reference object a Macbeth $Colorchecker^{TM}$ chart for the white and black level was used. A program has been developed to set the radiometric parameters available in the cameras used [1]. The colour balance determines the range $[E_{i,min}, E_{i,max}]$ of the image irradiance that is mapped to the dynamic range of the camera, which is 8-bit per channel after digitisation. All scene objects with a radiance lower than the black or higher than the white point are clipped.

The second step of the radiometric calibration is to estimate the camera response function f for each camera. We are using the method described in [11]. This method computes a discrete lookup table for the camera response function from a number of images taken under varied camera exposure[2] of a static scene. This process is also known as radiometric auto-calibration. The more recent method described in [10] has the advantage that it is assuming a monotone response function (which is usually the case) and a restricted number of parameters. This makes the estimation of the parameters better conditioned.

The illumination map of the lighting situation is created using a digital stills camera (Nikon D100) equipped with a spherical lens (Costal Optics 185^o field of view). Fig. 3 shows a picture taken in the experimental studio. A series of pictures is taken with different exposure times (as describe above) so that even the brightest lights are mapped into the dynamic range of the digital camera (i.e. not over exposed) and detail is retained in the dark areas. This method is called bracketing. After radiometric auto-calibration of the camera the series of

[1] The Sony DXC-9100P cameras allow only for setting of gain for red and blue and a 'master pedestal' (black level)

[2] Preferably the exposure time (integration time) is varied over changing the lens aperture. In addition a neutral density filter is used to allow the capture of highlights in the studio lamps.

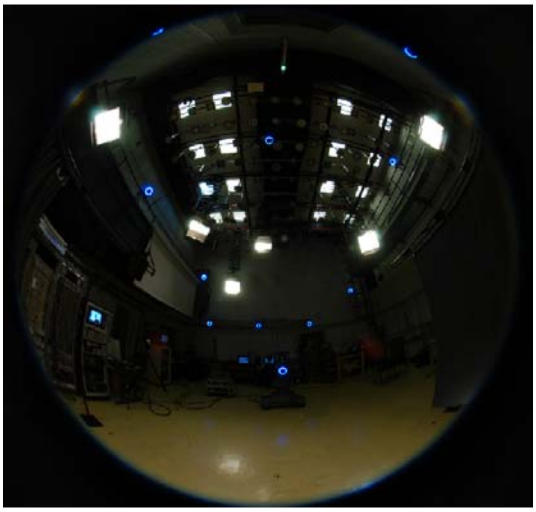

Fig. 3. A picture of the studio taken with a Costal Optics spherical lens showing all major light sources.

images is combined into one HDRI, as described in [11]. The spherical images are transformed into a latitude-longitude mapping for further use.

For the use as a model of the illumination the position and orientation of the spherical probe camera and its internal parameters have to been known. This is done by registering the camera to the coordinate reference system of the multi-camera system. For this purpose a number of known positions in the studio are manually extracted in the spherical camera images and a modified calibration method is used to compute the camera pose and internal parameters.

The illumination map will be taken roughly from the position of the actors. The approximation of the studio lighting is assuming that the extent of the acting area is small compared to the distance of the lights to the probing camera. For many practical configurations it is sufficient to use only one (half) sphere since all the major light sources can be captured. This is neglecting any light bouncing from the other half sphere which can not always be tolerated. In this case both hemispheres have to be provided.

4 Computation of radiometric surface properties

Using the configuration shown in Fig. 2 the expected appearance of a foreground object can be computed using its computed 3D shape O and the illumination map B by rendering a synthetic view for the viewpoint of camera i. Instead of using a general BRDF we are using a simplified reflection model, as common in many rendering systems to compute the intensity of a pixel I:

$$I = k_d I_d + k_s I_s + I_e \tag{3}$$

with the diffuse reflection coefficient k_d, I_d the diffuse component, I_s the specular component and I_e the error term that covers the (residual) model error. The determination of the specular reflection coefficient k_s requires quite accurate surface normals and is neglected here, ie. $k_s = 0$.

The diffuse component can be computed as the irradiance E by integrating all incoming light rays L_{in}:

$$I_d = E = \int_{\Omega(\mathbf{N})} L_{in}(\omega)(\mathbf{N}\omega)d\omega \tag{4}$$

in the direction ω and Ω upper hemisphere over surface point with the surface normal $\mathbf{N}$.

The intensity of the light rays L_{in} is taken from the illumination map. The computation of I_d can be done with standard rendering systems that provide global illumination rendering.

An often used strategy for computing global illumination is to generate a number of rays in random directions from the 3D point on the object surface. For accurate results many rays are necessary and this requires long computation times (up to several hours per frame). Therefore a specialised renderer has been developed that gives a fast computation from the illumination map: For each 3D point on the object surface the irradiance is determined by integrating directly over the related area in the illumination map taking occlusions into account. The area corresponds to the surface hemisphere, i.e. the surface normal marks the centre point of this area. This method is reducing the computational effort significantly (typically in the order of minutes on a recent PC).

The parameter k_d has three components (one for each colour channel) and can be approximated:

$$k_d = \frac{I - I_e}{I_d} \tag{5}$$

The error term I_e is used to compensate for errors in the illumination model and is a global parameter here.

In the case that the foreground object has a surface with significant specular reflection these will appear as highlights in the camera images, as depicted on the head in the left image of Fig. 6. A significant highlight is usually overexposed, ie. the image brightness values are clipped. Under these conditions it is not possible to recover the surface colour. However, it is possible to recover the colour from a different camera. This is done by determining the 3D coordinate of a pixel by reprojecting it onto the 3D model. In the next step this point is projected into all available images. Omitting all invisible surface elements (using a depth test) a set of colour values for the surface element can be retrieved. The highlight can be replaced by the median value from the list.

5 Results

Fig. 4 shows results of a test production. A scene with a boy was captured using 12 cameras simultaneously (just one camera image shown). Using a visual hull computation as outlined in section 2 this gives a 3D model of the scene as depicted in the top right of Fig. 4. The model is made of 3000 triangles and shows some typical artefacts due to principle limitations of the visual hull reconstruction.

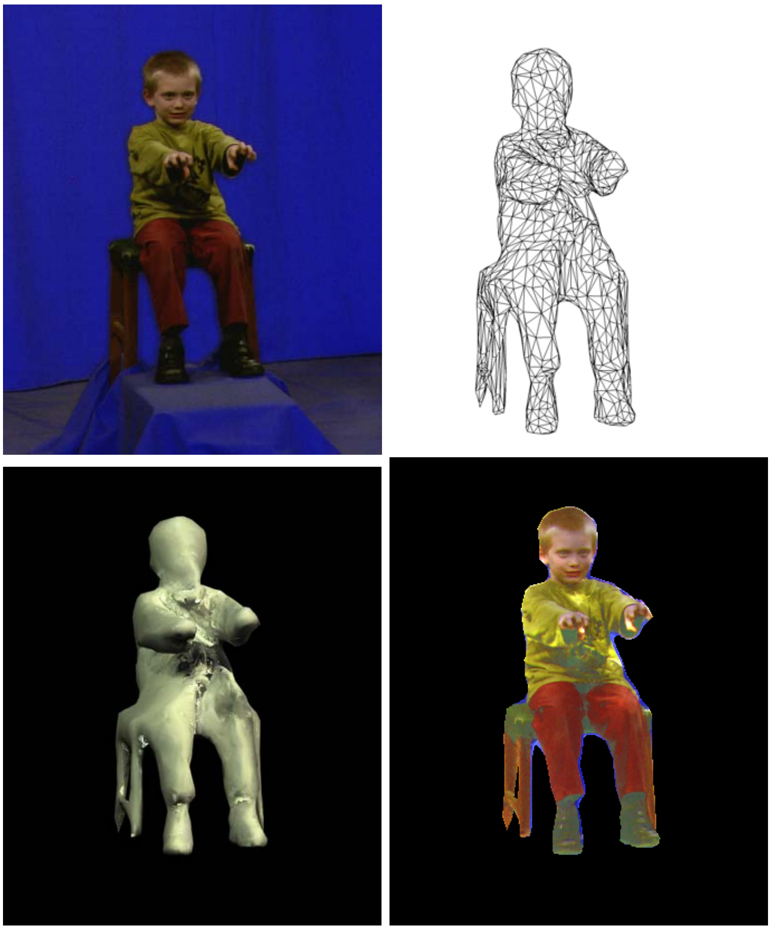

Fig. 4. Input image (top left) and wireframe rendered image of the 3D model with 3000 triangles (top right). The irradiance image (bottom left) and the estimated diffuse reflectance (bottom right).

The picture on the left bottom of Fig. 4 shows the irradiance image using the 3D scene model and an illumination map of the studio captured as described in section 3. The bottom right picture shows the computed diffuse reflectance for the input image. It can be clearly seen that most of the shading effects have been compensated. The remaining problems (like the area on the boy's chest) are mainly due to errors in the visual hull.

Fig. 5 gives an example of the usage of the computed reflectance model in a different illumination environment. The boy is sitting on a 'space scooter' and is moving forward from the inside of a room (top image) into bright sunlight. The pictures are rendered with Cinema 4D (a commercial animation and rendering package). The bottom left of Fig. 5 shows the use of the original camera image in this situation and the right image shows the use of the reflectance map computed with our method. It can be seen that the use of our reflectance model is producing more realistic results under these changed lighting conditions. A video with the results can be found in [12].

Fig. 5. An example of usage of the reflectance model. The top image shows the boy in a room. In the bottom he is rendered in bright sunlight using the original image (left) and our reflectance model (right).

Fig. 6 depicts an example of a person with a more shiny skin. In this case the specular components appear as highlights in the image (left). The image on the right shows the elimination of these highlights by using image information from the other cameras of the multi-camera system as described in the previous section.

Fig. 6. Compensation of specular hightlights by using image information from the multi-camera system.

6 Conclusions

This contribution described a method to estimate the diffuse reflectance parameters of actors captured with a multi-camera system by using an illumination map and a 3D model of the actors.

The results in the previous sections have shown that the proposed approach is increasing the range in which the illumination can be changed from the original studio lighting. The additional operational overhead for achieving that is relatively low since only the illumination map has to be captured in addition to the set-up of the multi-camera system.

A limiting factor of the method is the quality of the 3D models in this approach. In particular the surface normals that can be derived from the visual hull computation are not very precise. The diffuse component of the reflection can still be computed quite robustly since it is integrating over the hemisphere of each object surface point. The specular components are very sensitive to wrong surface normals. Therefore this paper was focussing on the diffuse components.

However more work will be carried out in the future to increase quality of the surface reflectance parameters. This will mainly target the accuracy of the 3D reconstruction that would allow better estimation of the surface normals and finally the consideration of the specular components.

Acknowledgements

I would like to thank my colleagues Susannah Fleming and Lloyd Lukama for their work on camera calibration.

References

1. Debevec, P.E.: Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Proceedings of SIGGRAPH 98, Computer Graphics Proceedings, Annual Conference Series, Orlando, USA (1998) 189–198
2. Sato, Y., Wheeler, M.D., Ikeuchi, K.: Object shape and reflectance modeling from observation. Computer Graphics **31**(Annual Conference Series) (1997) 379–388
3. Marschner, S.R., Westin, S.H., Lafortune, E.P.F., Torrance, K.E., Greenberg, D.P.: Image-based brdf measurement including human skin. In: In Proceedings of 10th Eurographics Workshop on Rendering, Granada, Spain (1999) 139–152
4. Lensch, H., Kautz, J., Goesele, M., Heidrich, W., Seidel, H.P.: Image-based reconstruction of spatial appearance and geometric detail. ACM Transactions on Graphics **22**(2) (2003) 234–257
5. Debevec, P., Hawkins, T., Tchou, C., Duiker, H., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: Proceedings of SIGGRAPH 2000, Computer Graphics Proceedings, Annual Conference Series. (2000)
6. Grau, O., Pullen, T., Thomas, G.A.: A combined studio production system for 3-d capturing of live action and immersive actor feedback. IEEE Transactions on Circuits and Systems for Video Technology **14**(3) (2004) 370–380
7. Potmesil, M.: Generating octree models of 3D objects from their silhouettes in a sequence of images. Computer Vision, Graphics and Image Processing **40** (1987) 1–29
8. Szeliski, R.: Rapid octree construction from image sequences. CVGIP: Image Understanding **58**(1) (1993) 23–32
9. Grau, O.: 3d sequence generation from multiple cameras. In: Proc. of IEEE, International workshop on multimedia signal processing 2004, Siena, Italy (2004)
10. Grossberg, M., Nayar, S.: Modeling the Space of Camera Response Functions. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(10) (2004) 1272–1282
11. Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: In Proceedings of SIGGRAPH 97. (1997) 369–378
12. Grau, O.: Capturing of light and surface reflection properties for re-lighting. http://www.bbc.co.uk/rd/projects/virtual/lightcapture (2006)