Volume 25 (2006), number 1 pp. 29–51 COMPUTER GRAPHICS *forum*

Classification of Illumination Methods for Mixed Reality

Katrien Jacobs and Céline Loscos

Department of Computer Science, University College London, UK
{K.Jacobs,C.Loscos}@cs.ucl.ac.uk

Abstract

A mixed reality (MR) represents an environment composed both by real and virtual objects. MR applications are used more and more, for instance in surgery, architecture, cultural heritage, entertainment, etc. For some of these applications it is important to merge the real and virtual elements using consistent illumination. This paper proposes a classification of illumination methods for MR applications that aim at generating a merged environment in which illumination and shadows are consistent. Three different illumination methods can be identified: common illumination, relighting and methods based on inverse illumination. In this paper a classification of the illumination methods for MR is given based on their input requirements: the amount of geometry and radiance known of the real environment. This led us to define four categories of methods that vary depending on the type of geometric model used for representing the real scene, and the sdifferent radiance information available for each point of the real scene. Various methods are described within their category.

The classification points out that in general the quality of the illumination interactions increases with the amount of input information available. On the other hand, the accessibility of the method decreases since its pre-processing time increases to gather the extra information. Recent developed techniques managed to compensate unknown data with clever techniques using an iterative algorithm, hardware illumination or recent progress in stereovision. Finally, a review of illumination techniques for MR is given with a discussion on important properties such as the possibility of interactivity or the amount of complexity in the simulated illumination.

Keywords: Augmented Reality, common illumination, relighting, inverse illumination

ACM CCS: 1.3.6. Computer Graphics: *Methodologies and Techniques* 1.3.7 Computer Graphics: *Three-Dimensional Graphics and Realism* 1.3.9 Computer Graphics: *Miscellaneous: Common illumination, relighting, inverse Illumination* 1.4.8 Image Processing and Computer Vision: *Scene Analysis*

1. Introduction

To understand the concept *mixed reality* it is necessary to classify the different types of environments that can be generated with a computer. Milgram and Kishino [1,2] present such classification based on the amount and type of virtual and real elements that constitute the resulting world. In their classification, all possible environments form one continuum called *reality–virtuality continuum (RV)*, see Figure 1. In this continuum, four worlds can be identified that have an outspoken character. These four worlds lie next to each other in the RV continuum and might even overlap. The first and most straightforward of these is the real world without any addition of virtual elements; it will be referred to as *reality* and it lies on the left end of the RV continuum. In the second world, virtual elements are added to a real scene. This world is referred to with the term *augmented reality (AR)* [3–5]. In an opposite scenario, the world consists of a virtual environment, augmented with real elements. This world is consequently called an *augmented virtuality (AV)*. The last and fourth world does not contain any real elements and is therefore labeled as a *virtual environment (VE)*; it lies on the right end of the RV continuum. The term *mixed reality (MR)* refers to those worlds that are a mix of virtual and real elements, or, MR spans the RV continuum. In general, methods that are developed for AR

Submitted November 2004
Revised August 2005
Accepted October 2005

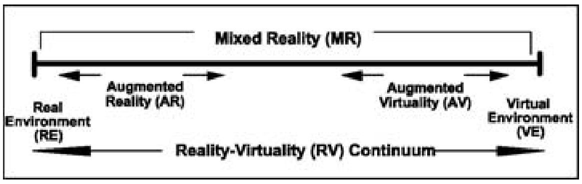

Figure 1: *Simplified representation of a reality–virtuality continuum [1,2]. (Courtesy of Milgram and Kishino)*

focus on real-time applications. Therefore they usually differ from methods that are specifically designed for MR applications whose focus can be on non-real-time applications. This paper will discuss the various existing illumination methods for MR applications in general.

Two different classes of AR exist; they differ in the realization of the AR [1]. The first class groups the methods for semi-transparent or see-through displays, examples are [6,7]. This first class contains two different see-through display methods. The first method (optical AR method [3]) projects the virtual objects on a transparent background, most likely the glasses of goggles. The second method (video AR method [3]) uses a head-mounted display: a head-mounted camera records the environment and this image is projected inside the display together with the virtual objects. The second class of AR replaces the expensive see-through devices with a nonimmersive display; it is usually called a computer-augmented reality (CAR). The quality of the immersion is higher for the first class than for the second. Nevertheless, see-through devices are not always required by the application: urban planning, architecture and some applications in the entertainment industry are satisfied with the second type of CAR display methods. In earlier approaches of AR, virtual objects were positioned on top of a real environment. Calibration and registration are difficult processes and for long the focus lied upon taking into account the possible occlusion and collision effects, while no further adaptations on real and virtual objects were carried out. In other words, after the inclusion, no resulting shadows were generated and no lighting changes were put through. An AR system of such kind does not yield a high level of realism, consistency between objects is restricted to geometric aspects. Nowadays, three illumination techniques can be identified that attempt to raise the quality of AR and in general MR: *common illumination*, *relighting* and *inverse illumination for relighting or common illumination*. These techniques vary in the quality of the illumination and in the consistency obtained between the illumination of the real and virtual objects.

The most straightforward method of these three results in the addition of shadows in the MR environments. Generating shadows is just as important as taking into account occlusions, since they help situating the objects in the scene and give information about the distance between different objects [8]. A higher level of realism can also be obtained when the local light interaction between real and virtual objects is incorporated in the MR scene. Simulating such effects results in *common illumination*. An example of an application that uses common illumination to improve the MR can be found in the movie industry. Special effects in movies make an effort to mix lighting effects and reflections as realistic as possible, resulting in brilliant graphical effects in recent movies such as *Jurassic Park*, *Harry Potter* and *The Lord of the Rings* trilogy. In these movies, computer-generated effects are blended entirely with the real footage; usually this is carried out by hand.

Some methods allow to change the original illumination, hereby influencing the appearance of virtual and real objects. An example of an application domain for this method is architecture. Being able to virtually change the lighting conditions in the real scene makes it possible to see the impact of a new building in a street under different lighting conditions without the need of recording the real environment under all these different conditions. Another application area is crime investigation [9]: a recording of a scene at a certain time can be changed to the illumination at a different daytime, making it possible to visualize the perception of the criminal at the time of the crime. Techniques that focus on virtually changing the illumination of an existing scene are simply known as *relighting techniques*.

The techniques brought together in a third category are based on more complex estimations of the reflectances in the environment in order to provide more accurate results. The process of estimating the reflectances (bidirectional reflectance distribution function or BRDFs) from an existing lighting system is called *inverse illumination*. It was originally developed to give more realism in computer graphics. Reflectance properties of objects were estimated in order to reproduce a more realistic simulation of virtual scenes. In the context of MR, inverse illumination techniques aim at making a correct estimate of the photometric properties of the objects in the scene. While other techniques search for acceptable solutions for the new illumination problem, inverse illumination makes it possible to produce relit images that aim to be an *exact* replica of the real conditions. A full discussion of the current state of the art of inverse illumination techniques can be found in [10], while Ramamoorthi and Marschner [11] present a tutorial on some of the leading research work in this area.

At the moment, fairly good techniques exist that can relight an augmented scene with a different illumination. It is getting more difficult to differentiate between virtual objects and real objects. The main limitation of most techniques is the tedious pre-processing time and the slow update rate, which excludes real-time applications. When a geometric model of the scene is required, the user will have to create one, usually in a semi-manual and error-prone manner. The scene update rate is often too slow to allow real-time user interaction, even with the current progress in computer hardware and software. The research focus is moving toward using hardware for the calculation instead of software to accelerate computation. Early results are promising, but more research needs to be carried out in this area. This paper does not review all existing work. Instead, it concentrates on illumination techniques in MR that are meant for large environments. When optimized and extended, these techniques will be widely applicable in real-time applications, for instance see-through display in AR. Several techniques exist for relighting human faces [12,13], or that focus on local objects or simple scenes [14,15]. These techniques are classified mainly in the domain of inverse illumination as the emphasis was placed on this aspect in the referenced papers. Although these techniques are designed for small objects they can be used to build extremely useful and strong methods for illumination in MR but they will not be further discussed in this paper.

This paper discusses the state of the art of those techniques that strive to solve the problem of illumination in MR environments and gives an objective evaluation of their quality. Section 2 describes in more detail the context of this review and the assessment of the criteria on which the classification is based. Section 3 gives a structured overview of all the illumination methods that were developed for MR. A further qualitative evaluation and discussion of the different techniques is given in Section 4. Finally Section 5 gives conclusions and presents the necessary future work for illumination methods for MR.

2. Problem Assessment

2.1. Objective and difficulties

The classes described above are not necessarily designed to lure the users into believing that what they see is real. For instance, VR often aims at trying to create the perception of a real world, without necessarily using convincing real imagery. Some AR systems merely add data displays to real scenes, making no attempt to mix the two seamlessly. This paper considers MR scenes that *do* try to convince the users of *believing* that a real world is surrounding them and will use this as a measure to assess the quality of the method.

An MR is convincingly real when it is impossible to separate the virtual elements from the real elements in the resulting environment. Four critical success factors were identified that need to be present in the MR in order to be convincingly real:

- **After including the virtual object(s), the resulting scene needs to have a consistent shadow configuration** [8]. The main difficulty to obeye this requirement is to find the correct appearance of the new shadows: their position in the scene, shape and colour. Sometimes these are estimated, but they can be calculated exactly if the geometry of the scene, the illumination characteristics and the material properties of all objects in the scene are known.
- **The virtual object(s) must look natural.** A cartoon-like virtual object is easily detectable and therefore efforts have been made to model objects that look realistic. One successful technique is *image-based modeling*, in which objects are rendered with textures based on real images.
- **The illumination of the virtual object(s) needs to resemble the illumination of the real objects.** There are two possible methodologies to achieve this requirement. Either the illumination pattern of the real scene is known, which in turn is used to illuminate the virtual object or all material properties of all objects in the scene are known or estimated, which allows the entire scene to be relighted with a consistent known illumination pattern.
- **If the user can interact with the MR environment, it is clearly important that all update computations occur in real time.** Any delay in the interaction will remind the user of the fact that what is seen is unreal [16]. The requirement of a real-time system is one of the most difficult to achieve, especially when no pre-processing time is allowed.

2.2. Assessment of existing techniques

The ultimate objective of the aforementioned techniques is defined by the amount of realism perceived by the user. This inherent subjectivity complicates an objective assessment of the various techniques. In this section, a few quality criteria are listed that will be used in Section 4 to assess the presented methods:

2.2.1. *Amount of realism*

In some cases, it is impossible to evaluate the amount of realism without using a statistical measure. For instance, a test audience can evaluate the technique, if the test group is large enough, a statistical value can be derived from the group evaluation. Alternatively, if the inserted virtual object is an exact replica of an existing real object, it is possible to give an exact value of the amount of realism in the produced scene. It suffices to compare the generated scene with an image of the real object in the same scene. The difference between the two gives a measure of the level of realism.

2.2.2. *Input requirements*

It is expected that the more input data are available, the higher the quality of the end result will be. On the other hand, the usability of the system reduces with the complexity of the input data. Possible input data are: the geometry, the light position, the illumination pattern and the material properties. This report gives a classification of the various techniques based on their input requirement.

2.2.3. *Processing time*

The time needed to create the end result is another important characteristic of the method. To offer the user a highly

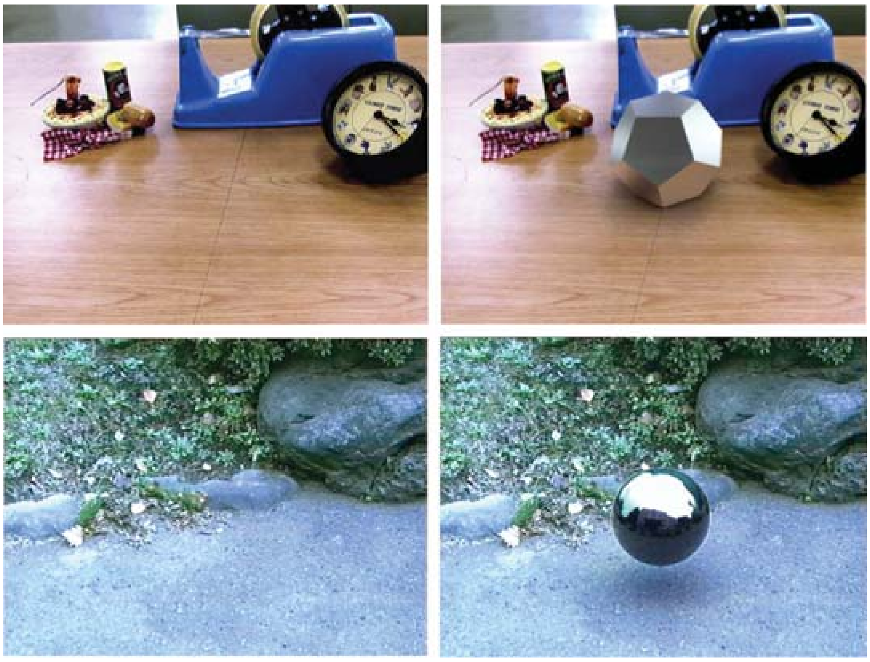

Figure 2: *Results for Sato et al. [17]. The top row shows results for an indoor scene, the bottom row for an outdoor scene. The images on the left are the input images, the images on the right illustrate the resulting MR. Soft shadows are produced using local common illumination. (Courtesy of Sato et al.)*

realistic interactive environment, the computations need to be done in real time. Unfortunately, this is very hard to achieve. If geometric and material properties of a scene need to be known, it is unavoidable that some pre-processing time needs to be incorporated. In general, the usability of the proposed techniques depends on the amount of pre-processing time needed and the computation speed of the illumination technique.

2.2.4. *Level of automation*

If the technique under consideration requires a considerable amount of manual interaction while processing the input data, the technique is less interesting than one that is automated.

2.2.5. *Level of interaction*

A technique can be judged based on its dynamic character: the possibility to change the viewpoint of the camera, or the possibility to let the user interact with the environment for instance by moving around objects. A higher degree of interaction gives a greater usability of the method.

2.3. Methodology

The various existing techniques can be grouped into three different classes, based on the methodology used to solve the problem. They were already listed in the introduction and are further discussed in this section:

2.3.1. *Common illumination*

To this category belong all methods that provide a certain level of illumination blending, like the addition of shadows projected from real objects on virtual objects and shadows cast by virtual objects on real objects. These techniques do not allow any modification of the current illumination of the scene. Two different types of common illumination can be considered: *local* and *global* common illumination, referring to the type of illumination simulated. For local common illumination, there is usually no requirement of any BRDF information. For global illumination, it is often important to have an estimate of the material properties of the real objects. The accuracy of this type of techniques depends on the accuracy of the known geometric model of the real scene. In Figure 2, an example is given of a rendering using global common illumination [17].

2.3.2. *Relighting after light removal*

Relighting techniques make it possible to change the illumination of the scene in two steps. Firstly, the current illumination effects of the real scene are analysed and possibly removed. Secondly, new illumination effects (shadows, intensity changes, addition of a new light, indirect lighting effects,

Figure 3: *Results for Loscos et al. [18]. The image on the left-hand side represents the real scene. The image on the right-hand side shows the relighted synthetic scene for which real light sources have been virtually turned off and a virtual light source is inserted. Global common illumination updates are performed at interactive rates using an adapted version of [21]. (Courtesy of Loscos et al.)*

Figure 4: *Results for Yu et al. [20]. Left: the original input scene. Right: the result of illuminating the original scene using a different illumination pattern. The specular and diffuse parameters of the real objects are calculated. (Courtesy of Yu et al.)*

etc.) are generated based on a new illumination pattern. These methods do not necessarily require an exact knowledge of the BRDF values of the real scene objects as for some methods the focus lies on generating a scene that *looks* realistic. These techniques require in general a detailed geometric model of the real scene. An example of a relighted scene using global illumination techniques [18] is given in Figure 3.

2.3.3. *Physically-based illumination*

This last category encloses those methods that make an attempt to retrieve the photometric properties of all objects in the scene often referred to by the term *inverse illumination*. These methods estimate BRDF values as correctly as possible as well as the emittance and positions of the light sources in the real scene. The BRDF values can be estimated using a goniometer [19] or can be calculated based on the photometric equilibrium equations [12,15]. The BRDF information can be used for both common illumination or relighting methods. The accurate BRDF estimation often permits to perform a complete and realistic relighting, which takes both reflections and global illumination techniques into account. Patow and Pueyo [10] give an in-depth overview of inverse illumination techniques. An example of inverse global illumination [20] is illustrated in Figure 4.

3. Classification of Illumination Methods for Mixed Reality

MR brings together those applications that create a new environment, around or in front of a user, containing both real and virtual elements. Sections 2.1 and 2.2 formulated the objectives, the difficulties encountered and the assessment criteria of the MR systems. One of these criteria, the type of input requirements, regulates the accessibility and accuracy of the technique. This criterion will be used to classify the different methods. An overview of the input information needed to calculate a common illumination, relighting or inverse illumination solution is given in Figure 5.

The classification put forward firstly takes into account the required geometric model of the real scene, starting with the techniques that require no geometric model and ending with techniques that require a precise geometric model. In this paper, a geometric model is defined as a reconstruction of a part of the (or the entire) real scene with significant detail. The pre-processing workload for techniques that extract a basic geometric model, e.g. the depth at a low resolution, is significantly lower than those methods that do require a high-level geometric model. Therefore techniques using basic geometric information are classified under that group of

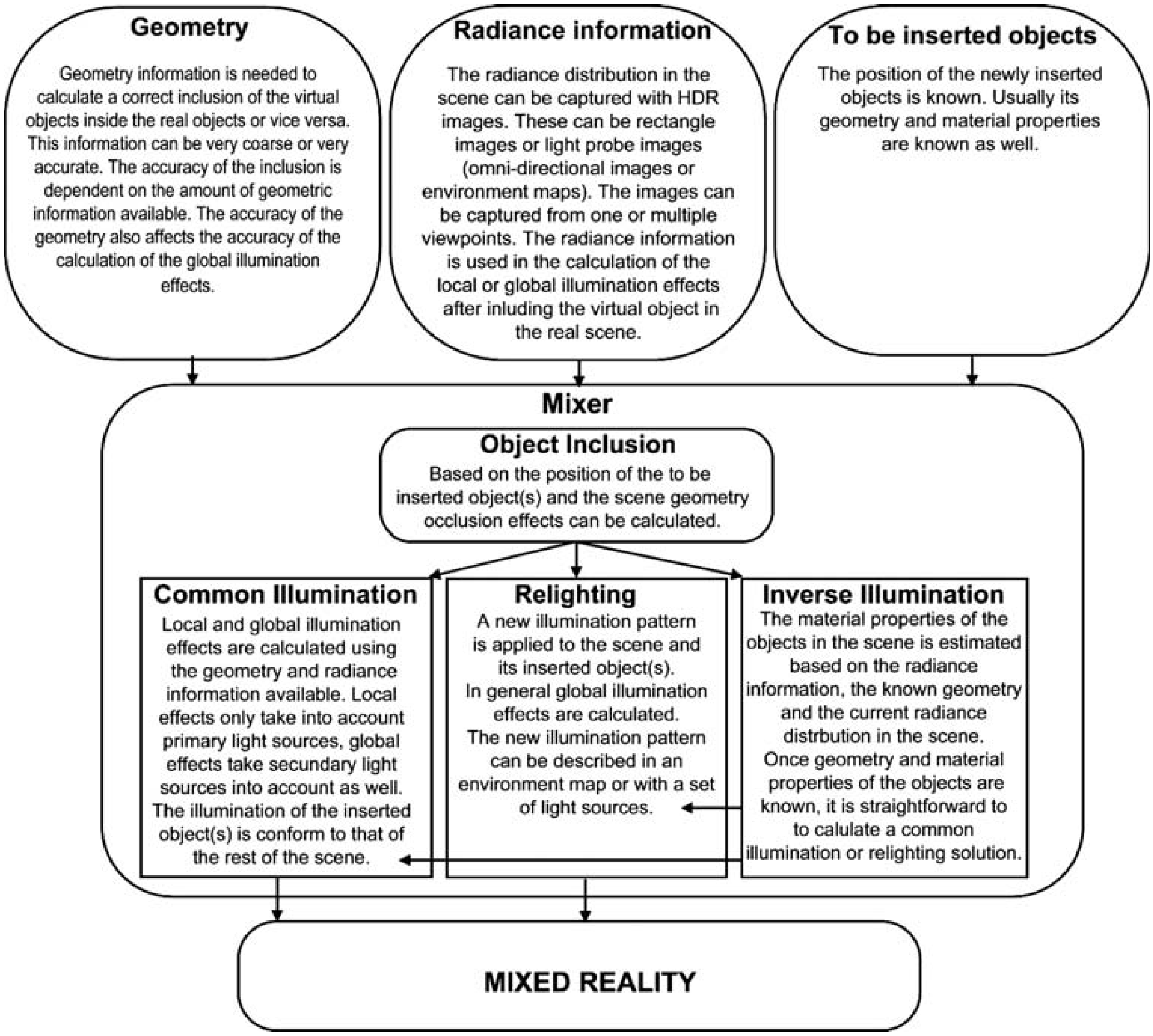

Figure 5: *An overview of the dataflow in illumination calculation for MR. Three input sources are considered: the scene geometry, the scene radiance and information about the to be inserted object(s). These are used to calculate the illumination solution. The accuracy of the integration depends on the amount of input information available.*

methods that do not require a geometric model, as this will give a better indication of the amount of pre-processing time required for each different class.

Two different approaches exist to reconstruct a geometric model of the real scene. Either the scene is scanned with a scanning device [22–24] or it is reconstructed using stereovision [25–28]. The first option of using a scanning device gives a precise geometric model but is expensive and tedious. Often the model will capture too much detail, which is not always necessary and is difficult to manage for real-time applications. Objects such as trees and objects with a highly specular surface are for some scanning techniques difficult to model accurately. Instead of using a scanning device, modeling techniques based on stereovision can be used to reconstruct the geometry of a scene. Most methods described in this survey requiring a 3D model of the scene make use of this low cost solution. In general, the 3D reconstruction requires at least two images from two different viewpoints. However, the entire geometry of a scene cannot be captured with one image pair only, this would create gaps in the known geometry. Usually more than one image pair is required for a complete geometric reconstruction. The ease at which this reconstruction can take place depends on the amount of information that is available for the camera(s) used. If the internal and external camera parameters are known, the reconstruction is easier. Internal parameters can be estimated in a relatively simple way. Recording the external parameters is more difficult and involves a precise and tedious capture process. Fortunately, the fundamental matrix of the stereovision system can be estimated based on the internal parameters only, if at least eight corresponding points are known [25,26]. This was reduced to six known points [27,28] for a calibration relative to a scale factor, which is often sufficient. Having the fundamental matrix can ease the reconstruction but does not make it trivial. Research led to different types of systems: nonconstraint systems [15, 29] and constraint systems [30–33]. Good commercial reconstruction software [34–37] exists, but most of them lack the option of reconstructing complex shapes and large environments. In general, we can conclude that systems requiring geometric information demand a long pre-processing time, and are not guaranteed to get an accurate geometric model. It is really important to recognize the geometric acquisition as a difficult problem, that still requires much more research efforts.

Our classification of methods is, parallel with the classification based on the geometric information, also based on this amount of image data needed to reconstruct a MR environment. Hereby excluding the image data needed for retrieving geometric information. More precisely, the classification will use the amount of different input images used for the rendering or calculation process. These processes might use more than one image (or texture containing radiance information) for the same point in the real scene, for instance for the BRDF estimation.

Some projects adopted the concepts of high dynamic range images (HDR images) [38] which can be computed using techniques such as [39,40]. Each HDR image is generated based on a set of images taken from the same viewpoint of the same scene, but with a different exposure. The end result is one image containing radiance values instead of ordinary RGB values. In other words, radiance values are not clamped in the RGB space. It may be argued that methods using HDR images should be classified under that class with methods that use more than one image for each point of the scene. However, this paper considers that a HDR image provides one radiance value per point, and methods that use only one HDR image for a certain point of the scene, are therefore classified as methods requiring only one input image. Similarly, techniques that require a few or many HDR images are classified as methods using respectively a few or many images.

The following classification is used throughout the rest of this section:

1. **Model of the real scene unknown, one image known** (Section 3.1): this category lists those techniques that do not require any model of the real scene, except for some low-level geometry like depth information. Any necessary radiance information of a certain point in the real scene is extracted from one single image.
2. **Model of the real scene known, one image known** (Section 3.2): a geometric model of the real scene is available. Any necessary radiance information is extracted from one image only.
3. **Model of the real scene known, few images known** (Section 3.3): again a geometric model of the scene is required. For a certain point in the scene, radiance information is available from a few different images.
4. **Model of the real scene known, many images known** (Section 3.4): this class groups those techniques that require both a detailed geometric model of the real scene and radiance information from a large set of different images.

The rest of this section lists the most significant methods based on the above mentioned classification and briefly discusses their methodology. A discussion of the techniques based on the assessment criteria mentioned in Section 2.2 is given in Section 4.

3.1. Model of real scene unknown, one image known

To this challenging category, in terms of output quality, belong those methods that require very little relevant information about the real scene. Since no geometric model of the scene is available it might be necessary to calculate depth information of the scene, to allow a correct inclusion of the virtual objects, or some lighting information. For this group, all radiance information is extracted from one single image.

Nakamae *et al.* [41] were the first to propose a method for composing photographs with virtual elements. Input photographs are calibrated and a very simple geometric model of the real scene is extracted. The viewpoints of the photograph and the virtual scene are aligned to ensure an appropriate registration of the virtual objects within the photographed elements. The sun is positioned within the system according to the time and date when the picture was taken. The sun intensity and an ambient term are estimated from two polygons in the image. The illumination on the virtual elements is estimated and adjusted to satisfy the illumination in the original photograph. The composition is done pixel by pixel and at that stage it is possible to add fog. All parameters are very inaccurate and therefore the results are limited in accuracy. However, they were one of the first to mention the importance of using a radiometric model to improve the image composition. Figure 6 gives an example of the obtained results.

Techniques exist in computer graphics that use *environment maps* to render objects in a scene. They were introduced to approximate reflections for interactive rendering [42–44]. These techniques can also be used to assist the rendering of glossy reflections [45–47] by pre-filtering a map with a fixed reflection model or a BRDF. At this moment, graphics cards

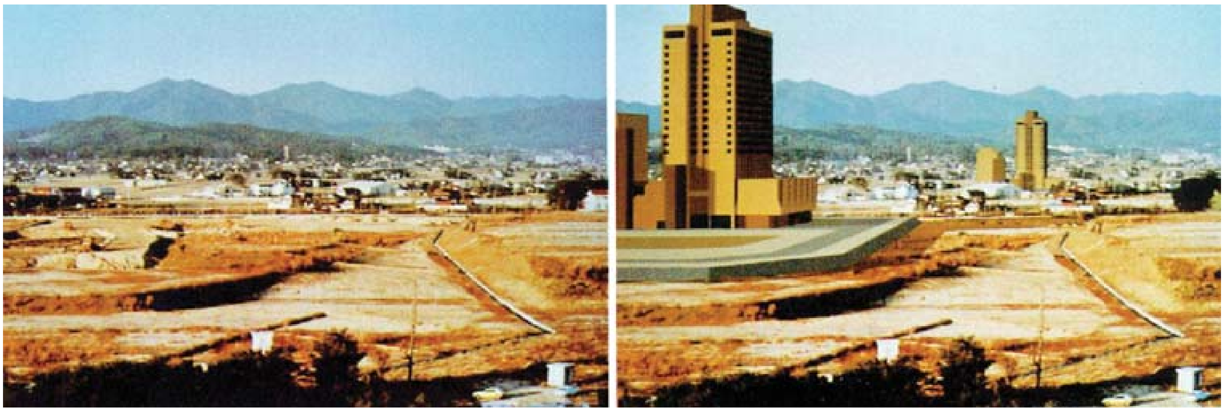

Figure 6: *Results for Nakamae et al. [40]. The first image depicts the background scene, the second image shows the augmented scene where the illumination of the augmented objects are matched to their surroundings and the shadows are cast accordingly. (Courtesy of Nakamae et al.)*

extensions support the real-time use of environment maps, this encourages its use even more. Graphics cards now support cube maps [48]; ATI [49] presented at SIGGRAPH 2003 a demonstration of a real-time application for high resolution. Environment maps can be used to represent the real scene in a MR environment as a panorama and the information from these images can be used to simulate reflections on a vertical object positioned at the center of the environment map [50].

Agusanto *et al.* [14] exploited the idea of environment maps to provide reflections in AR. They use HDR images of the environment captured by a light probe to create the environment map. These maps are filtered off-line to decompose the diffuse from the glossy components. The rendering is then performed with a multipass rendering algorithm that exploits hardware capabilities. After some pre-processing, like the inclusion of shadows, they present results for MR environments rendered on a desktop. An impressive aspect of their work is that the method also works for real-time AR. The implementation of their method works with ARToolKit [51] and the results show interactive reflections from the real scene on virtual objects at interactive frame rate. An example of such a projection is given in Figure 7. Although it should be feasible, they have not yet provided a shadow algorithm for the AR application.

Figure 7: *Results for Agusanto et al. [14]. The virtual objects are rendered with skin textures. The left object is blended with a diffuse map and no soft shadows. The right objects is blended with a glossy map and with soft shadows. (Courtesy of Agusanto et al.)*

Sato *et al.* [17] adopt a technique that extends the use of environment maps to perform common illumination. In their method, it is assumed that no geometry is known *a priori*. However, at least a few images are known from different but very restricted and known viewpoints, which can be used to estimate a very simple geometry of the scene and the position of the light sources. The obtained geometry does not offer a reliable occlusion detection and the positions of the virtual object are therefore restricted to lie in front of all real objects in the real scene. After this low-level geometric reconstruction, a set of omnidirectional images are captured with varying shutter speed. From these images, a radiance distribution is calculated, which in turn is mapped onto the geometry. To calculate the shadows and the local illumination a ray casting technique is adopted. The radiance values of the virtual objects are calculated using the information known about the light sources, the radiance values of the real scene, the geometry and the BRDF values of the virtual objects. To simulate the shadows cast by virtual objects on real objects, the radiance values of those points in the scene that lie in shadow are scaled. The simulated soft shadows look realistic, see Figure 2. Their geometric estimate is poor and therefore usability of the method is limited and the positions of the virtual objects are restricted. Nevertheless, the method produces convincing local common illumination.

3.2. Model of real scene known, one image known

Most of the existing illumination methods assume that a geometric model of the scene is available. The more detailed the geometric model is, the more reliable the occlusion detection will be. Although not all techniques explain where this model should come from, it is doubtful that a perfect geometric model can ever be acquired and this should be taken into account when evaluating a specific method. In this section, a discussion is given of those methods that take a certain 3D geometric model of the real scene as input and extract radiance information from one single image. All methods that belong to this category are further divided into three groups based on the type of illumination they produce:

- local illumination for AR applications: Section 3.2.2
- common illumination: Section 3.2.1
- relighting: Section 3.2.3

3.2.1. *Local illumination for AR*

As mentioned before, AR has long been an area wherein people focused on registration and calibration as these are still difficult problems to solve in that area. However, a few papers tried to introduce shadows in their systems, to show how well the registration was done and to improve the rendering quality. Recent improvements in graphics hardware for rendering shadows made it possible to perform real-time rendering of shadows on well-registered systems where the geometry is known. Early work was presented by State *et al.* [52] in which virtual objects are inserted in the see-through real scene. A real light source is moved around and tracked, and shadows of the virtual object due to this real light source are virtually cast onto real objects by using the shadow mapping technique [53]. In this case, the light source is assumed to be a point light source. It was very promising that some researchers in AR were interested in using local common illumination in their systems, but it was followed by a long period in which no innovative material emerged. Only recently, additional work of Haller *et al.* [54] and Jacobs *et al.* [55] was carried out to add shadows from a virtual object onto real objects. Both methods use shadow volumes, and in order to get good quality results knowledge about the scene geometry is essential. However, Jacobs *et al.* start from a rough approximation of the geometry and use image processing techniques to compensate for approximation errors. Other methods exist [7] that we will not discuss here, since they will not be applicable in general MR systems because these systems would require the capture of a large environment.

3.2.2. *Common illumination*

Jancene *et al.* [56] use a different approach to illuminate the virtual objects, they base their method, called RES (reality enriched by synthesis), on the principle of composition. The objective is to add virtual animated objects in a calibrated video sequence. The final video is a composition of the original video sequence with a virtual video sequence that contains both virtual objects and a representation of the real objects. The geometry of the real object is reconstructed *a priori* so that for each frame in the video the geometry is known. The rendering in the virtual sequence is performed using ray tracing. It is possible to modify the reflectance properties of real objects. Shadows are simulated in the virtual sequence, the impact of the shadows in the final video is acquired by modifying the original video with an attenuation factor. An occlusion mask is created to reflect occlusion between virtual and real objects. This method came quite early in the history of common illumination and video composition. Even though it is not applicable for real-time applications, it allows local common illumination and virtual modification of the reflectance properties of real objects. The images on the left in Figure 8 illustrate the original scene, the images on the right illustrate the composition.

Gibson and Murta [57] present a common illumination method, using images taken from one viewpoint that succeeds in producing MR images at interactive rates, by using hardware accelerated rendering techniques. Apart from constructing the geometry of the scene, the pre-processing involves creating a set of radiance maps based on an omni directional HDR image of the entire scene. New virtual objects are rendered via a spherical mapping algorithm, that maps the combination of these radiance maps onto the virtual object under consideration. Later shadows are added using a two-step algorithm. To simulate the shadows, a set of M-light sources are identified, which imitate the true, unknown illumination in the scene. Each light source is assigned a position and two parameters α_i and I_i, which define the colour of the shadow. For each light source, a shadow map is calculated using efficient hardware calculations (z-buffer). Shadow mapping is an intensive technique supported by the graphics hardware that helps to create shadows in a fast and efficient way. The shadows created with shadow maps are in nature hard shadows and therefore unsuitable for realistic shadow generation. Gibson and Murta combine the M-shadow maps in a specific way, using the above-mentioned parameters and now the system succeeds in simulating soft shadows, looking almost identical to the solutions obtained with a more computational and traditional ray-casting algorithm, see Figure 9. The system of M-light sources needs to be defined so that it represents a close replica to the current illumination system, an increase in number of light sources affects the rendering time. To demonstrate their method, Gibson and Murta used eight light sources to simulate an indoor environment. The position and the parameters of the light sources are defined via an optimization algorithm, which needs to be executed only once for each different scene.

Debevec [59] presents a more advanced common illumination technique that estimates the BRDF values for a small part of the scene. It is argued that if a virtual object is inserted

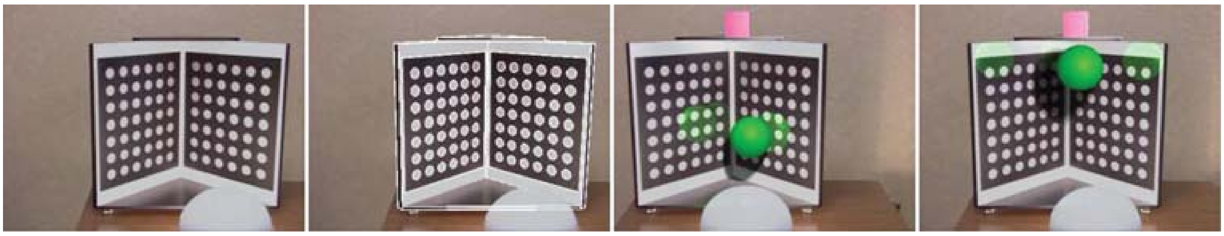

Figure 8: *Results for Jancene et al. [56]. The images on the left-hand side show the original scene and the registration of the cardboard box within this scene. The images on the right-hand side show two screen shots from the video sequence in which a virtual dynamic green ball and static pink cube have been added to the original scene. The reflection of the green ball is visible on the board behind it. (Courtesy of Jancene et al.)*

Figure 9: *Results for Gibson and Murta [57]. Comparison of a ray-traced (left) and a hardware generated image (right). The ray-traced image was generated using* RADIANCE *[58], the hardware-generated image made use of the rendering method described in [57]. The generation of the ray-traced image took approximately 2 hours, the generation of the hardware-rendered image took place at nearly 10 frames-per-second. (Courtesy of Gibson and Murta)*

into the scene, only a small fraction of the scene experiences an influence from that inclusion. Relighting techniques using inverse illumination therefore only require the BRDF values of those points that lie in this fraction. For most applications it is possible to know the position of the virtual objects and Debevec uses this position to divide the entire scene into two parts: the *local scene* and the *distant scene*. The local scene is that fraction of the scene whose appearance might alter after inclusion and the BRDF of the materials in that part need to be estimated. On the other hand, the distant scene is that part of the scene that undergoes no physical alteration after inclusion. A schematic overview of the division in local and distant scene and their corresponding influences is presented in Figure 10. The local scene is restricted to be diffuse only; the distant scene has no restrictions. An omni directional HDR image is captured using a mirrored ball. The resulting *light probe image* is used to present the illumination in the real scene. Based on the geometric model, the light probe image and the division into local and distant scene, the BRDF values in the local scene are estimated. The calculations are straightforward, since only diffuse BRDF values are considered. A *differential rendering* technique was developed to reduce the possible inconsistencies in the geometric model and the (specular) error on the BRDF estimates to an acceptable level. The rendering is a two-pass mechanism. First, the augmented scene is rendered using a global illumination technique, the result is denoted by LS_{obj}. Next the scene is rendered using the same global illumination technique, without including the virtual objects, denoted by LS_{noobj}. If the input scene is represented by LS_b, than the difference between LS_b and LS_{noobj} is exactly the error that results from an incorrect BRDF estimation. The differential rendering therefore calculates the final output rendering LS_{final} as:

$$LS_{final} = LS_b + (LS_{obj} - LS_{noobj})$$

This differential rendering technique removes most of the inaccuracies and in a certain way it is similar to the one of Jancene *et al.* [56] presented above. The results of this technique are promising, see Figure 10, but it still suffers

Figure 10: *Debevec [59]. Left: a diagram illustrating the relation between the different components presented in [59]. The real scene is divided into a local scene and a distant scene. The illumination from the distant scene influences the local scene and the virtual objects. The virtual objects influence the local scene. The local scene and the virtual objects do not have an influence on the distant scene. Middle: an image of the real scene. Right: an example of the differential rendering technique for an indoor scene after inserting virtual objects. Diffuse effects are simulated. (Courtesy of Debevec.)*

from a few deficiencies. Firstly, only diffuse parameters of the local scene are estimated, this introduces an error that should be compensated by the differential rendering. Secondly, the viewpoint can be altered but the technique is too slow to work at interactive rates. If the rendering could be accelerated using low cost graphics hardware, it could be possible to achieve interactive update rates for the MR.

Gibson *et al.* [60] developed a method to create soft shadows using a set of shadow maps. They created a rapid shadow generation algorithm to calculate and visualize the shadows in a scene after the material properties of the scene are calculated. A proper estimate of both the geometry and the radiance information in the real scene needs to be available. It is assumed that the BRDF for all materials is diffuse. This diffuse BRDF is estimated using geometry and radiance information (one radiance image per 3D point). In their method, the scene is divided into two parts: one part contains all patches in a scene that are visible from the camera, called the *receiver patches* and another part contains those patches in the scene that have a significant radiance, called the *source patches*. Then they organize these patches to build a shaft hierarchy between the receiver patches and the source patches. The shaft hierarchy contains information on which patches block receiver patches from other source patches. Next they render the scene from a certain viewpoint. This rendering is a two-pass mechanism. In a first pass, they go through the shaft hierarchy to see which source patches partially or completely illuminate a receiver patch. Once these source patches are identified, they set the radiance of each receiver patch to the sum of all irradiance coming from these source patches, without taking occlusions into account. The second rendering pass, takes the shadows in consideration. To calculate the portion of blocked light, they use the shadow mapping technique. In fact, they create a shadow map for each source patch. At each receiver patch, these maps are then combined and subtracted from the radiance value that was rendered in the first pass. This technique is capable of producing soft shadows in a fast and efficient way. In Figure 11, examples are given of synthetic scenes rendered using the above-described method. Renderings of the same synthetic scenes using a ray tracing method are given in the middle column. The images in the last column are photographic reference images. Another set of methods were built to exploit the structure of a radiosity method. Fournier *et al.* made pioneering work in this direction [61]. When this method was developed, facilities for modeling a geometric model from a real scene were not available. To overcome this issue, Fournier *et al.* decided to replace the geometry of the objects in the real scene by their bounding box, and an image of the object was applied on each of the faces of the box. An example of such a model is shown in Figure 12. To set up the scene for global common illumination computation, faces of the boxes representing the real objects are divided into patches. Using the information contained in the radiance textures, a diffuse local reflectance is computed by averaging pixels covered by each patch. Light source exitances are estimated and the radiosity of the patches are set as an average of the per pixel radiance covered by each patch. After insertion of the virtual objects and the virtual light sources in the model of the real scene, new radiosity values are computed for the elements in the scene using *progressive radiosity* [62]. The rendering is carried out by modifying the intensity of each patch with the ratio obtained by dividing the new radiosity by the original one. In Figure 12, an illustration of the result of this method is given. The results of this technique look promising but it suffers from the lack of a detailed geometry. This leads to misaligned shadows and other types of mismatching between real and virtual objects. The technique is slow and will not allow real-time interaction. Nevertheless, this pioneering method has influenced subsequent research work, e.g. Drettakis *et al.* [63] and Loscos *et al.* [18] as presented in the remainder of this section.

Drettakis *et al.* [63] present a method that builds on Fournier *et al.* [61], but uses a finer model of the real scene. The same equations are used to estimate the light sources emittance, the reflectance of the patches and the original

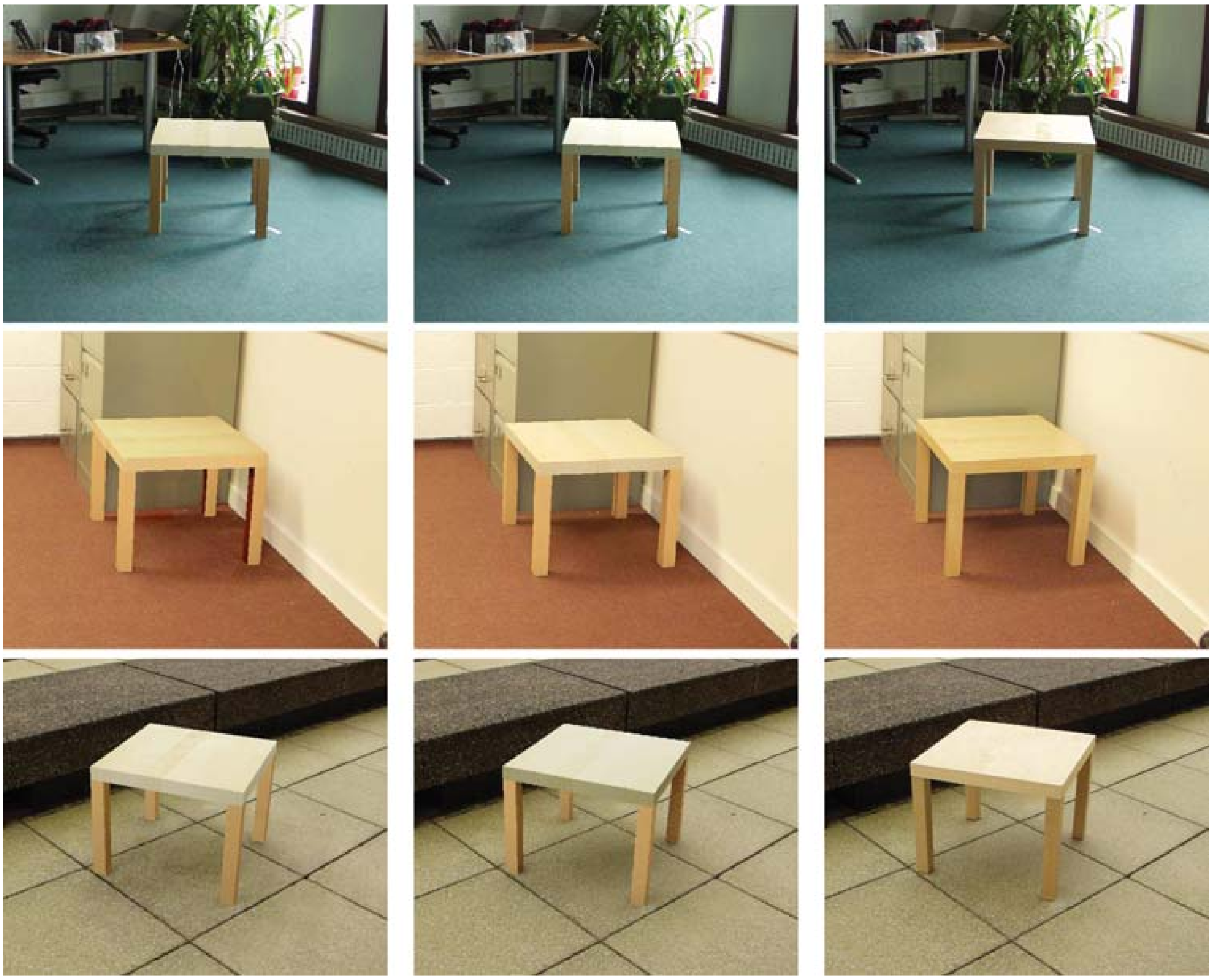

Figure 11: *Results for Gibson et al. [60]. A comparison of the rendering quality for three different scenes. The images in the left column are produced using the system presented in [60]. The images in the middle column are rendered using ray tracing. The images in the right column are photographic reference images. (Courtesy of Gibson et al.)*

radiosity. Drettakis *et al.* make use of the more recent hierarchical radiosity method hierarchical [64] accelerated by using clustering [65–67]. Based on [21] a hierarchy of shafts is built from the real scene model, which allows a local understanding when virtual objects are added. This permits an easy identification of all patches that need to undergo a radiosity alteration due to the influence of the newly added object. The advantage of this shaft hierarchy is that it permits interactive updates of the illumination in the augmented scene when virtual objects move. The final display is made similarly to the method of Fournier *et al.* [61]: the intensity of the patches is modified with the ratio defined by the modified radiosity divided by the original radiosity. This type of rendering is fast, compared to a ray tracing method, as it uses the hardware capability to render textured polygons. This method provides global common illumination with possible interaction. Unfortunately, the technique does not allow changing either the current illumination or the current viewpoint. In Figure 13, a screen shot is given of the 3D reconstruction and an example of the resulting MR using the above-explained method.

3.2.3. *Relighting*

In Loscos *et al.* [18], relighting is made possible, while keeping the framework set by Fournier *et al.* [61] and Drettakis *et al.* [63]. The scene parameters are extracted in the same way but all calculations are extended to the use of HDR images [68] as well. Since this technique focuses on relighting, a specific subdivision of the real scene is made to detect as much direct shadows as possible. The radiosity of each element is modified to simulate nonblocked radiosity, in other words, to erase the shadows from the textures. A factor is computed

Figure 12: *Results for Fournier et al. [61]. Left: wire-frame image, all objects in the scene are represented by a box, that narrowly fits the object. Middle: Image information is mapped on the boxes (note that for the ball, a more complex shape was used). Right: an example of an MR, the book on top of another book, lying on a table is virtual. Also a virtual light source is added. The global common illumination effects are generated with an adaptive progressive radiosity algorithm. (Courtesy of Fournier et al.)*

Figure 13: *Results for Drettakis et al. [63]. In the left image a screen shot is given of the 3D reconstruction of the real scene. The right image gives an example of the MR, the floating box is the virtual object. The virtual objects can be moved at interactive rate while keeping the global illumination effects. This is carried out by using an adaptation of hierarchical shafts for hierarchical radiosity [21]. (Courtesy of Drettakis et al.)*

using the radiosity method without taking the visibility in consideration. Then the new radiosity value is used to update the texture. Approximations of the estimation and of the input data lead to an inexact modification of the texture. In a second step, another factor is applied to automatically correct the imprecisions. This is done by using a reference patch that reflects the desired result. Once this is done, the new textures are used instead of the original ones, and reflectance and original radiosity values are updated accordingly. Shadows can be simulated using the factor of the newly computed radiosity solution divided by the original radiosity (without shadows). This technique also extends the method presented in [21] for the insertion of virtual lights. In the system of Loscos *et al.* [18], it is possible to virtually modify the intensity of real light sources, to insert virtual objects that can be dynamically moved and to insert virtual light sources. The problem that comes with inserting new lights or increasing light source intensity is that the value of the factor computed between the new radiosity value, divided by the original radiosity, may be greater than one. In that case, multipass rendering is used to enable the visualization of brighter illumination. This method allows interactivity and is fairly rapid in the pre-processing computation. However, the obtained results are inaccurate as the illumination of the real scene is not fully estimated. Firstly, because lit areas are not altered at all, and secondly, because it concentrates on the diffuse component only. An example of the results is shown in Figure 3 using HDR images as an input.

Although it does not seem feasible to estimate specular components of the BRDF from one single image, Boivin and Gagalowicz [69] present a technique that re-renders diffuse and specular effects based on radiance information from one single image and a full geometric model of the scene, including the light source positioning and the camera properties. With a hierarchical and iterative technique they estimate the reflectance parameters in the scene. In this method, the reflectance model of Ward [19] is used, which presents the entire BRDF with either three (isotropic materials) or five (anisotropic materials) different parameters. The BRDF estimation process starts by assuming that the BRDF values are all diffuse. A synthetic scene is rendered using the geometry, the current BRDF estimate and global illumination techniques. If the difference between the real scene and the synthetic scene is too large, the BRDF values are re-estimated using a more complex BRDF model. First specular effects are added and a roughness factor is estimated using an time-consuming optimization process. Later anisotropic effects are introduced and the optimization continues until a reasonable synthetic scene is acquired. This is very similar to the way parameters are estimated in [20]. However, in this case, only one input image is used, and anisotropic parameters are estimated as well. The method of Boivin and Gagalowicz relies on one single image to capture all photometric information. The advantage of such an approach is that the image capturing is relatively easy; the disadvantage is that only partial geometric information is available: there is no information for those surfaces that are not visible in the image. Nevertheless, the proposed technique allows changing the viewpoint. If a sufficiently large portion of a certain object is visible in the image, the reflectance properties of the missing parts of the object are calculated based on this portion. Grouping objects with similar reflectance properties makes this process more robust. On the other hand, this requires that not only the geometry needs to be known, but also a partitioning of the scene into objects with similar reflectance properties, which greatly compromises the operatability of this technique. Although optimized, the rendering algorithm is computationally expensive and therefore only a nonreal-time solution can be obtained. In Figure 14, an illustration is given of the output results of the described method.

3.3. Model of real scene known, few images known

If more information about the radiance of the points in the scene is available, a better BRDF estimate can be acquired. The radiance perceived at a certain point depends on the viewing angle, on the angle of incident light and the BRDF. Hence, it is possible to gain more information about the BRDF of a certain point in the scene if radiance information is available from images captured from a different viewing angle. Alternatively, if the viewpoint is kept the same but the position of the light sources is changed, extra BRDF information is captured as well. In this section, the methods are discussed that make use of this extra information.

Loscos *et al.* [70] developed a system that allows relighting, as well as virtual light source insertion, dynamic virtual objects inclusion and real object removal. They identified that it is difficult to estimate reflectance values in shadow regions due to saturation and because this estimate depends on the quality of the indirect light estimation. This is compensated for by adding extra photographs captured under different lighting. The geometry of the real scene is modeled from photographs. This geometric model is textured using one of the images, taken from the different viewpoints. A set of pictures is then taken from this chosen viewpoint while a light source is moved around the scene to modify the illumination. These pictures can be HDR images as used in [68]. Loscos *et al.* decided to mix a ray-casting approach to compute the local illumination and a radiosity approach to compute the indirect lighting. Two sets of reflectances are thus computed. First diffuse reflectance values are computed for each pixel of the viewing window. This is done with a weighted average of the reflectance evaluated with each input image differently lit. The applied weight is based on whether the 3D point associated to the pixel is in shadow relative to the light source position, and also whether the radiance value captured is saturated. The reflectance values are then used to initialize a radiosity system similar to those in [63,18]. This reflectance can be refined by an iterative algorithm [70]. With this reflectance, Loscos *et al.* are able to relight the scene using global illumination. Pixel values are updated by adding the local illumination value, computed by ray casting, to the indirect illumination value, computed by hierarchical radiosity using a rough subdivision of the scene. Local modifications are made after the insertion or moving of virtual objects by selecting the area of the window where local illumination will be affected. Indirect illumination is modified by adapting the technique of [21]. Similarly, virtual light sources can be added, and intensity of real light sources can be modified. A very interesting application of this method is the removal of real objects. The unknown information previously masked by the object is filled using automatic texture synthesis of a sample of the image of the reflectance values of the previously hidden object. The results show that the relighting and the interaction with virtual objects can be achieved in an interactive time. Image examples of the results are shown in Figure 15. The produced results are good but could be improved by considering specular effects. Due to the nature of the image capture process, it would be very difficult to apply this technique on real outdoor scenes.

A different approach taken by Gibson *et al.* [71] results in another relighting method, in which the reflectance of the material is roughly estimated based on a restricted amount of geometry and radiance information of the scene. In theory, only geometry and radiance information is needed for those parts of the scene that will be visible in the final relighted MR.

Figure 14: *Results for Boivin and Gagalowicz [69]. The top left image illustrates the original scene. The top right image is a relighted synthetic image. Diffuse and specular effects are simulated using an optimization algorithm. The bottom left image illustrates the possibility of changing the viewpoint by grouping objects with similar properties. The bottom right image illustrates the relighting of the original scene with a different illumination pattern. (Courtesy of Boivin and Gagalowicz.)*

Figure 15: *Results for Loscos et al. [71]. The left image is one of the input images of the real scene. The middle image is a relighted image of the real scene, using the calculated BRDF values. The left image illustrates the removal of an object (the door), the insertion of a new virtual object (the chair) and the insertion of a virtual light source. All manipulations are carried out at interactive update rates. The illumination is updated locally with ray casting. The consistency of the indirect illumination is kept using an adaptation of [21]. (Courtesy of Loscos et al.)*

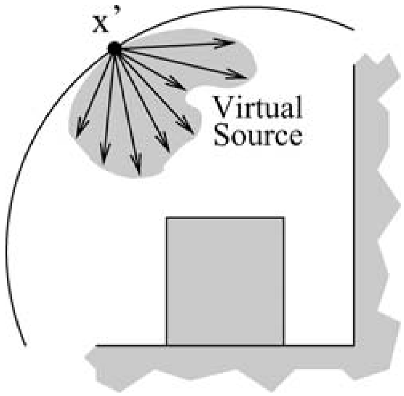

Figure 16: *Technique Gibson et al. [71]. The real illumination is approximated by a illumination surface. This illumination surface is covered by a set of virtual light sources. The parameters of these virtual light sources are estimated such that its effect resembles the real illumination. (Courtesy of Gibson et al.)*

In their approach a photometric reconstruction algorithm is put forward, that is capable of estimating reflectance and illumination for a scene if only incomplete information is available. To achieve this the direct illumination is modeled as coming from unknown light sources using virtual light sources, see Figure 16. The aim is not to produce an accurate illumination model, but rather a model that produces a similar illumination as in the original scene. The model used is a spherical illumination surface: a set of small area light sources that surrounds the known geometry. The parameters of this surface, the position and emission of the light sources, are estimated using an iterative minimization algorithm. Based on this model, the reflectance of the materials in the scene are estimated. The MR scene is rendered using a ray tracing algorithm. User interaction is impossible at real-time update rate but nevertheless the method illustrates the possibility of getting fairly realistic mixed realities, without limiting input requirements. This method is original, interesting and very practical to adapt to many situations where information on a real scene is partially known. Imprecisions and ambiguities are compensated for, resulting in a more accurate simulation of the existing illumination. An example of a relighted and augmented scene is given in Figure 17.

3.4. Model of real scene known, many images known

This category collects those techniques that require the most input information. Not only the geometry is known but also radiance information under many different geometric setups. Two significant methods could be identified that belong to this category of MR methods. They were selected from a broad set of techniques on inverse illumination because they provide a solution for a large group of objects, which is essential for MR. The first inverse illumination method [20] focuses on the BRDF estimation, using many HDR images from different viewpoints. The second [72] allows to relight outdoor scenes. The remainder of this section briefly discusses these two techniques.

Yu *et al.* [20] use a low parametric reflectance model, which allows the diffuse reflectance to vary arbitrarily across the surface while nondiffuse characteristics remain constant across a certain region. The input to their system is the geometry of the scene, a set of HDR images and the position of the direct light sources. An inverse radiosity method is applied to recover the diffuse albedo. The other two parameters in the reflectance model of Ward [19], the roughness and the specular component, are estimated through a nonlinear optimization. For the estimation of the specular BRDF, it is assumed that many HDR images are available from a different set of viewpoints. The estimation makes use of the position of the light sources and the possible highlights they may produce on a surface due to specular effects. It is therefore helpful to capture images of the scene with a various number of light sources, since this might increase the number of specular highlights. This precise estimate of the BRDF values in the scene allows to remove all illumination in the scene and a new illumination pattern can be applied. To render the scene they make use of Ward's radiance system [58]. No further steps were taken to speed up the rendering process. Figure 4 illustrates the results obtained for augmented images compared to photographs of the real scene. This technique is interesting for MR because it provides an algorithm to estimate an accurate complex BRDF of a complex real scene, resulting in an accurate representation of the illumination.

Yu and Malik [72] present a technique that allows relighting for outdoor scenes based on inverse illumination. As it is impossible to retrieve the geometry of the entire scene, they separate the scene into four parts: the local model, the sun, the sky and the surrounding environment. The illumination sources are the sun, the sky and the surrounding environment. Luminance due to the sun and the sky are estimated based on a set of input images. At least two photographs per surface of the local model are captured, which should show two different lighting conditions (directly and not directly lit by the sun). The local model is subdivided into small surfaces. Based on these two photographs, two pseudo-BRDF values are estimated per surface. One relates to the illumination from the sun, the other relates to the illumination from the integrated environment (sky plus surrounding environment). A least square solution is then used to approximate the *specular term* for each surface and for each lighting conditions (from the integrated environment and from the sun). This approach uses an approximation of the inverse illumination equation. It illustrates the difficulty of setting up a parameterized MR system for outdoor scenes. At rendering time, different positions of the sun are simulated. After extracting the sun and the local model from the background, sky regions are identified and they are mapped on a mesh supported by

Figure 17: *Results for Gibson et al. [71]. The left images illustrates the reconstructed scene from a novel viewpoint. The image in the middle is a synthetic image illuminated with virtual light sources. The right image illustrates the addition of virtual objects. Both specular and diffuse effects are simulated. (Courtesy of Gibson et al.)*

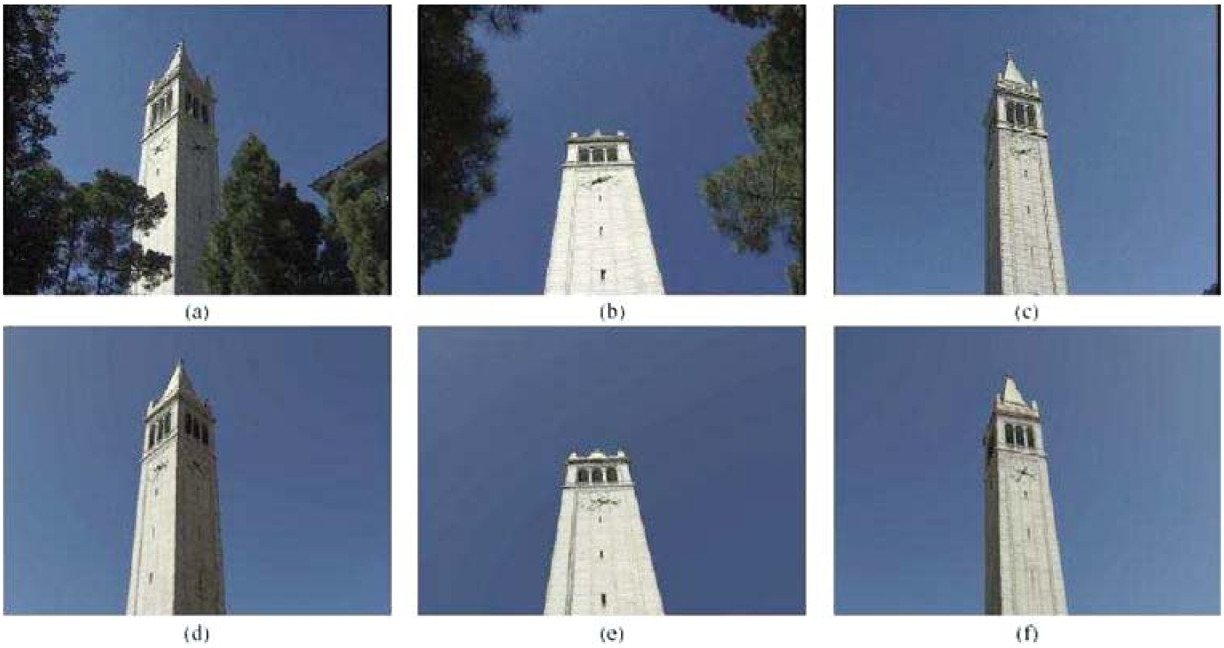

Figure 18: *Results for Yu and Malik [72]. The top row images illustrates the original tower from different viewpoints. The bottom row are synthetic images of the tower from approximately the same viewpoint. The real images were not used to generate the synthetic images, nevertheless the synthetic and real images look very similar. (Courtesy of Yu and Malik.)*

a hemisphere. Three parameters control the sky intensity. A first scale factor is applied when simulating sunrise and sunset; it is constant otherwise. The second parameter adjusts the intensity of the sky depending on the position of the mesh on the dome. The last parameter controls the sky intensity depending on the sun's position. Next, the radiance values and the pseudo-BRDFs are used to reproduce the global illumination on the local scene. This method is the first to present the possibility of relighting outdoor scenes. Results of these relighted scenes and a comparison image are shown in Figure 18. Although it is difficult to evaluate the quality of the relighting from the images provided by the authors, the images resemble the real conditions, and this can satisfy most of the MR applications for outdoor environments.

4. Discussion

In Section 2.2, we pointed out that the assessment of the various illumination techniques for MR comes with a certain degree of subjectivity. Fortunately, there are some aspects that can be evaluated in a rather objective way. Some of these measures will be used in this section to assess the methods from Section 3. Section 4.1 discusses the amount of *pre-processing* required. In Section 4.2 an evaluation of the *degree of interactivity* is given and in Section 4.3, the methods will be evaluated based on the *quality* of the results. Section 4.4 explains which methods are suitable for outdoor scenes. Finally an overview of the discussed methods is given in Section 4.5.

4.1. Pre-processing time

The term *pre-processes* refers to those steps, carried out once, that are required by the method before the merging of real and virtual objects takes place. The geometry reconstruction, image capturing and BRDF estimation, are considered as pre-processing steps.

A few methods do not require a full geometric model of the real scene: Sato *et al.* [17], Nakamae *et al.* [41] and Haller *et al.* [54]. All other methods require a geometric model. Some of these methods do not explain how this model can be constructed, others assume that it is constructed using semi-manual 3D reconstruction software, examples of such software were given in Section 3. Using reconstruction software usually results in a low resolution model and is in general error prone, this is due to the fact that no automatic, accurate 3D reconstruction software is yet commercially available. Scanning devices give a better resolution, but these devices are expensive and while the scanning of a small object might be straightforward, the scanning of a larger scene is tedious. As a summary we can say that a perfect geometric model is difficult to acquire and that reconstruction is always a tedious work.

Some methods require radiance information captured from several viewpoints [21,71] or under different types of illumination [70,72]. Taking several HDR images from different viewpoints and under different illumination delays the image capture time.

Many methods calculate a BRDF estimate, some use a diffuse model, some allow a more complex model. Often the calculation of the BRDF needs to be carried out off-line, due to timing issues and is therefore considered as pre-processing work. Methods that calculate a diffuse-only BRDF are: [59,61,63,18,70], methods that allow specular components are: [71,20,72,69].

4.2. Level of interactivity

Interactivity means:

- the possibility of navigating objects or viewpoints in the scene,
- the effort made to get an interactive rendering,
- the possibility to modify reflectance properties of real objects in the scene,
- the possibility to modify the illumination sources in the real scene.

A few methods allow to navigate the virtual objects or the viewpoints. These techniques have either enough BRDF information [69,20,61], enough geometry and illumination information [17,72] or use a different approach [14,52,56].

Only a few of the methods operate in true real time (RT) [14,52,60], others are near real time (near RT) [18,71,63] but most of them are nonreal time (NRT). However, it should be noted that some methods were developed years ago, when computer hardware and software were much slower than nowadays. Also, it should be pointed out that some methods did not make a special attempt in producing interactive systems. With a few modifications, it should be possible to speed up most of the described systems.

Some methods that specifically tried to speed up the computations are worth mentioning. Agusanto *et al.* [14] exploited the idea of environment mapping while State *et al.* [52] used shadow mapping and Haller *et al.* [54] shadow volumes. Gibson and Murta [57] developed a new technique to simulate soft shadows at interactive rates and Drettakis *et al.* [63], Loscos *et al.* [18] and Loscos *et al.* [71] made use of a hierarchical radiosity algorithm, that decreased the computation time to interactive rates as well. Gibson *et al.* [60] used shadow volumes.

Most methods that calculate the BRDF values are in principle capable of changing the BRDF values into something new. This can be used to modify the appearances of real objects in the scene. Relighting methods can use this BRDF information to relight a scene using a different illumination pattern. Table 1 gives an overview of the various different types of illumination the discussed methods allow.

4.3. Evaluation of the quality

Some of the described methods evaluate the quality of their method using one or more of the following evaluation methods:

- a comparison is made between a photograph reference of the real scene and a synthetic version of the same scene,
- the BRDF is measured using a device and these results are compared with the calculated BRDF values.

Gibson *et al.* [60] compare their shadow rendering technique with a ray traced rendering and an image of the real scene, see Figure 11. They are capable of producing realistic and similar shadows as in the real image and at a faster time than the ray traced rendering. In [57] the presented extended shadow mapping is compared with a ray-traced version using the same input parameters, see Figure 9. There are some differences between the two synthetic scenes, but the generated shadows look realistic.

Boivin and Gagalowicz [70] extract a full BRDF model and compare their rendering with an original image of the real scene, see Figure 14. In [72] the diffuse and specular components are calculated; the resulting rendering is compared with an original image of the real scene. Similarly, Loscos *et al.* [18,70] estimate the diffuse BRDF and

Table 1: *Overview of illumination methods for mixed reality.*

	Geometric model	Number of images	Methodology	Rendering	Computation time	Section
[14]	No	One	Global common illumination	Environment maps, multipass rendering	RT	3.1
[41]	No	One	Local common illumination	Ray casting	NRT	3.1
[17]	No	One	Global common illumination	Ray casting	NRT	3.1
[52]	Yes	One	Local common illumination	Shadow mapping	RT	3.2
[54]	Yes	One	Local common illumination	Shadow volumes	RT	3.2
[56]	Yes	One	Local common illumination	Ray tracing	NRT	3.2
[60]	Yes	One	Global common illumination	Shadow mapping	RT	3.2
[59]	Yes	One	Global common illumination	Differential rendering + ray tracing	NRT	3.2
[57]	Yes	One	Global common illumination	Extended shadow mapping	Near RT	3.2
[61]	Yes	One	Global common illumination	Radiosity + ray casting	NRT	3.2
[63]	Yes	One	Relighting using global illumination	Hierarchical radiosity algorithm	Near RT	3.2
[18]	Yes	One	Relighting using global illumination	Hierarchical radiosity algorithm	Near RT	3.2
[69]	Yes	One	Inverse global illumination	Ray tracing	NRT	3.2
[70]	Yes	Few	Relighting using global illumination	Hierarchical radiosity algorithm + ray casting	Near RT	3.3
[71]	Yes	Few	Relighting using global illumination	Ray tracing	NRT	3.3
[20]	Yes	Many	Inverse global illumination	Ray tracing	NRT	3.4
[72]	Yes	Many	Inverse global illumination	Ray tracing	NRT	3.4

compare a synthetic rendering with a original image of the real scene (see Figures 3 and 15). In both methods, the rendering occurs at interactive update rates.

Similarly, Gibson *et al.* [71], see Figure 17, compare an original and synthetic image and find that the error between the two images decreases drastically in the first three iterations. Both diffuse and specular reflectances are modelled.

Yu *et al.* [20], see Figure 4, estimate diffuse and specular BRDF values and compare these with measured BRDF values of objects in the scene. The estimates and the true values are similar.

We can also compare methods that use both specular and diffuse BRDF values for the rendering with those that have a more restrictive understanding of the BRDF. It is understood that systems based on a more complete BRDF model result in an MR of a higher quality than those based on diffuse BRDF values only or those that do not estimate BRDF values at all. For some methods, only a subjective user perceptive assessment can be made.

4.4. Usability on indoor and outdoor scenes

The reader may have noticed that most techniques were tested on indoor scenes. Outdoor scenes are more complex than indoor scenes. Not only is the geometry more difficult to model, the illumination is difficult to extract as well. Outdoor illumination is time and weather dependent and difficult to model and simulate. Only three methods from Section 3 explicitly used an outdoor scene to test their method [59,60,17,72] but this does not imply that the other methods are not suitable for outdoor scenes. For instance, one might argue that all methods that use environment maps are capable of capturing the outdoor illumination. But some caution is in place when interpreting this statement [73]. If HDR images are used to capture the environment map, which is in general the case, one needs to bare two things in mind. Firstly, the intensity of the sun is in general too bright to be captured in a HDR image without saturation, even at very fast shutter speeds. Secondly, if the sky is clouded and the clouds drift in the sky, there will inevitable be some movement in the low dynamic images used to compile the HDR image, making them worthless. It should be clear, that the extension from indoor to outdoor scenery is not straightforward. The current state of the art of MR shows no good solutions for the outdoor scenes.

4.5. Overview

Table 1 gives an overview of all methods discussed in Section 3. For each method, the overview discusses the following aspects:

- **Geometric model of the scene**: whether or not the method requires a geometric model of the scene.
- **Number of different images**: the number of different images needed per point in the scene, to calculate the MR.
- **Methodology**: the methodology used to create the MR. In Section 2.3 three different approaches were discussed:
  1. common illumination,
  2. relighting,
  3. inverse illumination.

  Further to this division, a distinction is made between local and global illumination techniques.

- **Rendering**: the rendering method used to compose the MR. Possible answers are: ray-casting, ray-tracing, radiosity, etc.
- **Computation time**: the *update* time of the method is real time (RT), non real time (NRT) or near real time (near RT).

5. Conclusions and Future Work

In the past few years, research has been motivated to consider a new type of simulation: the simulation of a new reality called *mixed reality* that refers to the concept of mixing a real scene with virtual elements. Mixed reality has now become very important for various applications ranging from entertainment with movie post-production and games, architecture, cultural heritage, education and training, etc. Several problems arise when composing reality with virtual elements. A first problem is how to automatically calibrate the relative position and the occlusion of virtual objects with virtual ones. As mentioned in this paper, this has been addressed successfully in two different ways. One can use a scanning device or one can use real-time stereovision to extract depth and shape information. A second problem is how to illuminate the virtual objects consistently with the original light conditions in the real scene. Research papers appeared already in the late 1980 to answer this last need of the industry, but it is only recently, within the last 10 years, that the international community made a more significant effort to provide more automated solutions for computing the illumination in mixed reality.

Although it is tempting to compare techniques relatively to the quality of the results achieved, this paper classifies them depending on the context and the goal of the method. Firstly, it is of course easier to compute illumination for mixed reality if a 3D model is available. Secondly, it may be that only a few images of the real scene are available from different viewpoints, and some available with different lighting conditions. The more images are available for the illumination extraction, the easier the computation procedure becomes. On the contrary, the fewer images are available the more difficult it is to perform an accurate estimation and therefore simulation. Consequently, it was decided that it would be fairer and more interesting to compare techniques using similar types of data. Four different categories were identified based on the amount of geometric detail and radiance information available. Different manners to compare the illumination techniques used for mixed reality were presented as well. For example, the methods were compared based on the type of the illumination achieved: local or global, diffuse or complex illumination effects. It was also pointed out if relighting was possible and if the user could interact with the scene.

An ideal conclusion of this paper would state which technique is the most perfect one. However, it is impossible to assess the methods without knowing the application at hand. It is therefore very difficult to describe the ideal method. It should be a real-time and automatic method with no pre-processing requirements. It would allow any type of virtual interaction: modification of lighting, removal of real objects, modification of material properties of real objects and addition of virtual elements. And the rendering quality would perfectly match with the real one. Research is heading toward this, and it is likely that this technology will become more accessible in future years. Progress in stereovision techniques, in automatic calibration, registration and in computer graphics will help in the progression in illumination for mixed reality. More automatic reconstruction methods of the geometry are needed, that will also model more complex details. Progress in augmented reality is heading toward systems being able to recognize shape and depth without markers. Computer graphics research needs to provide more precise description of reflection models and rendering software needs to be adapted to these more complex materials. Little work has been done in modeling the behavior of light sources, which are often assumed diffuse. It will be important for future work to consider more complex lighting in order to find a better estimate for the illumination in mixed reality. Finally, most of the methods have been designed for indoor environments that are easier to control. Outdoor environments present a real challenge, both in the capture and in the simulation. It is expected that more work for outdoor environments will appear in the near future.

Acknowledgments

We would like to thank Simon Gibson, Greg Ward, Anthony Steed, Mel Slater, Erik De Witte, Hila Ritter Widerfield and Bernhard Spanlang for reviewing this paper. This work was partly funded by CREATE (IST-2001-34231), a 3-year RTD project funded by the 5th Framework Information Society Technologies (IST) Programme of the European Union and a UCL graduate school research scholarship.

References

1. P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems E77-D*, 12, 1321–1329, December 1994.

2. Y. Ohta and H. Tamura. *Mixed Reality - merging real and virtual worlds*. Ohmsha and Springer-Verlag, Chapter 1, by Milgram Paul and Colquhoun Herman Jr, 1999.

3. R. Azuma. A survey of augmented reality. In *ACM Siggraph '95, Course Notes No. 9: Developing Advanced Virtual Reality Applications*. pp. 1–38, August 1995.

4. R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics Applications*, *21*(6), 34–47, 2001.

5. R. Behringer, G. Klinker and D. W. Mizell (Eds.). *Augmented Reality: Placing Artificial Objects in Real Scenes*, 1999.

6. A. State, D. T. Chen, C. Tector, A. Brandt, H. Chen, R. Ohbuchi, M. Bajura and H. Fuchs. *Case Study: Observing a Volume Rendered Fetus Within a Pregnant Patient*. Tech. Rep. TR94-034, University of North Carolina, Chapel Hill, 18, 1994.

7. O. Bimber, A. Grundheimer, G. Wetzstein and S. Knodel. Consistent illumination within optical see-through augmented environments. In *Proceedings of IEEE/ACM International Symposium on Augmented and Mixed Reality (ISMAR'03)*, San Diego, California, ACM Press, pp. 198–207, 2003.

8. M. Slater, M. Usoh and Y. Chrysanthou. The influence of dynamic shadows on presence in immersive virtual environments. In *Selected papers of the Eurographics workshops on Virtual environments '95*, Barcelona, Spain, Springer-Verlag, pp. 8–21, 1995.

9. T. Howard, S. Gibson and A. Murta. Virtual environments for scene of crime reconstruction and analysis. In *proceedings of SPIE Electronic Imaging*, vol. 3960, 2000.

10. G. Patow and X. Pueyo. A survey on inverse rendering problems. *Computer Graphics Forum*, *22*(4), 663–687, 2003.

11. R. Ramamoorthi and S. Marschner. Acquiring material models using inverse rendering. In *ACM Siggraph '02 , Course Notes #39)*. Organizer:Ramamoorthi Ravi, Marschner Steve, Lecturers: Samuel Boivin, George Drettakis, Hendrik P. A. Lensch, Yizhou Yu, 2002.

12. S. R. Marschner. *Inverse rendering in computer graphics*. PhD thesis, Department of Computer Graphics, Cornell University, Ithaca, NY, Program of Computer Graphics, Cornell University, Ithaca, NY, 1998.

13. P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of ACM Siggraph '00 (Computer Graphics)*, New York, NY, USA, ACM Press/Addison-Wesley Publishing Co., pp. 145–156, 2000.

14. K. Agusanto, L. Li, Z. Chuangui and N. W. Sing. Photorealistic rendering for augmented reality using environment illumination. In *proceedings of IEEE/ACM International Symposium on Augmented and Mixed Reality (ISMAR '03)*. pp. 208–216, October 2003.

15. Y. Sato, M. D. Wheeler and K. Ikeuchi. Object shape and reflectance modeling from observation. In *proceedings of ACM Siggraph '97 (Computer Graphics)*, pp. 379–388, 1997.

16. I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Amsterdam, The Netherlands, ACM Press, pp. 488–493, 1993.

17. I. Sato, Y. Sato and K. Ikeuchi. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Transactions on Visualization and Computer Graphics*, *5*(1), 1–12, 1999.

18. C. Loscos, G. Drettakis and L. Robert. Interactive virtual relighting of real scenes. *IEEE Transactions on Visualization and Computer Graphics 6*(3), 289–305, July-September 2000.

19. G. J. Ward. Measuring and modeling anisotropic reflection. In *proceedings of ACM Siggraph '92 (Computer Graphics)*, New York, NY, USA, ACM Press, pp. 265–272, 1992.

20. Y. Yu, P. Debevec, J. Malik and T. Hawkins. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *proceedings of ACM Siggraph '99 (Computer Graphics)*, ACM Press/Addison-Wesley Publishing Co., pp. 215–224, 1999.

21. G. Drettakis and F. X. Sillion. Interactive update of global illumination using a line-space hierarchy. In *Proceedings of ACM Siggraph '97 (Computer Graphics)*, pp. 57–64, 1997.

22. L. S. Nyland. Capturing dense environmental range information with a panning, scanning laser rangefinder. www.cs.unc.edu/ ibr/projects/rangefinder.

23. D. McAllister, L. Nyland, V. Popescu, A. Lastra and C. McCue. Real-time rendering of real-world environments. In *proceedings Eurographics Workshop on Rendering (Rendering Techniques '99)*, pp. 145–160, 1999.

24. 3Dscanners. www.3dscanners.com.

25. R. Hartley, R. Gupta and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of Computer Vision and Pattern Recognition*, IEEE Computer Society Press, pp. 761–764, 1992.

26. R. Hartley: In Defence of the eight point algorithm. *IEEE International Conference on Computer Vision 19*(6), 580–593, June 1997.

27. O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proceedings of the 2nd*

European Conference on Computer Vision, G. Sandini, (Ed.), vol. 588, pp. 563–578, May 1992.

28. O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. MIT press, 1993.

29. O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin and I. Zoghlami. 3-d reconstruction of urban scenes from image sequences. *CVGIP : Image Understanding 69*(3), 292–309, 1998.

30. P. Poulin, M. Ouimet and M. C. Frasson. Interactively modeling with photogrammetry. In *proceedings of Eurographics Workshop on Rendering (Rendering Techniques '98)*, pp. 93–104, June 1998.

31. P. E. Debevec, C. J. Taylor and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proceedings of ACM Siggraph '96 (Computer Graphics)*, pp. 11–20, 1996.

32. P. Debevec, G. Borshukov and Y. Yu. Efficient view-dependent image-based rendering with projective texture-mapping. In *Proceedings of the 9th Eurographics Workshop on Rendering (Rendering Techniques '98)*, Vienna, Austria, Springer-Verlag, pp. 105–116, June 1998.

33. V. Meas-Yedid, J.-P. Tarel and A. Gagalowicz. Calibration métrique faible et construction interactive de modèles 3d de scènes. In *Congrès Reconnaissance des Formes et Intelligence Artificielle*, Paris, France, AFCET, 1994.

34. Realviz. Image modeller. online.www.realviz.com.

35. MetaCreations. Canoma. www.metacreations.com/products/canoma.

36. Eos Systems Inc. Photomodeller. www.photomodeler.com.

37. Integra: Renoir. www.integra.co.jp/eng/products/renoir.

38. G. W. Larson. Real pixels. In *Graphics Gems II*, J. Arvo, (Ed.), pp. 80–83, 1991.

39. P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM Siggraph '97 (Computer Graphics)*, pp. 369–378, 1997.

40. T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Collins Fort, pp. 374–380, June 1999.

41. E. Nakamae, K. Harada, T. Ishizaki and T. Nishita. A montage method: the overlaying of the computer generated images onto a background photograph. In *proceedings of ACM Siggraph '86 (Computer Graphics)* (New York, NY, USA, ACM Press, pp. 207–214, 1986.

42. J. F. Blinn and M. E. Newell. Texture and reflection in computer generated images. *Communications of the ACM 19*(10), 542–547, 1976.

43. N. Greene. Environment mapping and other applications of world projections. *IEEE Computer Graphics Applications 6*(11), 21–29, 1986.

44. D. Voorhies and J. Foran. Reflection vector shading hardware. In *Proceedings of ACM Siggraph '94*, New York, NY, USA, ACM Press, pp. 163–166, 1994.

45. B. Cabral, M. Olano and P. Nemec. Reflection space image based rendering. In *Proceedings of ACM Siggraph '99 (Computer Graphics)*, New York, NY, USA, ACM Press/Addison-Wesley Publishing Co., pp. 165–170, 1999.

46. W. Heidrich and H.-P. Seidel. Realistic, Hardware-Accelerated Shading and Lighting. In *Proceedings of ACM Siggraph '99 (Computer Graphics)*, New York, NY, USA, ACM Press/Addison-Wesley Publishing Co., pp. 171–178, 1999.

47. J. Kautz and M. D. McCool. Approximation of glossy reflection with prefiltered environment maps. In *proceedings of Graphics Interface '00*, pp. 119–126, 2000.

48. NVidia. Cube environment mapping. www.nvidia.com/object/feature˙cube.html.

49. ATI. Car paint. www.ati.com/developer/demos/r9700.html, 2002.

50. S. E. Chen. QuickTime VR—an image-based approach to virtual environment navigation. In *Proceedings of ACM Siggraph '95 (Computer Graphics)*. pp. 29–38, 1995.

51. H. Kato, M. Billinghurst, R. Blanding and R. May. *AR-Toolkit*. Tech. rep., Hiroshima City University, 1999.

52. A. State, G. Hirota, D. T. Chen, W. F. Garrett and M. A. Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *proceedings of ACM Siggraph '94 (Computer Graphics)*. pp. 429–438, 1994.

53. J. Blinn. Me and my (fake) shadow. *IEEE Computer Graphics Applications 8*(1), 82–86, 1988.

54. M. Haller, S. Drab and W. Hartmann. A real-time shadow approach for an augmented reality application using shadow volumes. In *proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '03)*, 2003.

55. K. Jacobs, C. Angus, C. Loscos, J.-D. Nahmias, A. Reche and A. Steed. Automatic consistent shadow generation for augmented reality. In *proceedings of Graphics Interface '05*, 2005.

56. P. Jancene, F. Neyret, X. Provot, J.-P. Tarel, J.-M. Vezien, C. Meilhac and A. Verroust. Computing the interactions between real and virtual objects in video sequences. In *proceedings of 2nd IEEE Workshop on Networked Realities*, pp. 27–40, October 1995.

57. S. Gibson, A. Murta. Interactive rendering with real-world illumination. In *Proceedings of 11th Eurographics Workshop on Rendering (Rendering Techniques '00)*, London, UK, Springer-Verlag, pp. 365–376, June 2000.

58. G. J. Ward. The radiance lighting simulation and rendering system. In *proceedings of ACM Siggraph '94 (Computer Graphics)*, New York, NY, USA, ACM Press, pp. 459–472, 1994.

59. P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM Siggraph '98 (Computer Graphics)*, pp. 189–198, 1998.

60. S. Gibson, J. Cook, T. Howard and R. Hubbold. Rapid shadow generation in real-world lighting environments. In *proceedings of the 13th Eurographics workshop on Rendering (Rendering Techniques '03)* (Aire-la-Ville, Switzerland, Switzerland, Eurographics Association, pp. 219–229, 2003.

61. A. Fournier, A. S. Gunawan and C. Romanzin. Common illumination between real and computer generated scenes. In *proceedings of Graphics Interface '93*, Toronto, Canada, pp. 254–262, May 1993.

62. M. F. Cohen, S. E. Chen, J. R. Wallace and D. P. Greenberg. A progressive refinement approach to fast radiosity image generation. In *Proceedings of ACM Siggraph '88 (Computer Graphics)*, ACM Press, pp. 75–84, 1988.

63. G. Drettakis, L. Robert and S. Bugnoux. Interactive common illumination for computer augmented reality. In *Proceedings of the 8th Eurographics Workshop on Rendering (Rendering Techniques '97)*,Etienne Saint, France, June 1997.

64. P. Hanrahan, D. Salzman and L. Aupperle. A rapid hierarchical radiosity algorithm. In *proceedings of ACM Siggraph '91 (Computer Graphics)*, pp. 197–206, 1991.

65. H. E. Rushmeier, C. Patterson and A. Veerasamy. Geometric simplification for indirect illumination calculations. In *proceedings of Graphics Interface '93*, pp. 227–236, May 1993.

66. F. X. Sillion. A unified hierarchical algorithm for global iillumination with scattering volumes and object clusters. *IEEE Transactions on Visualization and Computer Graphics*, 1(3), 240–254, 1995.

67. B. Smits, J. Arvo and D. Greenberg. A clustering algorithm for radiosity in complex environments. In *proceedings of ACM Siggraph '94 (Computer Graphics)*, pp. 435–442, 1994.

68. C. Loscos. *Interactive relighting and remodelling of real scenes for augmented reality*. PhD thesis, iMAGIS-GRAVIR/IMAG-INRIA,. PhD thesis, 1999.

69. S. Boivin and A. Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proceedings ACM Siggraph '01 (Computer Graphics)*, New York, NY, USA, ACM Press, pp. 107–116, 2001.

70. C. Loscos, M.-C. Frasson, G. Drettakis, B. Walter, X. Granier and P. Poulin. Interactive virtual relighting and remodeling of real scenes. In *Proceedings of 10th Eurographics Workshop on Rendering (Rendering Techniques '99)*, New York, NY, D. Lischinski, G. Larson, (Eds.), vol. 10, Springer-Verlag/Wien, pp. 235–246, June 1999.

71. S. Gibson, T. Howard and R. Hubbold. Flexible image-based photometric reconstruction using virtual light sources. In *proceedings of Eurographics 2001*. Manchester, UK, September 2001.

72. Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *proceedings of ACM Siggraph '98 (Computer Graphics)*, New York, NY, USA, ACM Press, pp. 207–217, 1998.

73. K. Jacobs and C. Loscos. Relighting outdoor scenes,. BMVA one day symposium: Vision, Video and Graphics, July 2004.