Simplifying Real-Time Light Source Tracking and Credible Shadow Generation for Augmented Reality using ARToolkit

Bjarne Fisker Jensen, Jacob Sloth Laursen, Jacob Boesen Madsen and Thomas Wisbech Pedersen

Institute for Media Technology, Aalborg University

ABSTRACT

This paper presents an interactive vision based real-time system for estimating light source positions and generating credible shadows for augmented reality. The implementation uses ARToolkit as a basis for geometric tracking, and a reflective sphere for tracking light sources in the environment. The paper seeks to generate perceptually credible shadows. User testing was conducted in order to determine the minimum criteria for the credibility of the shadows. User testing showed that 64 shadows are sufficient and indistinguishable from more complex compositions and a real image. It was clear that users need a reference object to distinguish between real and virtual shadows.

The implementation and performance are tested using a consumer-grade web-camera and a regular laptop computer. The implementation can perform real-time with 256 generated shadows.

KEYWORDS: Image Based Lighting, Human Perception, Augmented Reality.

INDEX TERMS: D.2.6 [Programming Environments]: Interactive Environments; H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities;

1 INTRODUCTION

Augmented reality is a term for interactive and real-time visualization of virtual objects in a real scene [1]. In the real world, if an object is lit, it will cast a shadow. Despite shadows being important for easier three dimensional perceptions and the spatial relationship amongst objects [2], shadows are not always used in virtual or augmented settings.

Shadows most often reveal information of the scene, such as spatial relationships and revealing shapes of objects [2], [3]. To reconstruct a realistic scene in a virtual space is a complex task, in terms of parameters influencing the lighting and shading of a room. These parameters are i.e. the light sources, the luminance and the placement of objects, the geometry of the environment and the structural composition of the environment. People have a precise and well developed understanding of the underlying implications of shadows [1].

In this project, the research has only found Kanbara et al. [4] with a lightweight real-time shadow generating system, even though the volume of research in the field of shadow detection and generation is quite large and extensive. Therefore we seek to make a lightweight implementation for estimating light source positioning and credible shadow generation.

User testing will determine the minimum criteria for the credibility of the shadows, and so will the feasibility of algorithmically adding of visual effects that will be used to reduce the number of light sources necessary to generate credibility in the augmented shadows.

The project seeks implement an augmented reality system capable of tracking light sources using state of the art techniques from the scientific community. The implementation should focus on an easy-to-use approach and should make use of cheap and simple tools and equipments.

2 RELATED WORKS

A lot of research has been devoted to find illumination in the environment and to generate shadows based on this information for augmented reality. An often occurring trend in this research, is the need for a lot of different prerequisites, i.e. known geometry [5], [4], pre-computed environment map [5], stereo-camera setup [6], [7], etc.

Yan [7] proposed a method for estimating illumination parameters in an environment using two cameras to gather information of two lambertian spheres, each visible to the light source. This method generates parameters for a single point light source and ambient light.

Kanbara et al. [4] estimates the light source environment in real-time using a mirror ball for photometric registration. This method uses a generated light source map from the mirror ball and estimates the eight brightest light sources.

Supan et al. [6] uses an approach which tracks a mirror ball and generates a reflective environment map from the mirror ball information, as well as a diffuse irradiance map. Shadows are created by placing a reasonable amount of light sources in the scene that gets their intensity and color information from the environment map.

Based on Kanbara et al. [4], the method in this paper aims to create a real-time application for estimating light source positions and generate perceptually correct shadows using only general tools available to people, to stay in the spirit of ARToolkit.

The implementation should create shadows that are credible to the human perception with regards to placement, softness and length of the shadows. Thus, it is necessary to examine people's perception of shadows. The information gathered from testing will be compared to Nakano et al. [8], which states that 32-128 shadows as the acceptable threshold, above which people cannot distinguish between virtual and real shadows.

Bjarne Fisker Jensen, Jacob Sloth Laursen, Jacob Boesen Madsen and Thomas Wisbech Pedersen are masters' students at Medialogy, Aalborg University. E-mail: {bjen07, jlau07, jmad07, twpe07}@imi.aau.dk

3 METHODS

3.1 Implementation of the system

The implementation should track and process a known marker object and generate a user-defined number of light source positions, using the median cut algorithm [9]. This limits the implementation to use 2^n shadows, due to the median cut algorithm, dividing each area into two at each iteration.

OpenGL [10] is used for rendering the graphics to the screen, as it is the default renderer for ARToolkit [11].

In order to estimate light source positions, a modified ARToolkit marker is used with a reflective sphere at its center. The marker is tracked for position in space and the sphere has information of the light sources in the environment. An example of the setup can be seen in Fig. 1.

Fig. 1. The setup used for estimating light source position using a laptop computer and a standard consumer web-camera

The pipeline of the system is showed in Fig. 2. This serves as a basis for the implementation of the system.

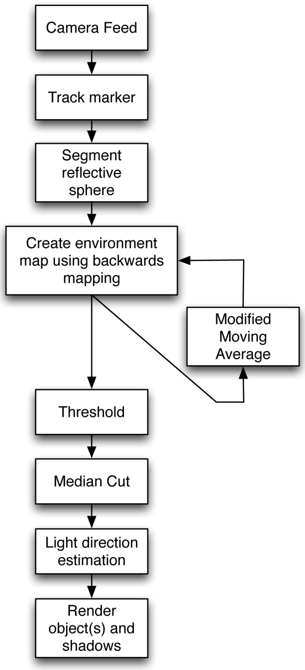

Fig. 2. Pipeline of the system

The light tracking marker consists of a table tennis ball painted with a glossy black paint that can be bought in most convenience stores.

Different colored mirror balls were analyzed to find the best color suited for light tracking without a dynamic range image. An example of these can be seen in Fig. 3.

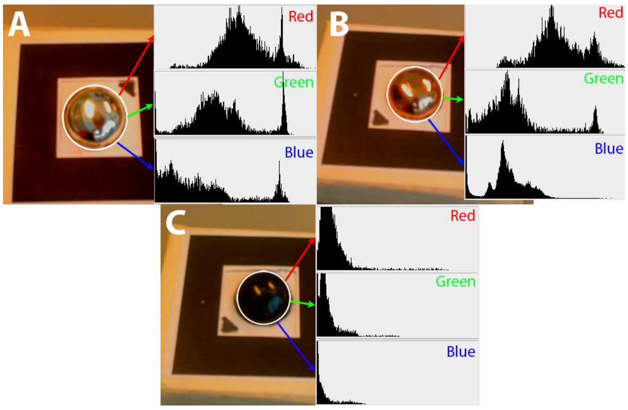

Fig. 3. Analysis of different colored spheres, to find optimal sphere coloring for non-HDRI

A black colored sphere with high gloss was selected, as it shifts less intense light to the lower end of the RGB scale, resulting in more data on higher intensity lights, making it better to track light sources.

As Kanbara et al. demonstrated in [4] the reflective sphere can be segmented , by projecting the center of the sphere and a point on the periphery onto the image plane. This can be done using the model-view matrix supplied by ARToolkit and the known relation between the marker and the reflective sphere. This method provides the center and the radius of the circle representing the sphere onto the image plane. The direction of a given reflection $\vec{r}_{reflection}(1, \theta, \varphi)$ can be calculated using the radius, rMax, of the circle in image coordinates, using the pixel coordinate P in the segmented circle. This is done by assuming that the camera is positioned at an infinite distance along the positive z-axis, thus making all camera rays parallel. The formulas (1) and (2) are used. An illustration of the approach can be seen in Fig. 4.

$$\theta = 2 * \sin^{-1}\frac{\sqrt{Px^2 + Py^2}}{rMax} \qquad ,\theta \in [0;\pi] \tag{1}$$

$$\varphi = \mathrm{atan2}(\mathrm{Py}, \mathrm{Px}) \qquad ,\varphi \in [-\pi;\pi] \tag{2}$$

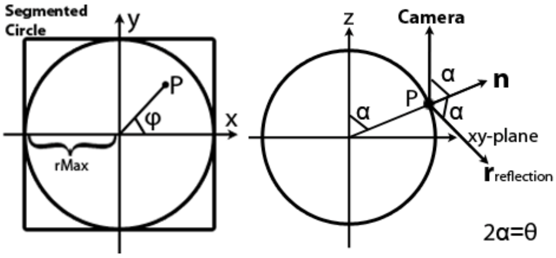

Fig. 4 Illustration of the approach for finding spherical coordinates from 2D representation of a sphere

An environment map with aspect ratio 2:1, of the reflections in the sphere is created using backwards mapping. By applying a modified moving average [12], [13] on the environment map, any sudden changes can be smoothed more or less depending on the sample size without raising computation time significantly. To acquire light directions, the median cut algorithm [9] is applied to the environment map. This is used to estimate 8 OpenGL light source positions and 2^n light source directions for shadow generation. The environment map and light source directions generated using median cut can be seen in Fig. 5.

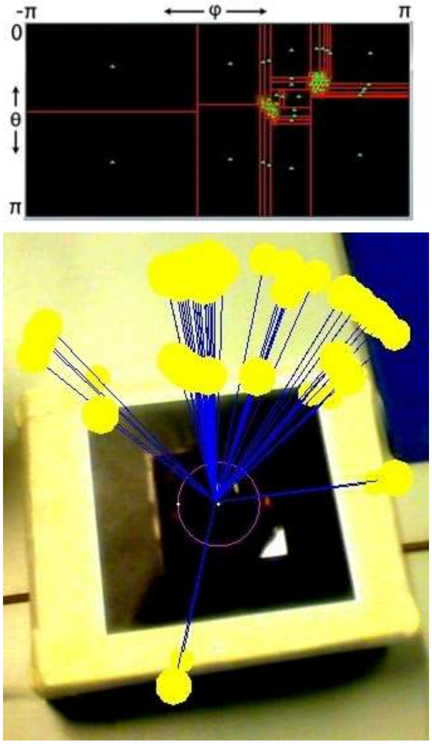

Fig. 5. Environment map of the black reflective sphere, with estimated light source directions, generated from the median cut algorithm.

3.2 Establishing criteria for realistic virtual shadows

A threshold for credibility of perceived virtual shadows as real ones must be established through testing. The test will seek to find whether it is possible to enhance virtual shadows by applying visual effects. It is assumed that an effective shadow generation algorithm can be produced, by finding an optimal compromise between realism and performance. The hypotheses for the test are:

- It is possible to create computer generated virtual shadows, that the user would rank as credible as real shadows.
- It is possible to find a lower bound of light source positions necessary to make the shadows believable to the user.
- By blurring the virtual shadows, it is possible to lower the bound of necessary light source positions.

From the hypotheses, two main problems for the test can be composed:

1. Determine the number of light source positions necessary to generate sufficient shadows.
2. Determine whether the use of visual effects can enhance credibility of virtual shadows.

The test is composed of still images taken in three different real environments with various light settings. Some of the shadows on the images were removed using Photoshop [14] and new virtual shadows were instead created and superimposed by using a HDR environment map created from a reflective sphere and the median cut algorithm [9] to calculate light positions. In each setting 10 images were composed with shadows generated from $[0, 2^1, 2^2, \ldots, 2^8]$ light sources, with and without a boxblur [15] applied, and 1 image with genuine shadows were used as a control image. The three different setting are shown in Fig. 6, each with 256 light sources. In B and C where three objects are present, only the middle object has virtual shadows. This gives a paper-cube, a book and a box setting.

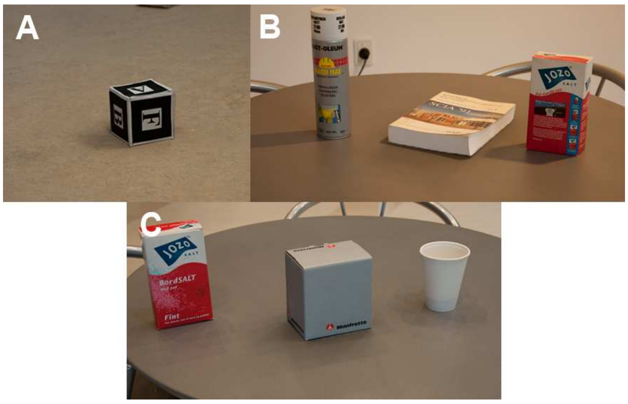

Fig. 6. Example images from the 3 different settings for use in the test. All with superimposed virtual shadows created with 256 light sources. In a) and b) only the middle objects have virtual shadows

The images are used to compare the imitated shadows against the real image and against each other, in order to find a boundary for credible shadows. The test will consist of two test stages:

1. An initial rank-order test to rank credibility of composited images. This first test act as a pilot test, to find any errors that might emerge from the images and test setting. Candidates from the initial test are used in a subsequent final test.
2. A second final test is a dose-response scenario [16] determining the percentage of the test population that judges an image as real or imitated. This test uses candidates from the first initial test.

4 EXPERIMENTS

4.1 Results from shadow criteria test

The initial rank order test was performed on 13 participants, who were set to rank each image from least credible to most credible.

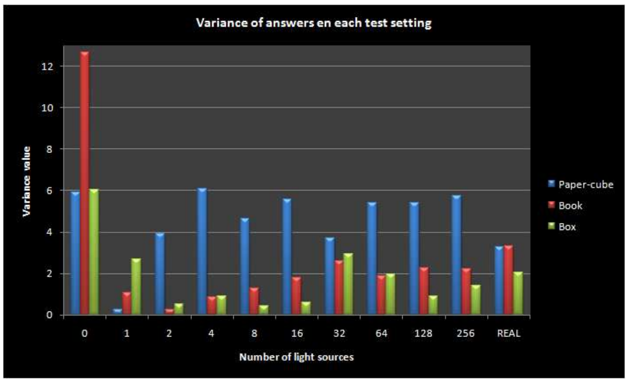

Fig. 7. Rank-order test and the plotted variance for the answers of each image in the different test settings

The diagram of Fig. 7 indicates two observations that are essential for our final dose-response test. First it indicates that an object that casts no shadows at all is very hard for people to rank. Second, the whole test setting with the paper-cube seems to confuse people as they have very scattered opinions on the ranking of each image. This could indicate that people it difficult evaluating shadows in images where no real shadows are present as a reference. Based on the results from the first test, the first test setting with no real shadows as reference together with the images with zero shadows is discarded for the final dose-response test.

The final dose-response test was performed as an online questionnaire and was posted on Facebook [17] and on the forum of Nordic-T [18]. This is done in order to reach a broad number of different people. This of course, could influence the results because both sites targets specific users, but overall it should give a fairly good understanding of how people perceive virtual shadows. 500 people took part in the test. To prevent the participant from studying an image for too long and not getting accustomed to the task, only 4 images from each of the two setting were shown, for a duration of 10 seconds. The participant then had to make a choice whether he would rate the shadow as obviously imitated or if he would not have noticed anything wrong. Also the participant had the possibility of submitting a “don’t know” answer, instead of choosing between credible and imitated.

In the diagram of Fig. 8, the results from the final dose-response test can be seen. Some data have been discarded after a thorough analysis. The data that have been discarded are that of participants that have taken a very long time to submit their answers. As the test is about finding the intuitive credibility criteria from test subjects, it makes sense that subjects, who answered did not know whether the shadows was imitated or real, is not seen as an obvious imitated shadow. As a consequence this data is plotted as if it was equal to a credible answer from the participant.

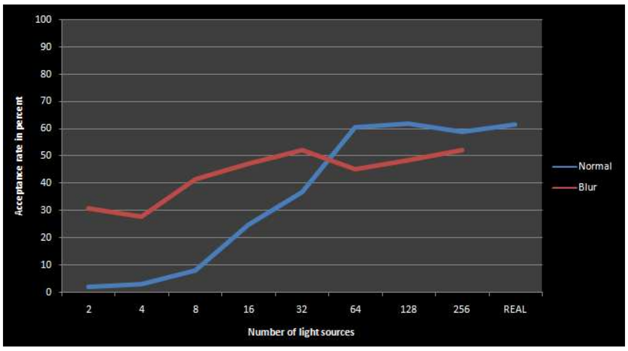

Fig. 8. Dose-response relationship between number of light sources and acceptance rate

The first apparent result is that generally people are very skeptic in an online questionnaire about shadows, as the maximum acceptance rate, even for real images, are just above 60 pct. This is in contrast to the initial first test, where many participants rated images with about 64 shadows as images that could have been real. But if we hold on to the fact that the real image indeed is real, then the test shows that 64 shadows or more should be enough to replicate a real scene. Also the visual effects seem to have a very positive effect on images with a low number of light sources. But as we pass 64 light sources or higher, the blur effect seems to have the opposite effect and just make the image less credible. On the other hand, when 64 light sources are enough to simulate a real image, it would not be necessary to use the blur effect anyway.

4.2 Test of the implementation

4.2.1 Performance test

The performance test was conducted using a Dell XPS M1530 laptop computer, using windows operating system running on a Core 2 Duo T7500 2.2 Ghz with 2.0 GB RAM, a NVIDIA GeForce GO 8600M GT graphics card with 256 MB RAM, and a web-camera with an 800x600 pixels resolution at maximum 15 frames per second, using a 20 stacks and 20 slices sphere as the virtual test object. The results can be seen in Table 1.

Table 1. Frame rate result of performance test. Camera is limited to 15 fps, which limited the performance

Nr. of shadows	Frame rate (fps)
0	15.0
8	15.0
16	15.0
32	15.0
64	15.0
128	15.0
256	15.0
512	12.1
1028	8.8
2048	5.4

The test results are limited due to the web-camera. In order to limit the need for special equipment, a consumer-grade web-camera has been used [19], which is limited to 15 fps with 800x600 resolution.

4.2.2 Qualitative user test

The final implementation has been evaluated by peer students at the department. Interested parties have volunteered to see and review the results.

The interview form was based on an unstructured interview [20]. The test subjects were presented with a few videos from the implementation, using 2-3 light sources. What we sought to investigate was whether direction, length, softness and general impression of the generated shadows lived up to the standards that they would expect.

The results are positive, based on feedback from peers. The direction and length of the shadows seems to be coherent with expectations. There are however some problems with the number of shadows generated, especially from a limited number of real world light sources. This is a problem due to the nature of the median cut algorithm, when estimating a high number of light source positions. The presence of reference real objects in the scene seems to make it harder to imitate the perception of shadows. Different shadow cast techniques could improve the implementation even further.

5 CONCLUSION

In this paper, a method for light source estimation has been presented, which estimate the positions of light sources in the environment and generate credible shadows of a virtual object augmented into the scene, in real-time. The system uses a reflective sphere for tracking light source direction in the environment, which is mathematically converted to an environment map. The median cut algorithm [9] is performed to find light source directions.

Virtual objects can then be placed in the scene using modified ARToolkit markers with shadows generated from the estimated light source directions.

Our implementation has succeeded in generating credible shadows with correct placement, and the implementation is interactive with frame-rate limited by the current web-camera.

The initial rank-order test confirmed our hypothesis which states that it is possible to imitate shadows using a number of light sources to create virtual ones. It also confirmed that an increasing number of shadows will increase perceptual realism, up until virtual and real shadows no longer can be distinguished from each other.

The test showed an acceptance rate of about 60% for a real image in the tests, which also is the level for 64 and more virtual shadows. This indicated that using 64 or more shadows in this implementation will generate shadows credible to human perception. This is within the range specified by Nakano et al. [8].

The qualitative interview with peers, who had seen a demonstration of the implementation, showed that the implementation had the desired general impression that was sought.

6 FURTHER DEVELOPMENT

The estimation of light source positions can be improved by using a high dynamic range camera and getting high dynamic range information of the reflective sphere. This will lead to more precise light source estimation. The coloring of the light should also be considered, as the current implementation only use RGB values for determining intensity through median cut, and not coloring. This information could be used to generate color-bleeding as well.

The current shadows of the scene should be tracked to avoid drawing shadows into shadows already in the scene, making the appearance look fake [21].

The type of shadows should be considered as well. Having the option to render projective shadows, shadow mapping and shadow volumes should be an option to allow users the ease of generating shadows suited for their application.

ACKNOWLEDGEMENTS

This paper is a semester project from 7th semester Medialogy, Institute of Media Technology, Aalborg University.
The authors would like to thank Claus B. Madsen for his assistance which helped improve the project.

REFERENCES

[1] C. B. Madsen, "The Importance of Shadows in Augmented Reality," p. 4, 2003.

[2] N. Sugano, *et al.*, "The effects of shadow representation of virtual objects in augmented reality," *Second Ieee and Acm International Symposium on Mixed and Augmented Reality, Proceedings,* pp. 76-83
363, 2003.

[3] J. M. Hasenfratz, *et al.*, "A survey of real-time soft shadows algorithms," *Computer Graphics Forum,* vol. 22, pp. 753-774, Dec 2003.

[4] M. Kanbara and N. Yokoya, "Real-time estimation of light source environment for photorealistic augmented reality," *Proceedings of the 17th International Conference on Pattern Recognition, Vol 2,* pp. 911-914
1048, 2004.

[5] M. S. Andersen, *et al.*, "Estimation of dynamic light changes in outdoor scenes without the use of calibration objects," *18th International Conference on Pattern Recognition, Vol 4, Proceedings,* pp. 91-94
955, 2006.

[6] P. Supan, *et al.*, "Image Based Shadowing in Real-Time Augmented Reality," *The International Journal of Virtual Reality,* vol. 3, pp. 1-7, September 2006.

[7] F. Yan, "Estimation of light source environment for illumination consistency of augmented reality," *Cisp 2008: First International Congress on Image and Signal Processing, Vol 3, Proceedings,* pp. 771-775
815, 2008.

[8] G. Nakano, *et al.*, "Generating Perceptually-Correct Shadows for Mixed Reality," *7th Ieee International Symposium on Mixed and Augmented Reality 2008, Proceedings,* pp. 173-174
211, 2008.

[9] P. Debevec, "A Median Cut Algorithm for Light Probe Sampling," presented at the International Conference on Computer Graphics and Interactive Techniques, Los Angeles, California, 2005.

[10] 20. December). *OpenGL.* Available: http://www.opengl.org/

[11] P. Lamb. (1999, 20. December). *ARToolKit* Available: http://www.hitl.washington.edu/artoolkit/

[12] P. J. Kaufman, *Trading systems and methods*, 3rd ed. New York ; Chichester: Wiley, 1998.

[13] 20. December). *Modified Moving Average*. Available: http://en.wikipedia.org/wiki/Moving_average

[14] (2009, 12. Dec). *Adobe*. Available: http://www.adobe.com/products/photoshop/photoshop/

[15] 12. Dec). *Wikipedia*. Available: http://en.wikipedia.org/wiki/Box_blur

[16] A. S. Kane. 12. Dec). *DOSE RESPONSE CONCEPTS*. Available: http://aquaticpath.umd.edu/appliedtox/dose-response.pdf

[17] (2004, 12. Dec). *Facebook*. Available: www.facebook.com

[18] 12. Dec). *Nordic-T*. Available: http://nordic-t.org/

[19] *Logitech C905*. Available: http://www.logitech.com/index.cfm/webcam_communications/webcams/devices/5868&cl=dk,da

[20] H. Sharp*, et al.*, *Interaction design : beyond human-computer interaction*, 2nd ed. Chichester ; Hoboken, NJ: Wiley, 2007.

[21] K. Jacobs*, et al.*, "Automatic generation of consistent shadows for augmented reality," presented at the Proceedings of Graphics Interface 2005, Victoria, British Columbia, 2005.