Eurographics Symposium on Rendering 2003
Per Christensen and Daniel Cohen-Or (Editors)

Wavelet Environment Matting

Pieter Peers Philip Dutré†

Department of Computer Science
Katholieke Universiteit Leuven

Abstract

In this paper we present a novel approach for capturing the environment matte of a scene. We impose no restrictions on material properties of the objects in the captured scene and exploit scene characteristics (e.g. material properties and self-shadowing) to minimize recording time and to bound the error. Using a CRT monitor, wavelet patterns are emitted onto the scene in order of importance to efficiently construct the environment matte. This order of importance is obtained by means of a feedback loop that takes advantage of the knowledge learned from previously recorded photographs. Once the recording process is finished, new backdrops can be efficiently placed behind the scene.

Categories and Subject Descriptors (according to ACM CCS): G.1.2 [Numerical Analysis]: Approximation; I.3.7 [Computer Graphics]: Three dimensional graphics and realism; I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture;

1. Introduction

Environment matting and compositing, an extension of the conventional matting process[12, 14], was first presented by Zongker et al.[17] and later extended by Chuang et al.[1]. Unlike conventional matting, an environment matte does not only represent the opacity of a pixel, but it also includes the reflection and refraction effects of the backdrop through the scene. To create an environment matte, a scene is photographed from a single vantage point against a series of known background patterns. Using the information from the recorded photographs, an approximation of the light transport from the background through the scene into the camera is computed. With this approximation, a new image of the captured scene can be generated (i.e. composited) with any backdrop image (figure 1).

Figure 1: *A dinner scene captured with our technique and composited with two novel backdrops.*

1.1. Environment matting

In the approach of Zongker et al.[17] horizontal and vertical stripe patterns are emitted onto a scene. For each emitted pattern a photograph of the scene is recorded from a fixed viewpoint. The environment matte, which encodes the reflection and refraction properties of the scene, is represented for each pixel by a single reflection coefficient and a normalized box filter on a rectangular support area on the backdrop. A least squares optimization procedure is used to extract the support area and reflection coefficient from the recorded pho-

† {pieterp, phil}@cs.kuleuven.ac.be

tographs. Compositing —i.e. applying a novel backdrop— is performed by filtering, for each pixel, the novel background over the support area and scaling the result by the reflection coefficient. Environment matting allows for backdrop replacement in presence of specular and transparent objects. The method itself is elegant and requires few photographs to be recorded.

This approach, however, as pointed out by Chuang et al.[1], has a few limitations. A single rectangular support area and a single reflection coefficient per pixel are not sufficient to capture the complex reflection and refraction effects of dielectrics or rough materials. In addition, the choice of a rectangular support area can cause excessive blurring in the final image. To address these problems, Chuang et al. sweep different oriented Gaussian stripes across the background to capture the environment matte. This resembles the space-time analysis used in 3D range scanning[2, 7]. The environment matte is approximated by a limited number of oriented elliptical Gaussian filters, each with a single reflection coefficient. Compositing is performed similar to Zongker et al. except that contributions from multiple supports for a pixel on the backdrop are added together.

Chuang et al.[1] also presented an environment matting method for real time acquisition, that uses a single color gradient as backdrop pattern. This method, however, is limited to perfect specular materials that do not modulate the emitted color. In this case, the environment matte is reduced to an image warping function. Wexler et al.[16] presented an environment matting extension that is able to work without knowledge of the exact form of the backdrop images used. It relies on having enough background samples or sufficiently rich backdrop images (e.g. by moving a backdrop image behind the scene) to successfully extract an environment matte.

1.2. Image-based relighting

Environment matting techniques can also be interpreted as image-based relighting methods and they are very practical methods which are able to capture the reflectance field (i.e. the description of the transfer of light through a scene) of objects containing specular materials. This is typically difficult for other image-based relighting methods that use quite a different approach than environment matting. Of interest is the approach followed by Nimeroff et al.[11] who represents the incoming illumination by steerable functions and combines weighted basis images lit by these functions. Based on this approach of combining weighted basis images is the Light Stage[3, 6, 9, 8] which samples a limited number of light source positions around the object. For each light source position a basis image is recorded. These methods can relight objects with material properties ranging from diffuse to glossy, the limiting factor being the relatively sparse sampling frequency of light source positions. Using a denser sampling increases the amount of data and the required time to capture these photographs upto a point that these methods become impractical. To overcome this problem, Matusik et al.[10] presented a clever hybrid solution that combines the Light Stage with environment matting. The reflectance field is split into two distinct parts. The part where the illumination is coming from behind the object is handled by an environment matte, whereas the illumination coming from the remainder of the hemisphere is handled by a coarse Light Stage approach.

The idea of linearly combining weighted basis images is a clean and elegant solution. In this paper we will transfer this idea into an environment matting context. We explore the difficulties and their solutions that are associated with this transfer.

1.3. Objectives

Capturing all lighting effects due to different materials is still impractical with existing relighting methods, because it requires an enormous amount of data to be captured. Environment matting presents a way to reduce this amount of data in case of specular and refractive materials, but suffers from some limitations before it can be used as a general image-based relighting method:

1. The error of the environment matte approximation is unknown, as is the error in the composited images. This error depends on scene properties, the filter on the support areas (e.g. a box filter vs. an elliptical Gaussian filter), the illumination patterns used during the recording process and the background image itself used during compositing.
2. Diffuse surfaces are still problematic, because an elliptical Gaussian filter is not sufficient to capture the effects of diffuse reflections. Diffuse materials have a large area of support which can be irregularly shaped because of occlusion and self-shadowing. These irregularly shaped support areas are difficult to approximate accurately with a limited number of elliptical Gaussian filters.
3. Finally, previous environment matting methods rely on non-linear optimization procedures, which require a significant amount of post-processing time, to compute the final environment matte approximation. Such methods usually depend on a number of parameters (e.g. error-thresholds) which greatly affect the quality of acquired results. Non-linear optimization procedures also require a significant amount of processing time. Increasing accuracy using better filters or more approximation terms, would increase post-processing time even more.

We present a novel method to acquire the environment matte of a scene, that does not suffer from these limitations. Our method is based on linearly combining basis images to create an environment matte, instead of non-linear optimization procedures used by previous environment matting techniques. Each basis image is a photograph of the scene lit by an illumination pattern, a 2D basis function of the incoming illumination. Key to our method is the use of wavelets as illumination patterns. A novel backdrop image is decomposed

using the same wavelet basis functions used for generating the basis images. The coefficient of each wavelet in this decomposition is used to weight the basis image lit by the corresponding wavelet pattern. The final composited image is obtained by summing all weighted basis images. A potential problem is the large number of basis images needed to create an accurate environment matte. However, the number of basis images can be limited by emitting only the patterns that are important for constructing the environment matte. The order of importance is estimated and this estimate is progressively refined during the recording process itself. Our method begins by emitting a few coarse wavelet patterns first. Based on the recorded photographs, a feedback loop determines which is the next most important pattern to emit.

More specifically our method addresses the following objectives:

1. No limitations on material properties are imposed in the scene.
2. Characteristics of a scene are exploited to reduce both the recording and compositing efforts and errors. For example, the user should be able to choose a stop criterion depending on the amount of data or the time spend during recording and still have the best solution possible for the performed work.
3. Relying on user input to specify scene characteristics is error prone and a daunting task; we want to automate the recording process as much as possible.
4. The environment matte should have a bounded error, or at least a reasonable estimate about the error should be available.
5. Finally, post-processing time (e.g. time needed to process the captured data) should be minimal and relatively independent with respect to the chosen accuracy.

In the next section we discuss the outline of our new method and the practical setup (section 2). Next, we develop a novel mathematical framework for environment matting in section 3. In section 4 we introduce the error-tree and show how it can be used to direct the recording process. Practical considerations are discussed in section 5. Finally we discuss the results in section 6 and conclude the paper in section 7.

2. Outline of the technique and practical setup

We use a similar setup as was used in previous environment matting papers. An object is placed in front of an emitter that is capable of displaying structured patterns (a plasma screen or a CRT monitor). In our setup we use a CRT monitor (figure 2). A series of illumination patterns is emitted and the resulting illumination of each pattern on the object is captured by means of a digital camera.

In our setup we opt for emitting wavelets as illumination patterns (section 4.1). When emitting these wavelet patterns we observe that not all cause an equal level of illumination on the scene. This is due to the properties of the scene and the locality of the wavelets in both the time and the frequency domain. Patterns that cause a great level of illumination are considered to be more important for the environment matte construction process. During acquisition we capture important patterns first. This enables us to stop the acquisition process prematurely when the contribution of the patterns to the illumination is below some threshold or when the acquisition time has exceeded a time-limit. We use a feedback loop to determine the next wavelet pattern which is important for the illumination of the scene.

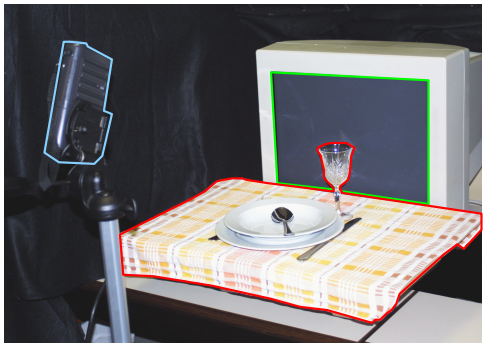

Figure 2: *The scene is highlighted in red, the camera in blue and the emitter in green.*

We use an error-tree (section 4.4) as a tool to determine which wavelet pattern is important. During the feedback loop this error-tree is constructed and refined with information obtained from newly recorded photographs. Each node in the error-tree contains information on how much an emitted wavelet pattern contributes to the received illumination from the scene. Using a tree-like structure to organize wavelet patterns is a natural choice since wavelets form a hierarchical basis. An overview of the recording process can be seen in figure 3.

Compositing an image using a novel backdrop is done by simply decomposing the backdrop image into wavelet coefficients and summing the recorded photographs weighted with the corresponding wavelet coefficients.

3. Mathematical framework

The environment matting equation presented by Zongker et al.[17] is well suited to represent specular and glossy reflections:

$$\mathbf{C} = \mathbf{F} + \sum_{i=1}^{n} \int R_i \mathscr{L}(\mathbf{B}, A_i), \tag{1}$$

where $\mathbf{C}$ is the composited image and $\mathbf{F}$ represents the ambient illumination. The reflection coefficient R_i denotes the amount of light reflected from an area of support A_i. $\mathscr{L}$ is a

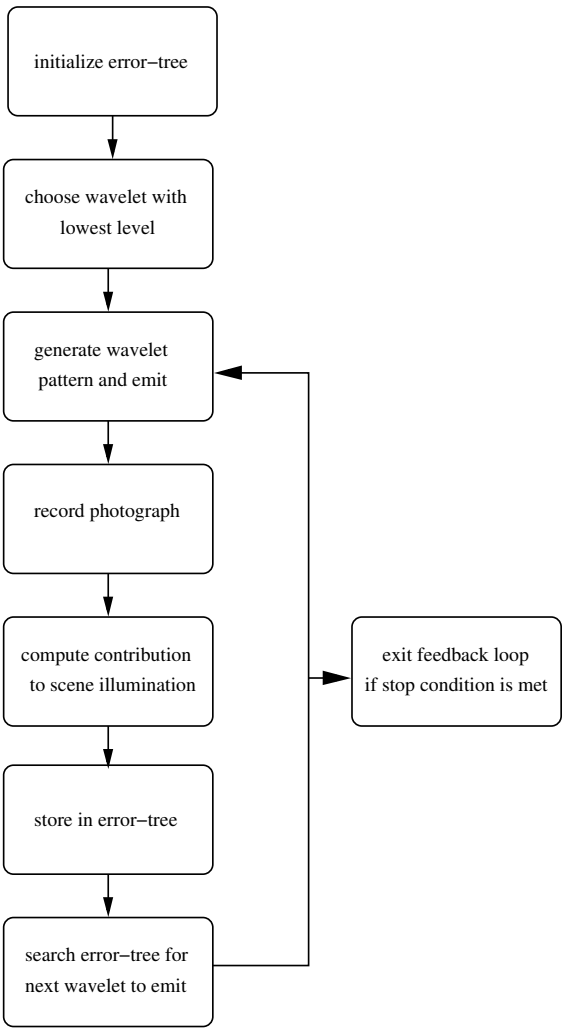

Figure 3: *An overview of the wavelet environment matting algorithm.*

normalized filter defined over A_i on the backdrop image **B**. In the implementation of Zongker et al. $\mathscr{L}$ was chosen as a box filter over a rectangular support A_i. Their implementation included a single reflection coefficient and support ($n = 1$) for the backdrop (n is set to 3 in case the side-drops are used). Chuang et al.[1] improved upon this by using multiple Gaussian filters for $\mathscr{L}$ and an elliptical support area A_i. Choosing more complex filters will not solve the problem of representing diffuse materials since the large area of support of these materials is irregularly shaped and more dependent on the scene properties (e.g. self-shadowing). Therefore, we use a more general mathematical description of the environment matte.

The resulting composite image can be seen as a collection of N pixels, stacked in a $N \times 1$ vector **C**. The M pixels in the backdrop image can also be stacked in a $M \times 1$ vector **B**. The matting process itself can now be written as:

$$\mathbf{C} = \mathbf{L}\,\mathbf{B} + \mathbf{F}, \tag{2}$$

where **L** is a $(N \times M)$ transfer matrix that represents the light transport from the background **B** through the scene into the camera. **L** is solely dependent on the characteristics of the scene. We assume that the effects on the illumination from the remainder of the environment is an invariable $N \times 1$ vector **F**. **F** is called the ambient illumination or foreground illumination. We assume **F** to be known* and therefore will act as if this term is zero for the remainder of this exposition. We will denote $\mathbf{C}(\psi)$ as the observed photograph of the scene illuminated by a pattern ψ.

Formula 2 is a more general mathematical notation of the environment matte, which encloses previous representations. Looking back at the classic matting equation[14], $\mathbf{C} = \mathbf{I}_\alpha \mathbf{B}$, one can see that it approximates **L** by a diagonal matrix of α-values or transparency values. Thus each pixel of the camera image is affected by only one backdrop pixel. The environment matting equation as presented by Zongker et al. (equation 1) expresses **L** in a clever and compact way. The matrix **L** is sparse for specular materials and each pixel is only affected by a localized area on **B**. This can be sufficiently approximated by a filter operation $\mathscr{L}$ with a limited number of parameters on **B**.

We now observe that the background image **B** can be written as a linear combination of M basis images $\mathbf{B_i}$:

$$\mathbf{B} = \sum_{i=1}^{M} a_i \mathbf{B_i},$$

where a_i are the weights or coefficients associated with each $\mathbf{B_i}$. Using formula 2 we can write **C** as:

$$\begin{aligned} \mathbf{C} &= \mathbf{L}\,\mathbf{B} \\ &= \mathbf{L}\Big(\sum_{i=1}^{M} a_i \mathbf{B_i}\Big) \\ &= \sum_{i=1}^{M} a_i (\mathbf{L}\,\mathbf{B_i}) \\ &= \sum_{i=1}^{M} a_i \mathbf{C_i}. \end{aligned}$$

The vectors $\mathbf{C_i}$ are therefore a set of M basis images of the composite image **C** (note that this basis is not necessary compact). A direct result of formula 2 is that each $\mathbf{C_i}$ can be

* **F** can be easily found by setting $\mathbf{B} = 0$ in formula 2.

measured by emitting $\mathbf{B_i}$ onto the scene, since $\mathbf{C_i} = \mathbf{L}\ \mathbf{B_i}$. This is an interesting result, since it implies that we do not need to know the exact form of the transfer matrix $\mathbf{L}$.

To illustrate, assume we have a novel backdrop $\mathbf{B'}$ to composite. We can decompose $\mathbf{B'}$ into the basis images $\mathbf{B_i}$ by projecting $\mathbf{B'}$ onto each dual basis* image $\hat{\mathbf{B}}_\mathbf{i}$ resulting in the coefficients a'_i:

$$a'_i = \langle\ \mathbf{B'} \mid \hat{\mathbf{B}}_\mathbf{i}\ \rangle.$$

The final composite image $\mathbf{C'}$ is then:

$$\mathbf{C'} = \sum_{i=1}^{M} a'_i \mathbf{C_i}.$$

The number of basis images $\mathbf{B_i}$ required to represent $\mathbf{B}$ and consequently the number of $\mathbf{C_i}$ to observe, is enormous. Backdrop images typically have resolutions of $2^{10} \times 2^{10}$ which results in a space of dimension 2^{20}. If each photograph $\mathbf{C_i}$ would take one second then the recording the complete set of basis vectors would last approximately 12 days. Also, assuming an equal resolution for the camera image and the backdrop image would require to store $2^{(10+10+10+10)} = 2^{40}$ pixels!

In the next section we investigate wavelet patterns as a set of basis vectors for $\mathbf{B}$ and try to exploit their hierarchical nature to efficiently handle this large dimensionality.

4. Wavelets and the error-tree

Wavelets are a class of multilevel basis vectors, best known for their applications in image compression. A very useful property is the local support in *both* the time domain —in this case the primal image dimension— and the frequency domain. For more information on wavelets we refer the interested reader to the extensive literature available on this subject (e.g. Stollnitz et al.[15]).

In this section we will motivate the use of wavelets for $\mathbf{B_i}$ (section 4.1). In section 4.2 we argue that the principles used in image compression can also be used in our wavelet environment matting framework (section 4.3). Finally in section 4.4 we introduce the error-tree, which is used to decide which subsequent wavelet pattern is most important for the construction of the environment matte.

4.1. Effects of scene characteristics on the environment matte

It is important to consider the properties of the scene when choosing a specific set of basis vectors as $\mathbf{B_i}$. Ramamoorthi and Hanrahan[13] showed that (unoccluded) diffuse materials act as a low pass filter for incoming illumination. This makes it possible to represent the effects of the incident illumination on diffuse materials with a limited number of coefficients in the frequency domain. For capturing unoccluded diffuse reflections, this implies that a good choice for $\mathbf{B_i}$ should be local in the frequency domain in order to minimize the number of required basis vectors.

On the other hand, previous environment matting methods showed that specular reflections can be compactly represented by a small support area on the backdrop. A compact support area on the backdrop implies locality in the time domain (and hence a non-compact footprint in the frequency domain). Thus for specular materials a good choice for $\mathbf{B_i}$ should be local in the time domain.

Representing both cases with equal ease requires a set of basis vectors that is local in both domains, which leads us to wavelets.

For clarity we will use the Haar wavelet to demonstrate our method, but it can be used with any type of wavelet. The effects of using other wavelets are discussed in section 6. In this paper we will assume that all wavelets are normalized to a DC (low frequency) and Nyquist (high frequency) gain of one.

4.2. Wavelets for image compression

Capturing all possible basis images $\mathbf{C_i}$ is not feasible, when the resolution of $\mathbf{B}$ is large. To overcome this problem we turn to techniques presented in (lossy) image compression literature. In general, an image $\mathbf{I}$ is decomposed into a set of basis vectors $\mathbf{I_i}$ (e.g. using Fourier series, DCT or wavelets), resulting in a set of corresponding coefficients w_i:

$$\mathbf{I} = \sum_i w_i \mathbf{I_i}.$$

Not all weights w_i are equally large. Large weights indicate that the associated basis vector $\mathbf{I_i}$ contributes more to the image $\mathbf{I}$. An approximation $\mathbf{I'}$ of the original image $\mathbf{I}$ can be created by:

$$\mathbf{I} \approx \mathbf{I'} = \sum_i w'_i \mathbf{I_i},$$

where:

$$\begin{aligned} w'_i &= w_i \quad if\ w_i > t \\ &= 0 \quad otherwise. \end{aligned}$$

* $\hat{\mathbf{B}}_\mathbf{i}$ and $\mathbf{B_j}$ are a dual basis iff $\forall i, j : \langle\ \hat{\mathbf{B}}_\mathbf{i} \mid \mathbf{B_j}\ \rangle = \delta_{i,j}$. If $\mathbf{B_j}$ is an orthogonal set of basis vectors then $\hat{\mathbf{B}}_\mathbf{j} = \mathbf{B_j}$.

The threshold t determines which weights are considered important enough for the image reconstruction. Of course leaving out weighted basis vectors introduces an error.

Wavelets have interesting properties that make them very well suited for image compression. First of all, wavelets form a hierarchical basis, which means that the coefficients can be easily sorted into a tree-structure (where the depth of a node in the tree equals the level of the wavelet). DeVore et al.[5] noted that for natural images (i.e. photographs of real scenes) these coefficients decay, and that this decay is dependent on the level j or resolution of the wavelet, the local order of continuity l of the image, and the number of dual vanishing moments* d of the wavelet used:

$$decay \sim 2^{-j\,\mathbf{max}(l,d)}. \tag{3}$$

In the section 4.3, we will explore how we can use this decay of wavelet coefficients in our environment matting setup.

DeVore et al. also noted that if the root of a branch in the coefficient-tree has a low value, then the probability is high that all other coefficients in that branch are also low. In section 4.4 we investigate a similar property in our environment matting method.

4.3. Wavelets for environment matting

We apply the knowledge of the previous section to the environment matting setup in order to reduce the number of photographs $\mathbf{C_i}$ to be recorded.

To begin, define the L_p-norm on an image $\mathbf{I}$ as:

$$L_p(\mathbf{I}) = (\sum_{x,y} |\ \mathbf{pixel_I}(x,y)\ |^p)^{\frac{1}{p}}, \tag{4}$$

where p is usually set to 1 or 2. We can state that the norm $L_p(\mathbf{C_i})$ is an indication of the importance of the emitted illumination pattern $\mathbf{B_i}$, and we can use this norm to sort $\mathbf{C_i}$ in order of importance and record only the important ones.

A major difference between image compression and our environment matting setting is that the backdrop images $\mathbf{B}$, that will be used during compositing, are unknown. The coefficients a_i of the wavelet decomposition of $\mathbf{B}$ in basis images $\mathbf{B_i}$ are thus also unknown.

Formula 3 gives the rate of decay for coefficients a_i, assuming $\mathbf{B_i}$'s to be wavelet patterns and if $\mathbf{B}$ is a natural image. Suppose a_i is a coefficient of a wavelet at level j_i, then we can use:

$$weight(j_i) = 2^{-j_i \times s}, \tag{5}$$

as an upper bound for the coefficients a_i, where s is a constant indicating the general smoothness of the wavelet patterns and the backdrops used. We use $s = 1$ in our examples, but if it is a priori known that the backdrop images and the wavelets are smooth, then a larger s could be chosen. Selecting a larger constant s favors wavelet patterns with low level (low frequency wavelet patterns) over patterns with a high level (high frequency wavelet patterns).

Combining equations 2, 4 and 5 gives us:

$$\begin{aligned}
L_p(\mathbf{C}) &= L_p(\mathbf{LB}) \\
&= L_p(\sum_i a_i(\mathbf{LB_i})) \\
&= L_p(\sum_i a_i \mathbf{C_i}) \\
&\leq \sum_i L_p(a_i \mathbf{C_i}) \\
&\leq \sum_i L_p(weight(j_i) \times \mathbf{C_i}) \\
&\leq \sum_i weight(j_i) \times L_p(\mathbf{C_i}) \\
&\leq \sum_i W_i(\mathbf{C_i}).
\end{aligned}$$

We denote $weight(j_i) \times L_p(\mathbf{C_i})$ as $W_i(\mathbf{C_i})$. Thus we can bound the norm of $\mathbf{C}$ by the sum of $W_i(\mathbf{C_i})$ (being the result of emitting a wavelet patterns $\mathbf{B_i}$).

To apply the same principle as in image compression we need to sort $W_i(\mathbf{C_i})$ and only emit the important $\mathbf{B_i}$ (i.e. with large value for $W_i(\mathbf{C_i})$). The problem is that we do not know the norms $L_p(\mathbf{C_i})$ in advance and hence do not know the order of importance.

4.4. The error-tree

As mentioned in section 4.3, large wavelet coefficients tend to cluster together in a tree-like manner. Low coefficient values in the root of a branch usually indicates low coefficient values throughout the entire branch. A similar observation can be made in our environment matting setup with respect to the weighted norms $W_i(\mathbf{C_i})$.

In this paper we opt for a progressive refining algorithm to find the order of importance of $W_i(\mathbf{C_i})$ for the construction of the environment matte.

In order to do this we define an error-tree. Each node in the error-tree contains a $W_i(\mathbf{C_i})$, in which j_i (formula 5) is equal to the depth of the node. The error-tree is constructed in a top-down manner. Suppose we already have the n most important weighted norms measured and stored in the error-tree. All of the leaf nodes in this error-tree are candidates for

* Dual vanishing moments: the order of polynomials that can be approximated by the *dual* scaling functions of the wavelet.

starting a new branch in the error-tree. We can now state that the leaf node with the largest node-value will probably be the root of a new (not yet measured) branch of the error-tree which is important for the construction of the environment matte.

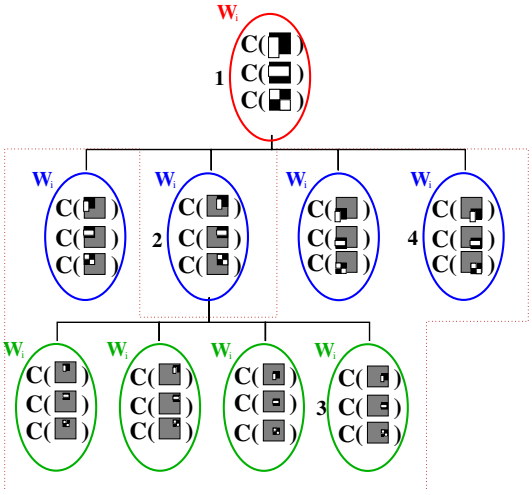

Figure 4: *An example of an error-tree. First the red encircled photographs* $\mathbf{C}(\mathbf{B_i})$ *are recorded and their weighted norm is computed and stored in the error-tree. Next, the error-tree is searched for the largest weighted norm, in this case being (1) and its wavelet dilations (blue encircled) are emitted and recorded. Again the weighted norm is computed and stored. The leaf nodes are searched (blue nodes) for the largest weighted norm. Suppose this is (2), then the green encircled* $\mathbf{C}(\mathbf{B_i})$ *are recorded and their weighted norm computed. Now all the leaf-nodes (within the dark-red dotted line) are searched for the largest weighted norm. If this is (3), then its wavelet dilations are emitted. Note that if the largest weighted norm were (4) then the depth-first order is broken.*

We start up the error-tree by emitting the coarsest wavelet patterns first, the root of the whole error-tree, and storing their weighted norm $W_i(\mathbf{C_i})$. It will be obvious that the dilations of these wavelet patterns will be selected next as patterns to emit. After emitting these wavelet dilations of the coarsest wavelet patterns, the error-tree is searched for the largest leaf-node and the dilations of the wavelets in this node are selected as the next wavelet patterns to emit.

This procedure is repeated until some stop-criterion is met. Note that since we work with 2D wavelet patterns, we have 3 different wavelets per location and level ($\psi_x\varphi_y$, $\varphi_x\psi_y$ and $\psi_x\psi_y$, where φ is the scaling function and ψ the wavelet function and the index denotes the axis on which the function is defined). Therefore we store the weighted norm $W_i(\mathbf{C}(\psi_\mathbf{x}\varphi_\mathbf{y})) + W_i(\mathbf{C}(\varphi_\mathbf{x}\psi_\mathbf{y})) + W_i(\mathbf{C}(\psi_\mathbf{x}\psi_\mathbf{y}))$ in each node in the error-tree. An example of an error-tree is depicted in figure 4.

The error-tree itself is limited in depth by defining a maximal resolution possible for a backdrop, because this limits the maximum level of the wavelet patters used.

The use of the error-tree enables the feedback loop to estimate which subsequent wavelet patterns will contribute most to the illumination on the scene. The feedback loop will ensure that the error-tree is constantly refined by adding newly acquired information, increasing the accuracy of the estimate.

5. Practical considerations

Some practical considerations have to be accounted for, before we can implement the method discussed in the previous section. Our setup consists of a scene placed in front of a CRT monitor. (Note that any emitter capable of displaying structured patterns can be used.) A digital camera is used to capture the effects of emitting wavelet patterns onto the scene.

5.1. Emitting and capturing wavelet patterns

High dynamic range photographs are needed to capture the environment matte, because of the differences in dynamic range between the reflectance of specular and diffuse materials. The camera response curve for the digital camera is obtained by the method introduced by Debevec and Malik[4] and each recorded photograph is converted to a high dynamic range image using this camera response curve. Multiple photographs with different shutter times are recorded if the dynamic range of the scene is too large to be captured with only one photograph.

The dynamic range of a wavelet pattern usually does not fit within the range of the emitter, nor is the range of the emitter linear in radiance space. Scaling the wavelet patterns solves the first problem, whereas calibrating the emitter solves the second.

We need to inversely apply the gamma curve of the emitter to transform the non-linear range of the emitter to a linear range in radiance space. Measuring the gamma curve of the emitter is done simular to Chuang et al.[1], where the average radiance emitted by solid patterns with different intensities is measured and a gamma curve is fitted through the acquired data.

Also, wavelets have positive and negative values. Therefore we need to map each wavelet pattern to a completely positive range, since emitting negative light is not possible. Lets assume that the scaled wavelet ψ has a range of $[-1,+1]$ and the range of the calibrated emitter is linear in radiance space $[0,1]$. Two mappings are possible:

1. translating the wavelet and scaling it: $\psi' = \frac{\psi+1}{2}$,
2. splitting it into two patterns $\psi_\mathbf{p}$ and $\psi_\mathbf{n}$ which contain respectively the positive and negative part of ψ.

We obtain the resulting photograph $\mathbf{C}(\psi)$ from the captured data as follows:

1. $\mathbf{C}(\psi) = 2\mathbf{C}(\psi') - \mathbf{C}(\mathbf{W})$, where $\mathbf{W} = 1$ is a solid white pattern.
2. or respectively $\mathbf{C}(\psi) = \mathbf{C}(\psi_\mathbf{p}) - \mathbf{C}(\psi_\mathbf{n})$.

We have chosen for the latter approach, because of the fact that exposing a CRT monitor a long period with the same color introduces significant extra noise caused by the afterglow from these pixels. An advantage of the second approach is that the dynamic range of an emitted wavelet is doubled at the cost of an extra photograph for each pattern.

5.2. The feedback loop

The feedback loop consists out of taking photographs of the scene lit by different wavelet patterns. The norm of the high dynamic range photographs ($W_i(\mathbf{C_i})$) is used to refine the error-tree. In our implementation we choose for the squared norm L_2, since it weights low radiance values (which are more susceptible to noise) less then the L_1-norm.

5.3. Directly visible pixels

Directly visible backdrop pixels from the emitter should not be included in the computation of $W_i(\mathbf{C_i})$. An alpha-matte is computed in order to exclude these direct visible elements. This alpha-matte is constructed using the method proposed by Zongker et al. The overhead of recording these extra photographs of progressively finer stripe patterns is minimal. Uncovered pixels in the matte should be replaced in the final image by the correct pixels from the backdrop. An image warp of the backdrop image should be computed since we do not require that the camera is perpendicular with respect to the emitter. The image warp can be easily computed using the information in the recorded reference stripe patterns (i.e. without the scene in front of the emitter) used to compute the alpha-matte. It is possible to construct a warped grid representing the image warp by using an edge detection algorithm on the highest resolution horizontal and vertical reference photographs. Each grid line has a known relative position on the backdrop. Using the lowest resolution horizontal and vertical stripe reference image, it is possible to absolutely determine which of the grid lines is in the middle and thus identifying each grid line absolutely.

6. Discussion and results

Our environment matting method can handle diffuse surfaces, as can be seen in figure 5 where four different colored cubes are placed upon a diffuse surface. We used approximately 400 wavelet patterns for the depicted scene. Note the shadows which form high-frequent edges in the environment matte. These edges are hard to represent using smooth filters like elliptical Gaussian filters.

In section 4.1 we noted that other wavelets then the Haar wavelet could be used. Of special interest are biorthogonal wavelets* (e.g. Daubechies (9,7) wavelet). These wavelets result in a smoother approximation, as can be seen in figure 6, because they are smoother in shape. Using these smooth wavelets gives pleasing results if the number of photographs is very limited with respect to the resolution of the backdrop, opposed to the Haar wavelet which gives blocky results. The advantages of these smoother wavelets becomes less obvious when the number of recorded wavelet patterns is increased. The reason is that smoother wavelets require more coefficients to represent high frequency details, and thus require more photographs of high resolution wavelets to represent these fine details. This number of (high resolution) wavelet patterns quadruples with every increase in level.

In figure 7, a glass candy jar filled with little candy bears is depicted. The environment matte is captured using 2400 photographs (or 1200 wavelet patterns split in a negative and positive part). Different backdrops are applied to the scene. The smaller pictures on the right how the result would look like after respectively 100, 300, 600 and 900 emitted wavelet patterns. To give a better idea about the process we did not replace directly visible pixels by the correct pixels from the backdrop image, nor did we show the ambient illumination.

There exists an interesting relation between the L_p-norm of $\mathbf{C_i}$ and $\mathbf{B_i}$:

$$L_p(\mathbf{C_i}) \leq L_p(\mathbf{B_i}).$$

This formula is a direct result of the fact that a material cannot reflect more light then it receives. This is an important observation since it means that the error on $\mathbf{C}$ is bounded by the error on $\mathbf{B}$ from approximating it using a limited number of $\mathbf{B_i}$. This is an upper-boundary for the error on $\mathbf{C}$ and is, in general, an overestimation of the real error. It also implies that increasing the number of emitted $\mathbf{B_i}$ will have a positive effect on the error of $\mathbf{C}$, and in the limit this error will vanish.

Our method is significantly different from previous environment matting methods. It does not rely on non-linear optimization procedures to minimize error, instead it uses a feedback loop to instantaneously process the recorded images. A theoretical comparison between Chuang et al.[1] and the presented method results in some interesting conclusions:

1. Previous environment matting methods[17, 1] result in visually pleasing images for specular and glossy materials, with a fixed number of photographs. The relation between the number of photographs and the error on the composite image is not clear. Our method can control the

* Biorthogonal wavelets: the wavelet and scaling functions of the composition are crosswise orthogonal with the (dual) wavelet and scaling functions of the decomposition.

error and number of photographs more closely. The number of photographs can be adjusted to bound the error and visa versa.
2. Previous environment matting methods utilize a clever brute force attack with respect to the number of photographs to be recorded, which does not take into account the characteristics of the scene, except for the assumption that material properties range from specular to glossy. It cannot represent diffuse materials. The feedback loop in the presented method directs the recording process. The method decides on previously recorded photographs which subsequent wavelet pattern is to be emitted and thus implicitly uses the scene characteristics. It is possible to capture scenes with all kinds of material properties with the presented method.
3. Previous environment matting methods, however, have less storage requirements since they do not require each recorded photograph to be stored, whereas the presented methods requires that each $\mathbf{C_i}$ is stored.

The time to converge to a visual pleasing solution is in general short. Large specular objects, however, can slow the convergence since a large amount of high resolution wavelet patterns have to be emitted and recorded to fully capture these effects.

We used a time limit (12 hours for each scene) as the stop criterion in the feedback loop. The total recording time could be improved by using an optimized wavelet generator (currently the bottleneck in our implementation) and a better synchronization between the digital camera (Canon EOS D30) and the feedback loop. A digital video camera could reduce the time to capture an environment matte even more. Each environment matte requires an average of 2.5GB to store all photographs (RLE compressed). Using more advanced compression algorithms (e.g. JPEG2000) could reduce the required storage even more.

7. Conclusion and future work

In this paper we presented a novel environment matting technique. The method uses wavelets as illumination patterns. An error-tree is constructed during the recording process by means of a feedback loop. Using this feedback loop the contribution of each wavelet pattern on the illumination of the scene is recorded and stored in the error-tree. The feedback loop adapts the recording process automatically to the characteristics of the scene by optimally choosing the next wavelet pattern to emit. Our method can handle scenes composed of any material and requires minimal user interaction. Looking back at the objectives stated in section 1.3 we can see that:

1. The developed method can handle any kind of material properties. Large areas of highly specular materials are problematic due to the slow convergence rate, but are still possible to capture. In natural scenes this situation usually does not occur often.
2. The feedback loop directs the recording process. Knowledge of the scene is accumulated during the recording process itself and is used to minimize work or error.
3. Using a feedback loop implies minimal user intervention which is limited to choosing which kind of wavelet pattern and stating a stop criterion.
4. In section 6 we showed that the approximation error can be bounded.
5. Original environment matting papers required an optimization procedure per pixel. The presented approach uses the idea of linearly weighting and combining basis images, which requires minimal post-processing since the recorded images from the feedback loop can be directly used. The idea of linearly combining basis images is more elegant and easier to implement than non-linear optimization procedures.

Future work includes solving the problem of the slow convergence for large specular objects. This could be solved by using a hybrid solution of the developed method in which upper wavelet resolution is bounded at a fairly low resolution and followed by a classical environment matting step to capture the effect from specular materials. This would ensure the correct capture of diffuse and glossy materials and faster capturing of specular materials.

Other future work includes investigating the effects of different wavelet patterns on the convergence rate and on the approximation error. Wavelets could also be used as a filter in a classical environment matting setup, paving the way for a more elegant hybrid solution.

Better heuristics need to be developed to create a more intelligent stop-criterion. Such a stop-criterion could decide to stop the recording process if the remaining $W_i(\mathbf{C_i})$ falls below some threshold. A head to head comparison with other environment matting methods (e.g. Chuang et al.) can give a better idea when the presented method is preferred and when perhaps a less accurate, but possibly faster non-linear environment matting method is required.

Finally, the presented method could be extended to a fully fledged relighting method by placing the object in a closed cube of emitters and replacing the concept of backdrop images by environment maps.

Acknowledgments

We would like to thank Eyetronics for letting us use their digital camera when ours broke down. Furthermore we would like to thank Jo Simoens and Evelyne Vanraes for answering questions related to wavelets. We also like to thank the people in our research group: Karl vom Berge, Frederik Anrys, Ares Lagae, Bart Adams and Vincent Masselus for proofreading and especially Frank Suykens for his invaluable help. A big “thank you” to the reviewers for their helpful and constructive suggestions. The first author would also

like to thank Saskia Mordijck and Yves D. Willems for believing in me.

Finally we would like to thank Murphy for being right when it comes to things going wrong. This work was partially supported by K.U.Leuven Grant #OT/01-34.

References

1. Yung-Yu Chuang, Douglas E. Zongker, Joel Hindorff, Brian Curless, David H. Salesin, and Richard Szeliski. Environment matting extensions: Towards higher accuracy and real-time capture. In Kurt Akeley, editor, *SIGGRAPH 2000, Computer Graphics Proceedings*, Annual Conference Series. Addison Wesley, 2000.

2. Brian Curless and Marc Levoy. Better optical triangulation through spacetime analysis. In *IEEE International Conference on Computer Vision*, pages 987–994, 1995.

3. Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In Kurt Akeley, editor, *SIGGRAPH 2000, Computer Graphics Proceedings*, Annual Conference Series, pages 145–156. ACM SIGGRAPH, Addison Wesley, July 2000.

4. Paul Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In Turner Whitted, editor, *SIGGRAPH 97, Conference Graphics Proceedings*, Annual Conference Series, pages 369–378. ACM SIGGRAPH, Addison Wesley, August 1997.

5. Ronald A. DeVore, Bjorn Jawerth, and Bradley J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory*, 38(2):719–746, 1992.

6. Tim Hawkins, Jonathan Cohen, and Paul Debevec. A photometric approach to digitizing cultural artifacts. In *In 2nd International Symposium on Virtual Reality, Archaeology, and Cultural Heritage, Glyfada, Greece, November 2001.*, 2001.

7. Takeo Kanade, Andrew Gruss, and L. Carley. A very fast vlsi rangefinder. In *IEEE International Conference on Robotics and Automation*, pages 1322–1329, 1991.

8. Vincent Masselus, Philip Dutré, and Frederik Anrys. The free-form light stage. In *Rendering Techniques EG 2002*, Annual Conference Series. EG, 2002.

9. Wojciech Matusik, Hanspeter Pfister, Addy Ngan, Paul Beardsley, Remo Ziegler, and Leonard McMillan. Image-based 3D photography using opacity hulls. In *SIGGRAPH 2002 Conference Proceedings*, Annual Conference Series, pages 427–437. ACM SIGGRAPH, 2002.

10. Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard McMillan. Acquisition and rendering of transparant and refractive objects. In *Rendering Techniques EG 2002*, Annual Conference Series. EG, 2002.

11. Jeffry Nimeroff, Eero Simoncelli, and Julie Dorsey. Efficient re-rendering of naturally illuminated environments. In *Eurographics Rendering Workshop 1994*, Darmstadt, Germany, June 1994. EG, Springer-Verlag.

12. Thomas Porter and Tom Duff. Compositing digital images. In Hank Christiansen, editor, *Computer Graphics (SIGGRAPH '84 Proceedings)*, volume 18, pages 253–259, July 1984.

13. Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In Eugene Fiume, editor, *SIGGRAPH 2001, Computer Graphics Proceedings*, Annual Conference Series, pages 497–500, 2001.

14. Alvy Ray Smith and James F. Blinn. Blue screen matting. *Computer Graphics*, 30(Annual Conference Series):259–268, 1996.

15. Eric J. Stollnitz, Tony D. DeRose, and David H. Salesin. *Wavelets for computer graphics: theory and applications*. Morgan Kaufmann Publishers, Inc., 1996.

16. Yonatna Wexler, Andrew W. Fitzgibbon, and Andrew Zisserman. Image-based environment matting. In *Rendering Techniques EG 2002*, Annual Conference Series. EG, 2002.

17. Douglas E. Zongker, Dawn M. Werner, Brian Curless, and David H. Salesin. Environment matting and compositing. In Alyn Rockwood, editor, *SIGGRAPH 1999, Computer Graphics Proceedings*, Annual Conference Series, pages 205–214, Los Angeles, 1999. ACM SIGGRAPH, Addison Wesley.

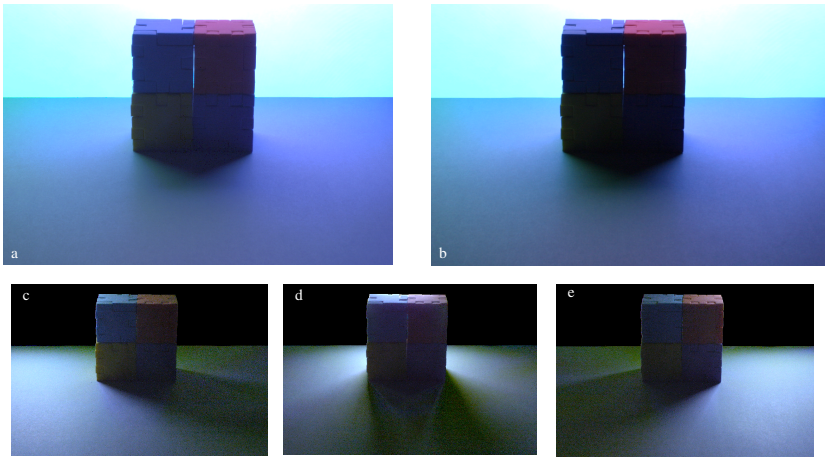

Figure 5: *A scene containing colored cubes placed on a diffuse surface. The scene, composited with a low frequency plasma backdrop, is shown in figure b. A reference image is shown in figure a. In figure c, d and f the same scene is composited with different backdrops containing a white square at different locations (respectively located on the left, middle and right).*

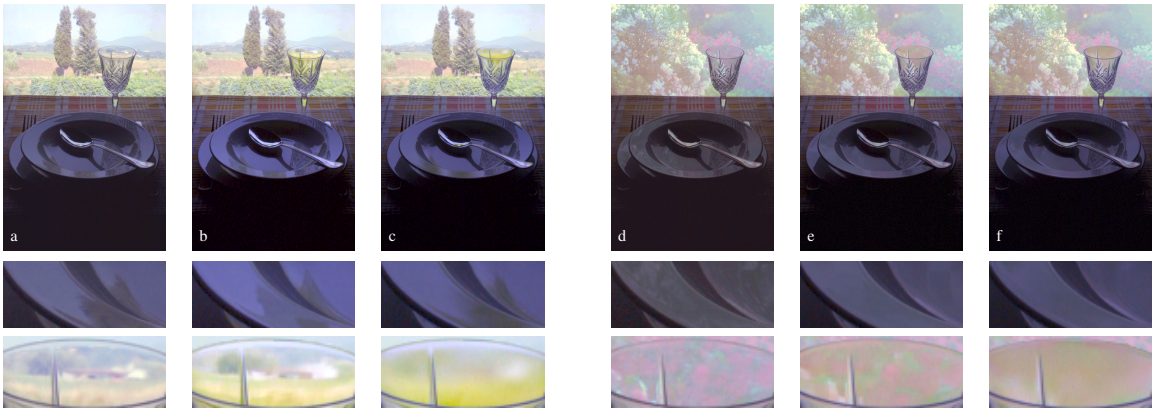

Figure 6: *A dinner scene composited with two different backdrops. Figure a and d show the reference images. Figure b and e are captured (and composited) using 1000 Haar wavelet patterns. Figure c and f are captured using 1000 Daubechies (9,7) wavelet patterns. Details of a part of the plate and the glass are shown for each figure. Note that the colors do not completely match due to a slight calibration error during color correction.*

Figure 7: *A scene containing a glass candy jar filled with little candy bears, composited with two different backdrops. Figure a and g show the reference images, whereas figures b and h were captured (and composited) with 1200 Haar wavelet patterns. On the right is the same scene shown without foreground illumination or without direct visible pixels replaced. One can see the effects of compositing with 100 (c and i), 300 (d and j), 600 (e and k) and 900 (f and l) basis images. Note that the colors do not completely match due to a slight calibration error during color correction.*