Adaptive Grid Optical Tomography

Ivo Ihrke

Graphics - Optics - Vision

Max - Planck - Institut für Informatik

66 123 Saarbrücken, Germany

Marcus Magnor

Graphics - Optics - Vision

Max - Planck - Institut für Informatik

66 123 Saarbrücken, Germany

Abstract

Image-based modeling of semi-transparent, dynamic phenomena is a challenging task. We present an optical tomography method that uses an adaptive grid for the reconstruction of a three-dimensional density function from its projections. The proposed method is applied to reconstruct thin smoke and flames volumetrically from synchronized multi-video recordings. Our adaptive reconstruction algorithm computes a time-varying volumetric model, that enables the photorealistical rendering of the recorded phenomena from arbitrary viewpoints. In contrast to previous approaches we sample the underlying unknown, three-dimensional density function adaptively which enables us to achieve a higher effective resolution of the reconstructed models.

Categories and Subject Descriptors (according to ACM CCS):
I.4.5 [Image Processing and Computer Vision]: ReconstructionI.4.8 [Image Processing and Computer Vision]: Scene AnalysisShapeI.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

Certain dynamic phenomena in nature are challenging to model realistically [FSJ01, NFJ02, ZWF*03]. We propose to use real images of such phenomena to obtain computer models that are suitable for photorealistic image synthesis. In [IM04] optical tomography is introduced as a suitable method to reconstruct volumetric phenomena from camera images. The method is applied to reconstruct three-dimensional volumetric models of flames. This article presents an advancement of this scheme that is more memory-efficient which allows increasing the accuracy of the reconstruction. Besides flames, we reconstruct another class of volumetric phenomena, i.e. thin smoke.

Image-based modeling of transparent phenomena has received only little attention in computer vision. There have been approaches to extend surface reconstruction by taking transparency into account [BV99, SG98]. Computerized tomographic methods have been applied to rigid body reconstructions [GW99]. Transparent, volumetric phenomena are treated by Hasinoff et al. [HK03, Has02]. While the similarity of the reconstruction problem to computerized tomography (CT) is pointed out, the CT method is deemed not applicable in the sparse view case. In Ref. [HK03] the *flame sheet decomposition* algorithm is developed, which recon-
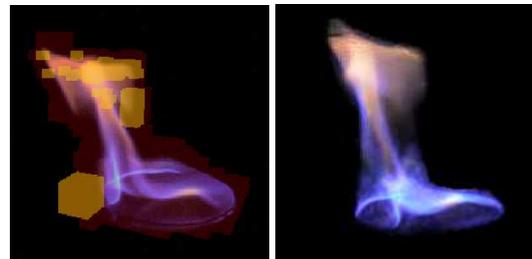

Figure 1: *The adaptive reconstruction process to the left, reconstruction result to the right.*

structs a surface (the flame sheet) with varying transparency and color. Ihrke and Magnor [IM04] use 3D - CT reconstruction to generate time-varying volumetric models of flames.

In contrast, this paper presents an adaptive grid computerized tomography technique that has the advantage of requiring less memory and allowing for higher resolution reconstructions. It uses an octree data structure [Sam90] to manage the inhomogenous set of basis functions. The closest work to this are multigrid CT reconstruction [LMM93, HLMR96] and wavelet-based multi-resolution tomogra-

phy [BKW96,RFLBW97,SY93]. The latter works use multi-resolution techniques to restrict the number of X-ray exposures [RFLBW97], to supress noise in smooth regions of the reconstructed image [SY93] and to establish an error bound on the reconstructed image [BKW96]. In the CT literature the term adaptive computerized tomography is applied to methods that perform the data acquisition process adaptively e.g. [HFO04]. Octree-based adaptive methods have also been successfully used in fluid dynamic simulations to increase the effective resolution of the simulation [LGF04].

The paper is organized as follows: Section 2 reviews the uniform grid algorithm and its underlying assumptions. A discussion of the applicability to the reconstruction of smoke follows. In Section 3 we derive the adaptive algorithm. Section 4 presents experiments and results. In Section 5 we conclude the paper and present directions for future work.

2. Review of the basic Algorithm

In this section we review the CT algorithm presented in [IM04]. In this paper a basic computerized tomography (CT) technique is described, that is based on an algebraic formulation of the inversion of a simplified image formation model for fire. We discuss its applicability to smoke in Section 2.4.

2.1. Image Formation Model

Hasinoff et. al. [HK03] present a simplified image formation model for fire. The fire is modeled as a 3D density field ϕ of fire reaction products i.e. soot particles. Image intensity is related to the density of luminous particles in the fire. The model has the form

$$I_p = \int_c \phi \, ds + I_{bg}. \tag{1}$$

Here I_p is pixel p's intensity, c a curve through 3D space, ϕ is the density field of soot particles and I_{bg} is the background intensity. Curve c is the backprojected ray of pixel p. We approximate every pixel by one ray through the density field. The underlying assumptions of this simplified model are

- Negligible absorption/scattering - this assumption is valid for fire not substantially obscured by smoke, and
- Proportional self-emission - the brightness depends on the density of the soot particles only

2.2. Mathematical Derivation

In order to invert (1) we have to make an assumption on the structure of ϕ. We do this by assuming that ϕ can be represented as a linear combination of basis functions ϕ_i:

$$I_p = \int_c \left(\sum_i a_i \phi_i \right) ds + I_{bg} \tag{2}$$

The sum and the coefficients a_i can be moved out of the integral and we get

$$I_p = \sum_i a_i \left(\int_c \phi_i \, ds \right) + I_{bg}. \tag{3}$$

Eq. (3) describes a linear system of equations,

$$\mathbf{p} = \mathbf{Sa} + \mathbf{b} \tag{4}$$

The rows represent the equations for one pixel and the columns contain the integrals of the pixel's backprojected rays over the basis function ϕ_i. The choice of the basis functions ϕ_i is essential for the tractability of the problem. The box basis function

$$\phi_i^{Box}(x,y,z) = \begin{cases} 1 & \begin{array}{l} x_{min}^i < x \le x_{max}^i \\ y_{min}^i < y \le y_{max}^i \\ z_{min}^i < z \le z_{max}^i \end{array} \\ 0 & \text{else} \end{cases} \tag{5}$$

is a popular choice because of its simplicity. This is especially true in an adaptive setting. Furthermore it is non-negative which ensures a non-negative density field if a non-negative solution to Eq. (4) is found. This is necessary to ensure a physically plausible reconstruction.

2.3. Implementation Issues

The system of equations (4) is generated in a similar way to volume raytracing [Max95]. We generate the complete system of equations, that means we incorporate one equation for every pixel in all camera images that contains all basis functions.

For a particular frame of a multi-video sequence we remove columns of the matrix in Eq. (4) corresponding to basis functions with non-zero support completely outside the visual hull [Lau94]. This effectively sets their coefficients to zero and makes the sparse view reconstruction process possible. This process is described in detail in section 3.4. The resulting smaller linear system is solved in a least squares fashion:

$$\mathbf{a} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T(\mathbf{p} - \mathbf{b}) \tag{6}$$

The inversion is carried out using the CGLS variant [Han98] of the conjugate gradient method. To obtain a non-negative solution a projection to the subspace of non-negative solutions is carried out in every iteration. This is done by setting all negative entries of the solution vector $\mathbf{a}$ to zero.

2.4. Application to Smoke

To apply the presented method to smoke we have to make sure the image formation model is more or less valid. Both assumptions stated in section 2.1 are obviously not true for smoke in general. We tackle the problem by making the following assumptions

- The smoke is uniformly and diffusely lit, and
- Scattering takes place in a uniform manner .

These assumptions make it possible to treat the smoke as a self-emissive medium. We found this model to be applicable for thin smoke reconstruction.
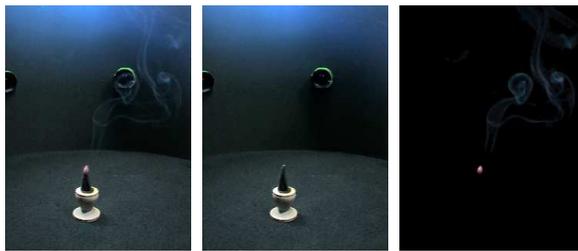
(a) Original   (b) Background   (c) Foreground

Figure 2: *Background subtraction for thin smoke: The original image (a) contains a burning incense that produces thin smoke. It is diffusely lit with daylight. (b) is a background image recorded beforehand and (c) is the background subtracted image that contains only the smoke column.*

In the case of flame reconstruction it is possible to record in a dark setting, avoiding the complication of background subtraction for transparent phenomena, i.e. vector **b** of Eq. (4) is zero. Since the smoke has to be lit uniformly we have to perform background subtraction which involves some image processing. We record a sequence of background images and compute the median background image as well as the standard deviation for every pixel. This allows us to classify pixels into foreground and background in the smoke sequences. For pixels classified as foreground we subtract the backgound value which corresponds to the additive nature of the image formation model. All other pixels are set to zero. This step computes the difference $(\mathbf{p} - \mathbf{b})$ in advance and is depicted in Fig. 2. In the following we regard vector **p** as the background-subtracted pixel vector.

In [IM04] it is argued that image processing might be potentially dangerous for the reconstruction process because it corresponds to an image filter in the 3D domain

$$\hat{\mathbf{p}} = \mathbf{S}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{p} \qquad (7)$$

and the preprocessing of the images might not correspond to a valid filter of this form. This follows from

Eqs. (4) and (6). Fortunately, background subtraction and noise reduction are not observed to have adverse effects on the visual quality of the reconstruction.

3. Adaptive reconstruction

3.1. Motivation

An adaptive reconstruction technique for three-dimensional computerized tomography is motivated by

- better memory efficiency, and
- better regularization properties .

Both improvements are achieved by using less basis functions than in the uniform subdivision case. Using a uniform grid, the memory limit of 2 GB on a 32-bit machine is reached relatively fast. In our experiments we have found that using 8 cameras with images taken at 320x240 pixel resolution, we can achieve a reconstruction resolution of 128^3 voxels.

In general we have n_p rows with $O(\sqrt[3]{n_b})$ non-zero elements each in matrix **S**, where n_b is the number of basis functions needed to approximate the density field and n_p the number of pixels in a frame (a frame is one time frame of a multi-video sequence and contains n_c images, where n_c is the number of cameras). A uniform discretization of the reconstruction space is assumed. We use an index-stored sparse matrix as a representation for **S**. We store the two indices and the matrix value for each non-zero entry. Even though this is the most memory efficient storage scheme for an unstructured sparse matrix, the available memory fills up quite fast when using a uniform grid.

Better regularization is achieved by representing large regions with a uniform density by only a few basis functions. This reduces the number of degrees of freedom in that region.

3.2. Basic Iteration

An adaptive reconstruction algorithm has to proceed iteratively.

1. Estimate the coefficients **a** of Eq. (4), then
2. Project the residual error, i.e. image plane error onto the basis functions to get a measure where to split, followed by
3. Splitting of the k basis functions that are responsible for the largest error. This is implemented as an augmentation of matrix **S**.
4. If the estimate is not good enough we go back to step 1.

The following subsections cover the single steps in detail. The estimation process (step 1) has already been described in Section 2.2. It is independent of the shape of the basis functions and therefore directly applicable to the adaptive algorithm.

3.3. Error Projection

The main difficulty in the adaptive estimation process is to relate the residual error

$$\mathbf{r} = \mathbf{p} - \mathbf{Sa} \qquad (8)$$

to the interpolation error

$$|u - P_{\phi_i}u| \qquad (9)$$

Here u is the perfectly reconstructed function and $P_{\phi_i}u$ its projection onto the subspace of functions representable by the basis functions ϕ_i. A relation between the two errors allows for the identification of the coefficients $\mathbf{a_i}$ that contribute most to the residual error. We present a heuristic for this projection step and show the feasability of an adaptive computerized tomography reconstruction.

The basic idea of our heuristic is based on the projection of the basis function's regions of support into the camera images, and on the accumulation of the residual error of the affected pixels. This yields an intuitive way of relating the error caused by a particular basis function to the residual error in the image plane. This error measure is efficiently implemented using sparse matrix - vector multiplications.

Our main observation is that the complete geometry of the problem is encoded in the matrix $\mathbf{S}$. The system of equations (4) has the following structure:

$$\begin{pmatrix} \mathbf{p_1} \\ \mathbf{p_2} \\ \vdots \\ \mathbf{p_{n_p}} \end{pmatrix} = \begin{pmatrix} \int_{c_1}\phi_1 ds & \dots & \int_{c_1}\phi_{n_b}ds \\ \int_{c_2}\phi_1 ds & \dots & \int_{c_2}\phi_{n_b}ds \\ \vdots & \vdots & \vdots \\ \int_{c_{n_p}}\phi_1 ds & \dots & \int_{c_{n_p}}\phi_{n_b}ds \end{pmatrix} \mathbf{a} \qquad (10)$$

n_p is the number of pixels and n_b denotes the number of basis functions used to represent the solution. Since the basis functions ϕ_i have local support the matrix $\mathbf{S}$ is sparsely populated. Note that every column of the matrix corresponds to a particular basis function. The rows are the equations for one particular pixel. Therefore an entry $\mathbf{S_{ij}}$ is non-zero only if the support of basis function ϕ_j projects onto pixel $\mathbf{p_i}$. We use this observation to formulate the projection of the basis functions into the images and the accumulation of residual errors per basis function in matrix notation.

$$\mathbf{e}_{\phi_j} = \frac{1}{n_{\phi_j}} \sum_i \left[(\mathbf{S_{ij}} \neq 0)?1:0 \right] |\mathbf{r_i}| \qquad (11)$$

The basis function ϕ_j is visible in n_{ϕ_j} cameras, and vector $\mathbf{e}$ contains the error measure for all basis functions. This computation requires either a specialized function in the implementation of the algorithm or a second copy of matrix $\mathbf{S}$ with all non-zero entries set to one. The specialized function approach might not be possible if the linear algebra package is closed source as, for example, MATLAB is. The copied matrix approach, on the other hand requires twice the amount of main memory. Therefore it is advisable, to find a better measure that does not require matrix $\mathbf{S}$ to be changed. By using the coefficients of matrix $\mathbf{S}$ directly we can incorporate a weight corresponding to the pathlength of the back-projected ray c_i that is affected by basis function ϕ_j. This in itself is not sufficient to capture the influence of basis function ϕ_j on the error in pixel $\mathbf{p_i}$, because the coefficient $\mathbf{a_j}$ might scale basis function ϕ_j in an arbitrary way. A better version is therefore

$$\mathbf{e}_{\phi_j} = \frac{\mathbf{a_j}}{n_{\phi_j}} \sum_i \mathbf{S_{ij}}|\mathbf{r_i}|. \qquad (12)$$

3.4. Splitting and Basis Function Independent Visual Hull Restriction

Using the error heuristic from the previous section, we iteratively split those k basis functions that cause the largest residual error. k is an arbitrary number that influences the convergence of the adaptive scheme. We perform a uniform splitting on the box basis functions. For different types of basis functions the splitting process becomes more complicated. E.g. for the linear basis function, asymmetric basis functions have to be introduced where basis functions of different splitting level overlap.

To incorporate the splitting into matrix $\mathbf{S}$ it does not have to be recomputed completely. Rather the columns corresponding to the split basis functions are removed and replaced by columns corresponding to the new basis functions. This, together with the error heuristic expressed in matrix form results in an efficient implementation of the iterative method described in Section 3.2.

The visual hull restriction of matrix $\mathbf{S}$ can be performed in an efficient and accurate way. The idea is shown in Fig. 3. In step 1 we extract all rows that have non-zero entries in pixel vector $\mathbf{p}$. These represent the rays that are inside the silhouette of camera c. In step 2 we identify the columns that have zero entries only, i.e. the basis functions that do not affect the silhoutte in camera c. Therefore they cannot be part of the visual hull. This step has to be performed per camera. Therefore it is necessary to keep track where the pixels in vector $\mathbf{p}$ originated. We compute a binary vector for each camera, marking all basis functions that are potentially contained in the visual hull. These correspond to non-zero columns in the submatrices extracted in step 1. The basis functions marked with one have non-zero support in the generalized cone backprojected from the silhouette of camera c.

By taking the elementwise logical AND of all binary vectors, we compute the intersection of the generalized cones
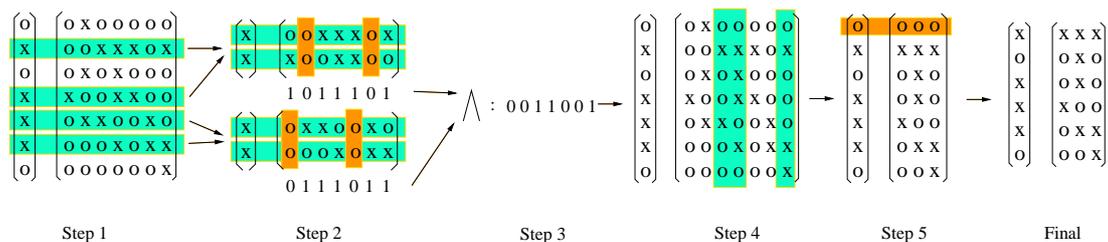
Figure 3: *Visual hull restriction of matrix* **S**.

of all cameras and thus the visual hull. This computation is accurate up to the discretization in the image plane, i.e. up to the pixel level. Step 4 restricts the original matrix **S** to columns corresponding to basis functions that have non-zero support in the visual hull. The resulting linear system has zero rows for some of the pixels in vector **p** that are outside the silhouette. Note that not all rows that have a zero right hand side get removed. This is because the basis functions might not be completely contained in the visual hull. Therefore the zero pixels have to be accurately included in the estimation process. Basis functions on the boundary of the visual hull are very likely to be split, so the boundary of the visual hull gets represented accurately after some iterations of the adaptive scheme.

The splitting process cannot be performed infinitely. It is advisable to set a maximum split level. A suitable criterion is the Nyquist limit, i.e. if a basis function projects to less than two pixels in all images the splitting can be stopped. A useful number for k is the square root of the number of basis functions currently used. This choice results in a sub-exponential growth in the number of basis functions but converges faster than a constant number k.

3.5. Implementation

The whole adaptive process can be efficiently implemented using basic matrix - vector operations. A simple indexed, unordered sparse matrix representation has been found to be suitable for the purposes of this algorithm. This allows for a straight-forward implementation and is also easily parallelizable. The memory requirements are typically only 20 to 25% of the uniform grid case while achieving comparable reconstruction accuracy. This allows for higher resolution input images and a higher resolution of the reconstructed model.

We use a minimalistic octree data structure to keep track of the splitting process. For each column index of matrix **S** we store the split level from the root of the tree and the index that the corresponding leaf would have in a uniformly split octree under a fixed order of traversal. This enables us to identify the positions and sizes of the newly inserted basis functions and to generate the integral values for the new columns of matrix **S** in every iteration.
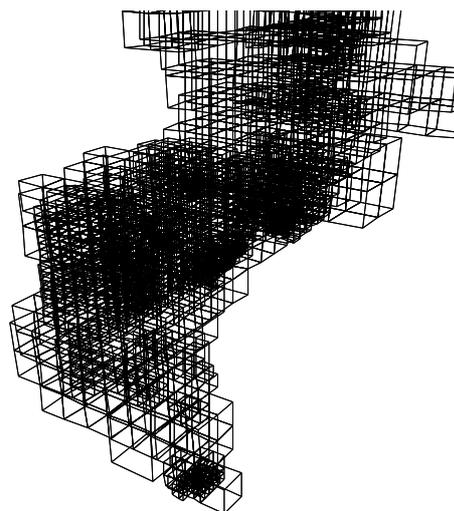

Figure 4: *A visual hull restricted adaptive grid (intermediate step in the iteration).*

4. Results

For our experiments we use a calibrated multi-camera setup with 8 cameras. We record multi-video sequences at a resolution of 320x240 pixels with 15 frames per second. In the case of smoke we perform the background subtraction as a preprocessing step. Because the background subtraction is not perfect, we use an alpha-matte in step 1 of the visual hull restriction process Fig. 3. The matte is created using morphological operators on the thresholded foreground images. The fire sequences are recorded in the dark and do not need to be preprocessed. We performed a convergence study of the adaptive algorithm. The results are shown in Fig. 5. The residual error decreases as expected with the number of iterations. The decrease is not monotonic though. This is because our error measure is based on a heuristic. An interesting graph is the plot of the number of rows in matrix **S** versus the number of iterations. It shows that the basis functions adapt to the pixel perfect visual hull. A visualization of the results after different numbers of iteration and the convergence of the solution is shown in Fig. 7. Along with Figs. 8
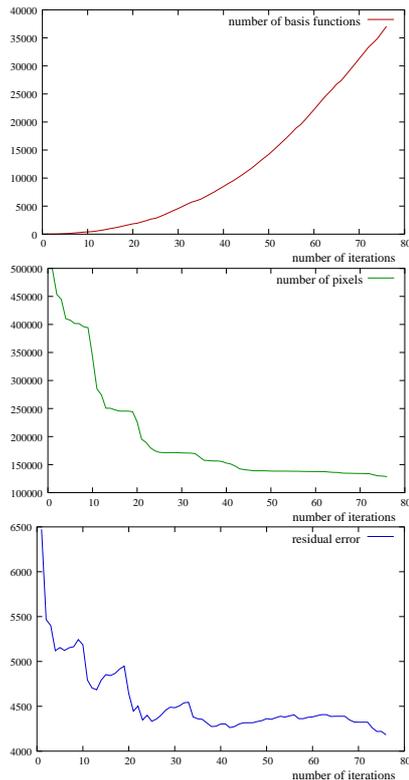
Figure 5: *Behaviour of n_b the number of visual hull consistent basis functions (columns) in matrix **S** (top), n_p the number of pixels (rows) in matrix **S** (middle) and the residual error (bottom) versus the number of iterations of the adaptive algorithm.*

and 6 it shows reconstructions we have obtained by applying our algorithm to different multi-video sequences.

5. Conclusions and Future Work

We have presented an adaptive algorithm for optical tomography. The algorithm is based on an octree hierarchy of piecewise constant basis functions. We propose a heuristic that enables the projection of errors in the image plane into the domain of the basis functions. This allows us to iteratively split basis functions that cause large residual errors in the image plane. Using this algorithm we are able to reconstruct dynamic, volumetric models of flames and thin smoke. Additionally we presented an efficient scheme for the accurate computation of the visual hull. This scheme is independent of the choice of basis functions and accurate up to the pixel level.

We believe that our adaptive tomography algorithm is applicable to other tomography problems as well. We would like to test it on real X-ray data to measure its performance

compared to more traditional methods like filtered backprojection. Future work includes the derivation of a mathematically sound error projection as well as the use of different classes of basis functions. Wavelet bases provide an interesting option but it is more difficult to ensure a nonnegative density field because the basis functions are not non-negative. This research would also provide a connection to wavelet-based multi-resolution schemes [BKW96, RFLBW97, SY93].

Further experiments regarding the combination of fire and smoke in recorded sequences should be conducted. Modeling occluding objects in flames and smoke is another possibility.

References

[BKW96]   BATHIA M., KARL W. C., WILLSKY A. S.: A Wavelet-Based Method for Multiscale Tomographic Reconstruction. *Transactions on Medical Imaging 15*, 1 (Feb. 1996), 92–101.

[BV99]   BONET J. S. D., VIOLA P. A.: Roxels: Responsibility Weighted 3D Volume Reconstruction. In *Proc. International Conference on Computer Vision (ICCV '99)* (1999), pp. 418 – 425.

[FSJ01]   FEDKIW R., STAM J., JENSEN H. W.: Visual Simulation of Smoke. *Proceedings of SIGGRAPH* (August 2001), 15–22.

[GW99]   GERING D. T., WELLS III W. M.: Object Modeling using Tomography and Photography. In *Proc. of IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes* (June 1999), pp. 11–18.

[Han98]   HANSEN P. C.: *Rank-Deficient and Discrete Ill-Posed Problems*. Society of Industrial and Applied Mathematics, 1998.

[Has02]   HASINOFF S. W.: Three-Dimensional Reconstruction of Fire from Images. MSc Thesis, University of Toronto, Department of Computer Science, 2002.

[HFO04]   HOFMANN J., FLISCH A., OBRIST A.: Adaptive CT scanning - mesh based optimisation methods for industrial X-ray computed tomography applications. *NDT&E Internaitonal 37* (2004), 271–278.

[HK03]   HASINOFF S. W., KUTULAKOS K. N.: Photo-Consistent 3D Fire by Flame-Sheet Decomposition. In *In Proc. 9th IEEE International Conference on Computer Vision (ICCV '03)* (2003), pp. 1184 – 1191.

[HLMR96]   HENSON V., LIMBER M., MCCORMICK S., ROBINSON B.: Multilevel image reconstruction with natural pixels. *SIAM J. Sci. Comp. 17* (1996), 193–216.

[IM04]   IHRKE I., MAGNOR M.: Image-Based Tomographic Reconstruction of Flames . *ACM Siggraph / Eurographics Symposium Proceedings, Symposium on Computer Animation* (June 2004), 367–375.

[Lau94] LAURENTINI A.: The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Recognition 16*, 2 (Feb. 1994), 150–162.

[LGF04] LOSASSO F., GIBOU F., FEDKIW R.: Simulation of Water and Smoke with an Octree Data Structure. *ACM Transactions on Graphics 23*, 3 (August 2004), 457–462.

[LMM93] LIMBER M. A., MANTEUFFEL T. A., MC-CORMICK S. F.: Optimal Resolution in Maximum Entropy Image Reconstruction from Projections with Multigrid Acceleration. In *Proceedings of the Sixth Annual Copper Mountain Conference on Multigrid Methods* (1993).

[Max95] MAX N.: Optical Models for Direct Volume Rendering. *IEEE Transactions on Visualization and Computer Graphics 1*, 2 (June 1995), 99–108.

[NFJ02] NGUYEN D. Q., FEDKIW R., JENSEN H. W.: Physically Based Modelling and Animation of Fire. *ACM Transactions on Graphics 21*, 3 (July 2002), 721–728.

[RFLBW97] RASHID-FARROKHI F., LIU K. J. R., BERENSTEIN C. A., WALNUT D.: Wavelet-Based Multiresolution Local Tomography. *Transactions on Image Processing 10*, 6 (Oct. 1997), 1412–1430.

[Sam90] SAMET H.: *The Design and Implementation of Spatial Data Structures*. Addison - Wesley Publishing Company Inc., 1990.

[SG98] SZELISKI R., GOLLAND P.: Stereo Matching with Transparency and Matting. In *Proceedings Sixth International Converence on Computer Vision* (1998), pp. 517–524.

[SY93] SAHINER B., YAGLE A. E.: Image Reconstruction from Projections under Wavelet Constraints. *Transactions on Signal Processing 41*, 12 (Dec. 1993), 3579–3584.

[ZWF*03] ZHAO Y., WEI X., FAN Z., KAUFMAN A., QIN H.: Voxels on Fire. In *Proc. 14th IEEE Visualization Conference (VIS'03)* (October 2003), pp. 271–278.

Figure 6: *A volumetric model of smoke rendered from different viewpoints.*

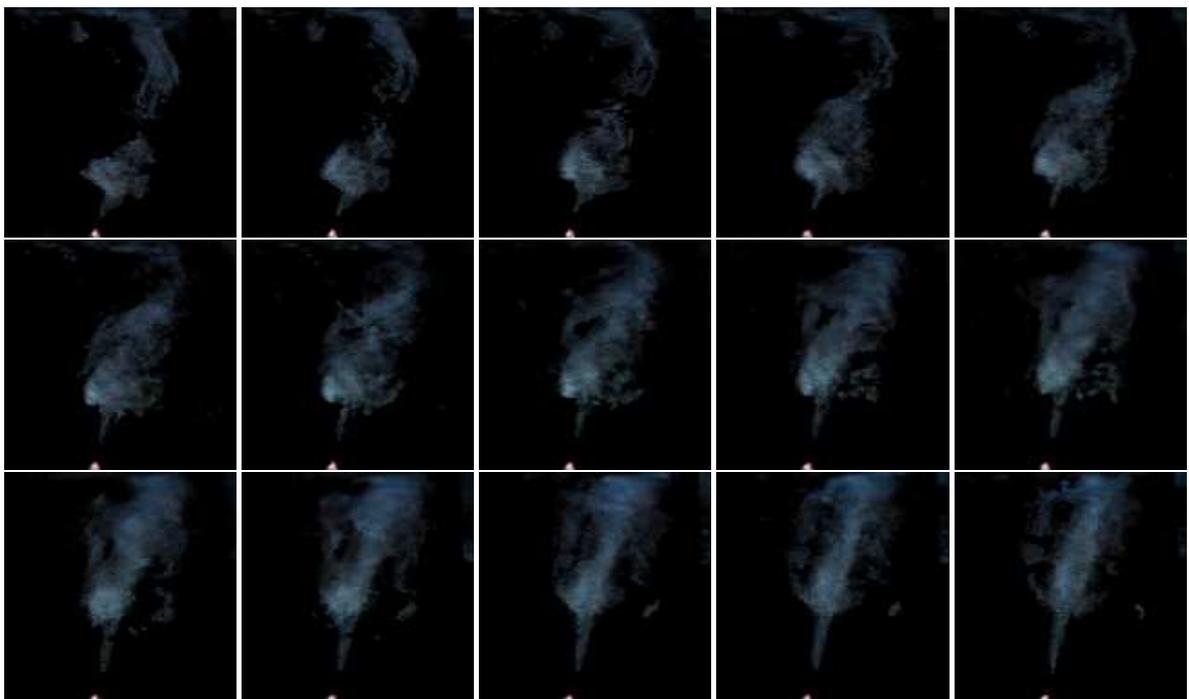**Figure 7:** *Visualization of reconstruction results after 1, 2, 14, 28 and 100 iterations.*

Figure 8: *Reconstructions of 15 consecutive frames of a smoke sequence.*