Modeling Hair from Multiple Views

Yichen Wei[1] Eyal Ofek[2] Long Quan[1] Heung-Yeung Shum[2]

[1]The Hong Kong University of Science and Technology
{yichenw,quan}@cs.ust.hk

[2]Microsoft Research Asia
{eyalofek,hshum}@microsoft.com

Figure 1: From left to right: one of the 40 images captured by a handheld camera under natural conditions; the recovered hair rendered with the recovered diffuse color; a fraction of the longest recovered hair fibers rendered with the recovered diffuse color to show the hair threads; the recovered hair rendered with an artificial constant color.

Abstract

In this paper, we propose a novel image-based approach to model hair geometry from images taken at multiple viewpoints. Unlike previous hair modeling techniques that require intensive user interactions or rely on special capturing setup under controlled illumination conditions, we use a handheld camera to capture hair images under uncontrolled illumination conditions. Our multi-view approach is natural and flexible for capturing. It also provides inherent strong and accurate geometric constraints to recover hair models.

In our approach, the hair fibers are synthesized from local image orientations. Each synthesized fiber segment is validated and optimally triangulated from all visible views. The hair volume and the visibility of synthesized fibers can also be reliably estimated from multiple views. Flexibility of acquisition, little user interaction, and high quality results of recovered complex hair models are the key advantages of our method.

Keywords: Hair modeling, 3D photography

CR Categories: I.3.5 [Computer Graphics]: Object modeling—Modeling packages; I.4.5 [Image Processing and computer vision]: Reconstruction.

1 Introduction

Many computer graphics applications, such as animations, computer games or 3D teleconferencing, require 3D models of people. Hair is one of the most defining characteristics of human appearance, but the capturing of hair geometry remains elusive. The geometry of hair, with hundreds of thousands of thin fibers, is complex and hardly perceivable at normal image resolutions. The intriguing reflectance properties of hair make active scanning methods unreliable. Most prior dedicated modeling tools [Kim and Neumann 2002] required costly and tedious user interaction. The recent work by [Paris et al. 2004; Grabli et al. 2002] demonstrates the potential and possibility of an image-based hair modeling approach combining synthesis and analysis. Impressive results are presented in [Paris et al. 2004]. But the capturing of the images via this approach has to be done in a controlled environment with a known moving light source. This requirement limits its applicability. The subject to be captured must remain still during the entire process of capturing. This process excludes the possibility of capturing a subject in motion. Moreover, this approach cannot be used for existing captured videos.

In this paper, we propose a more practical and natural method of modeling from images of a given subject taken at different viewpoints. A handheld camera or video camera is used under natural lighting conditions without any specific setup. It offers greater flexibilities for image capture; moreover, it may produce more complex and accurate models of sophisticated hair styles. This is mainly due to the inherent strong and accurate geometric constraints from multiple views. Our approach is efficient and involves little user interaction. This makes it a useful solution as a starting point for other interactive modeling systems, thereby reducing the amount of work needed for generation of a final model.

2 Related Work

Active 3D laser scanners cannot handle hair reconstruction due to the complex geometry and reflection properties of hairs. Traditional image-based vision techniques, such as stereo methods or photometric stereo approaches [Faugeras 1999] at best generate a smooth surface boundary of the hair volume without the fiber geometry.

Most existing methods [Gelder and Wilhelms 1997; Kim and Neumann 2002; Chang et al. 2002; Ward and Lin 2003] are highly interactive and take a long time to produce hair models. Nakajima et al. [Nakajima et al. 1998] started to use images for bounding the hair volume, yet the synthesis of the hair fibers is heuristic and does not follow the observed images.

More recently, Grabli et al. [Grabli et al. 2002] and Paris et al. [Paris et al. 2004] pioneered a new image-based hair modeling approach. Note that hair structure consists of *fibers* of a spatial frequency much higher than that can actually be captured by the relatively low image resolution of cameras. This makes the direct recovery of individual fibers ill-posed, even impossible. Yet, it is observed in [Grabli et al. 2002; Paris et al. 2004] that neighboring hair fibers tend to have similar directions, and a group of such closely tied hair fibers, called a *strand*, generates structured and detectable 2D orientation information in the images. The fiber direction in space is first constrained by the local orientation at one viewpoint, and finally determined from the scattering properties [Marschner et al. 2003] observed from multiple images at the same viewpoint with the known light positions.

Paris et al. [Paris et al. 2004] proposed an effective filtering method based on the oriented filters to produce a robust, dense and reliable orientation map at one viewpoint. Good results have been produced by repeating the procedure from several different viewpoints to cover the entire hair. The main disadvantage of the method is that the capturing procedure is rather complex under a controlled environment and a fixed viewpoint approach *per se* restricts the visibility and geometric constraints.

3 Approach

3.1 Overview

Our approach is to use images captured from multiple views to recover the hair geometry. A hand-held camera is used under uncontrolled illumination conditions to capture images from multiple viewpoints. Then, the camera geometry is automatically computed using the method developed in [Lhuillier and Quan 2005]. Next, we detect a local orientation per pixel in each image. Each hair fiber is represented by a sequence of chained line segments. The key step of the reconstruction is triangulation of each fiber segment using image orientations of multiple views. If fibers were observable in images, the recovery would be geometrically equivalent to a 3D line reconstruction from multiple 2D image lines [Faugeras 1999; Quan and Kanade 1997]. However, since they are invisible at normal image resolutions, each fiber segment is first synthesized in space, then validated by orientation consistency from at least three views, and finally optimally triangulated from all its visible views. The multi-view setting naturally allows a robust and accurate estimation of the *hair volume* that bounds the hair synthesis, by computing a visual hull from multiple views.

The major difference between our method and the most relevant work by [Paris et al. 2004] lies in our multiple view approach versus their *per se* single view approach. This results in a different reconstruction method, different capturing techniques and different results. The practicality of the capturing method and the inherent strong 3D geometric information are clear advantages of our multi-view approach over a fixed view one.

3.2 Hair volume definition and determination

The hair volume is the volume bounded by two surfaces, the hair surface, S_{hair}, and the scalp surface, S_{scalp}, as illustrated in the violet area in Figure 2.b. A good estimate of the hair volume is important to bound the synthesis of the hair fibers. Although this is difficult for the fixed viewpoint approach [Paris et al. 2004], it can be reliably estimated through a multiple view approach.

One possible way to compute a surface representation of the subject, can be using the quasi-dense approach developed by [Lhuillier and Quan 2005]. However, to make our approach more general and practical, we approximate the surface of the subject S_{hair} by the visual hull [Laurentini 1994]. The visual hull is usually a good approximation when using images of full turn around the subject and can be automatically computed from the subject's silhouettes in a practical and robust manner [Matusik et al. 2000; Szeliski 1993].

The scalp surface S_{scalp} underneath the hair is invisible, thus unobtainable by cameras. A generic head model or an ellipsoid can be used as an approximation of it. Another good systematic way of approximating the scalp surface is to use an inward offset version of S_{hair}.

The hair area, $H(S)$, of a given surface S can be defined as the union of all patches whose projection onto images is in the hair mask of those images. The extraction of the hair mask is described in Section 4. The hair volume is then approximated to be the volume between $H(S_{hair})$ and $H(S_{scalp})$. The synthesized hair fibers start from a regular distribution over $H(S_{scalp})$. They are terminated when they exceed a pre-defined maximum length or reach $H(S_{hair})$. All these quantities are illustrated in Figure 2.a.

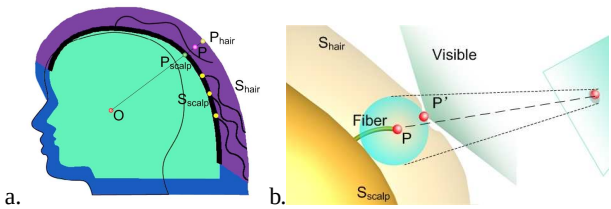

Figure 2: a. Hair volume, marked in violet, is approximated from a visual hull S_{hair} and an inward offset surface S_{scalp} of the visual hull. b. The visibility of a fiber point P is determined from that of the closest point P' on the hair surface. The projection of P onto an image is not a point, but a small image area that is the projection of a ball around P with radius PP'.

3.3 Fiber visibility determination

Given a space point, P, through which a hair fiber is to be synthesize, we first need to determine its visibility with respect to all views.

The visibility of a point on the hair surface S_{hair} is determined by the surface itself and can be stored off-line. However a point P inside the hair volume is obviously invisible from all views. Its visibility is defined to be the same as that of the closest point P' on the hair surface (see Figure 2.b). This may be considered as a convenient heuristic. On the other hand, when the 3D orientation of P is computed, it average the orientation of the small image area that is the projection of the ball defined in Figure 2.b. Intuitively, the observed image information is smoothly diffused from the visible surface to the interior of the hair volume, and the direction of inside invisible fibers is an interpolation of visible fibers on the hair surface. This procedure naturally results in more smoothed hair fibers inside the volume: the further away from the visible hair surface the fibers are, the more smoothed the reconstructed directions of the fibers are.

3.4 Three-view orientation consistency

Let P be a point in space. In each view, the 2D orientation at the projection of P defines a line in the image, **a**, a homogenenous

3-vector. Back-projecting this line into space from the viewpoint gives a plane going through the camera center and the line. The normal vector $\mathbf{n}$ of the plane is $\mathbf{n} = \mathbf{A}^T\mathbf{a}$, where $\mathbf{A}_{3\times3}$ is the 3×3 submatrix of the 3×4 projection matrix $(\mathbf{A}_{3\times3}, \mathbf{t}_{1\times3})$.

If P is a valid fiber point in space, its 3D orientation should be consistent with the multiple 2D orientations obtained on its projections in these views where it is visible. All back projection planes should intersect into the common supporting line of a fiber segment at P. This constraint can be expressed in a minimum of three views, resulting in two independent algebraic constraints. One of them, important to our application, is expressed only in terms of the orientation quantities [Faugeras 1999]: $\mathbf{n} \cdot (\mathbf{n}' \times \mathbf{n}'') = 0$.

Therefore, a space point P is a valid fiber point if the orientations $\mathbf{a}$, $\mathbf{a}'$, and $\mathbf{a}''$ at the projections of P in three views represented by $\mathbf{A}$, $\mathbf{A}'$ and $\mathbf{A}''$ satisfy

$$\mathbf{A}^T\mathbf{a} \cdot (\mathbf{A}'^T\mathbf{a}' \times \mathbf{A}''^T\mathbf{a}'') = 0. \quad (1)$$

This is indeed one of the trilinear constraints [Hartley and Zisserman 2000; Faugeras and Luong 2001] and can be used to efficiently validate the positions of the fiber points in space.

3.5 Triangulation of the 3D orientation of fibers

A hair fiber is a space curve that is approximated by a sequence of chained line segments. The position of one end-point of a fiber segment is synthesized in space, and the direction of the fiber segment is optimally triangulated from all visible views as long as the fiber segment is validated in multiple views.

Let $\mathbf{n}_j$ be the normal vector of the plane π_j defined by the line of orientation $\mathbf{a}_j$ in the jth view. Assume that the direction vector in space of the fiber segment is $\mathbf{d}$. We have $\mathbf{d} \cdot \mathbf{n}_j = 0$. The direction vector $\mathbf{d}$ can therefore be recovered from at least two views by solving this linear homogeneous equation system. When more than two views are available, the direction vector can be optimized over all visible views j as follows. A line through the point P can be parameterized by any other point Q of the line. The position of Q is optimized by minimizing the sum of all distances, $\Sigma d_j(Q, \pi_j)^2$, from Q to each plane π_j. Taking the uncertainty of orientation estimation into account, we solve for $\mathbf{d}$ by minimizing

$$\Sigma_j \frac{1}{\sigma_j^2 ||\mathbf{n}_j||} (\mathbf{n}_j \cdot \mathbf{d})^2, \text{ subject to } ||\mathbf{d}|| = 1, \quad (2)$$

where σ_j is the variance of the response curve in different orientations of the given filter at that position. Its inverse encodes the strength of the detected orientation edge. This is a linear optimization problem that is efficiently solved by singular value decomposition. A small line segment from P to $P \pm \lambda\mathbf{d}$ can be created as a portion of the fiber. The scalar λ is a globally fixed constant such that the segment is of unit length, *i.e.*, the projection of $||\lambda\mathbf{d}||$ onto an image should roughly be the size of a pixel. The sign is determined such that the current segment forms a larger angle with the previous segment to keep the fiber as smooth as possible.

Given the fact that the computed 2D orientations are noisy, we first systematically discard those with low confidence. Second, instead of directly using the algebraic constraint given by Eq. 1 to validate a synthesized fiber point P, we use the average unweighted residual error of Eq. 2, computed from at least three views, $\frac{1}{n}\Sigma_j \frac{(\mathbf{n}_j \cdot \mathbf{d})^2}{||\mathbf{n}_j|| ||\mathbf{d}||}$, for the validation.

4 Implementation

Image capture and camera calibration The image capturing in our approach is conveniently achieved using a hand-held digital camera under natural conditions. Typically, about 30 to 40 images are taken to cover the whole head of the subject. We either move around the subject that remains fixed to make a full turn or turn the subject around itself. This acquisition process does not require any ad hoc setup. Necessary precautions are that a short shutter time should be used to avoid motion blurring of the hair images and the environment should be sufficiently illuminated to make sure that all parts of the subject are visible, particularly with very dark hair areas.

The camera geometry through correspondences and autocalibration is fully automatically recovered using the quasi-dense approach developed in [Lhuillier and Quan 2005]. A more standard sparse approach using the points of interest as described in [Hartley and Zisserman 2000; Faugeras and Luong 2001] may require more images for camera geometry determination.

2D orientation map computation We adapted the filtering method developed in [Paris et al. 2004] to compute a dense orientation map at each pixel for each image. We choose only one oriented filter, a Canny-like first derivative of Gaussian, and apply it at discrete angles for each pixel. The angle at which the filter gives the highest response is then retained for the orientation of that pixel. The *confidence* of the orientation at each pixel is taken as the inverse of the variance of the response curve.

The 2D orientation estimation for each pixel obtained in [Paris et al. 2004] is usually better than the one estimated by us, as the estimate for each pixel is taken to be the most reliable response among *all* images under different lighting conditions for a given viewpoint, while we process only a single image under a fixed natural lighting for each viewpoint. Yet, instead of using the 2D orientation information to directly constrain one degree of freedom of the fiber direction in space, as done by [Paris et al. 2004], we can gain robustness by taking advantage of redundancy in multiple views. A direction of a fiber segment is typically triangulated from about 10 views.

We also found that choosing the best response out of multiple filters instead of using a single one does not significantly improve the quality of the results.

Hair mask extraction An approximated binary *hair mask* for each image is used to determine the hair area of a surface $H(S)$. Silhouettes of the subject are used to compute a visual hull as an approximation of the hair surface, bounding the synthesized hair fibers.

In our current implementation, these masks are drawn manually. This is essentially the only interactive part of our approach, that can be automated as well. For example, the subject silhouettes could be recovered using techniques such as space-carving [Kutulakos and Seitz 1999] for a general case, and chroma keying or background accumulation for a case of static camera and a moving head. Separating the boundary between the hair and face could be achieved by advanced skin tone analysis.

Hair fiber synthesis We fix an object-centered coordinate frame at the center O of the volume bounded by S_{hair} and align it with the vertical direction of S_{hair}. We use spherical coordinates for points on the scalp, and angular coordinates, $\theta \in [0, 180]$ and $\phi \in [0, 360]$, for surface parameterization. We also define a mapping between these surfaces by drawing a line through the center O and intersecting it with a surface at points P, P_{hair}, and P_{scalp}, as illustrated in Figure 2.a. This mapping may not be one to one if one of the surfaces is not star shaped. In that case, we systematically choose the intersection point that is the shortest to the center.

The starting points of hair fibers are uniformly sampled points in the hair area of the scalp surface $H(S_{scalp})$. A point on S_{scalp} is in the hair area $H(S_{scalp})$ if its projection in the most fronto-parallel view lies in the hair mask of that view. It is sufficient to discretize the parameter ϕ by one degree sampling from 0 to 360 degrees, and

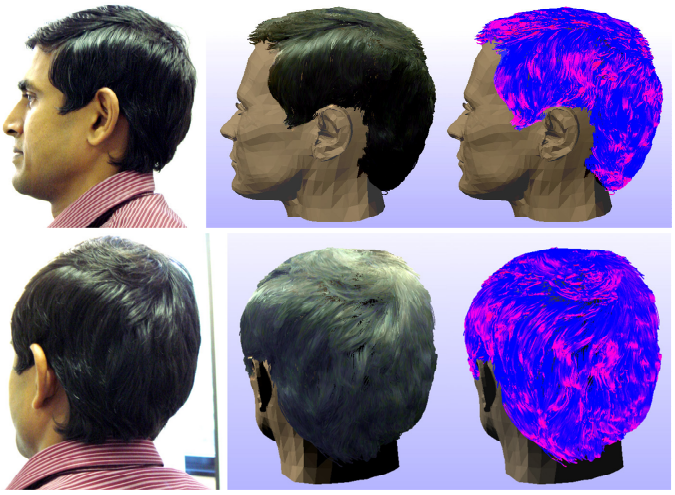

Figure 3: Example of a typical man's short and dark hair. In each row, one of the original images on the left, the recovered hair rendered with the recovered diffuse colors in the middle, and rendered with a curvature map on the right. Middle bottom image's levels are set brighter to better show details.

the parameter θ, pointing to the top of the head with $\theta = 0$, from 0 to 120 degrees at each degree.

Each hair fiber is a sequence of chained small line segments, and each of them roughly corresponds to an imaged pixel size. The segments are generated one by one. The procedure terminates when the total accumulated length of all segments exceeds a pre-defined threshold or the segment reaches the boundary of the hair volume. The fiber segment direction is optimally triangulated from all visible views as described in Section 3.5. Those visible views at extreme viewing angles are discarded to avoid uncertainty due to occlusion and blurring around the silhouettes. To maintain smooth fibers, we avoid sharp changes of growth direction. In such cases, we prefer to keep the growth direction of previous segments. Repeated failure to get a newly triangulated direction will terminate the fiber growth. Finally, the fibers are smoothed where the curvature is too large and the very short ones are discarded.

5 Results

Examples We tested different hair styles and hair colors to demonstrate the flexibility and the accuracy of our hair geometry recovery. For a typical man's example with very dark hair, shown in Figure 3, 38 images of resolution 480×640 are captured in the corner of our lab. For the woman's styles shown in Figures 1, 5 and 6, we used short and long wavy wigs worn on a real person to capture 40 images of resolution 768×1024. We also tested a difficult curly hair style shown in Figure 4. For that, a wig was put on a dummy head on a turntable with a black background to capture 36 images of resolution 1024×768. Additional views of the recovered model can be seen in the accompanying video.

All these results are fully automatically computed except for the extraction of the hair masks, clearly demonstrating the robustness and the quality of the reconstruction with a multi-view approach. Many long strands from the long wavy hair style have been successfully recovered in good order. In the curly example, the recovered density is slightly lower due to the intriguing complex curls.

Hair rendering The recovered hair fibers are rendered as OpenGL lines without anti-aliasing to better visualize the separate fibers. A diffused color for each fiber segment is collected from all its visible views. Then, a median of all sampled colors is retained to remove directional illumination. Yet, same specularity is retained, and visible on the top part of the head, due to light sources on the ceiling. capturing from additional viewpoint directions, such as top views, or modeling of the original illumination structure can be used to get a better estimation of the hair colors, if needed. For display purposes, a generic head model is displayed, roughly scaled to fit the subject head (although some discrepancies can be seen).

To emphasize the structure of the recovered fibers, we use other rendering methods. The *curvature rendering* uses a curvature map in which a redder color encodes a higher curvature. The *sparse rendering* only renders a fraction of the longest recovered fibers so that the hair threads are better illustrated. The long wavy hair style is also rendered, shown in Figure 6, with the self-shadowing algorithm developed in [Bertails et al. 2005].

Running time The image capturing takes only about a few minutes. The registration is automatically done within 5 minutes on a 1.9GHz P4 PC. The orientation map computation is implemented in Matlab and takes a few hours. Extraction of hair masks which could be automatic is currently done interactively for each view. The visual hull computation takes about 1 minute. We typically synthesize about 40K fibers, consisting of 4 to 8 million line segments, in about 15 minutes.

Limitations As we use natural images, we can model only what we can see. Some parts of the hair might be saturated or shadowed in the images, so the orientation cannot be reliably estimated. Some interpolation techniques like inpainting could be used to fill in these areas. The invisible fibers are now synthesized from a lower resolution orientation map or neighboring ones, but there exist multi-layered hair styles in which one layer is completely occluding the others. Since each layer already contains invisible strands, it is impossible to recover the occluded parts of the inner layer. Heuristics or user interactions are necessary for the improvement and enhancement of the recovered hair.

6 Conclusion

We have presented a method for modeling hair from a real person by taking images from multiple viewpoints. It typically recovers a complete hair model for complex hair styles from about 30 to 40 views. The method offers many practical advantages. Images can be captured at any place and any time as long as the subject is sufficiently illuminated. It opens up possibilities for capture of hair dynamics, using a setup of multiple cameras. It is highly automated with little user interaction. The only user interaction in our current implementation is the extraction of hair masks from images that could have been automated with a more advanced image analysis

Figure 4: Example of a difficult curly hair style. One of the original image on the left. The recovered hair rendered with the recovered diffuse colors in the middle, rendered with a curvature map on the right.

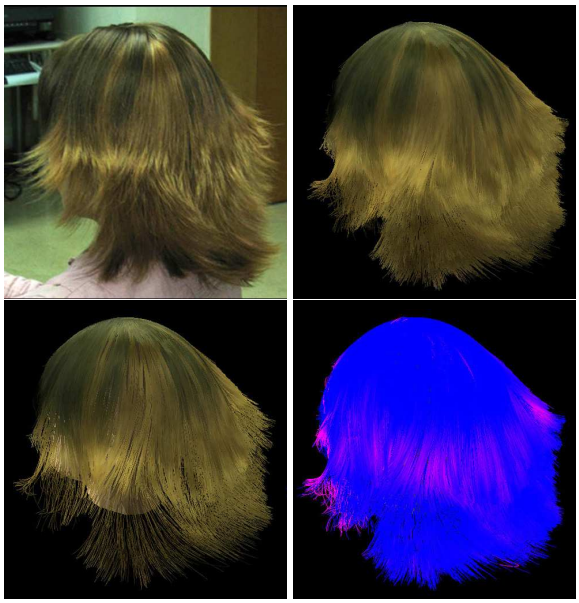

Figure 5: Example of a typical woman's style. One of the original images on the top left. The recovered hair rendered with the recovered diffuse colors on the top right, rendered with a fraction of long fibers on the bottom left, rendered with a curvature map on the bottom right.

Figure 6: Example of a typical long wavy hair style. One of the original images on the top left. The recovered hair rendered with recovered diffuse colors on the top right, rendered with a constant hair color using the self-shadowing algorithm on the bottom left, rendered with the recovered diffuse colors using the self-shadowing algorithm on the bottom right.

method. We will focus our future research directions on the following issues: the immediate extension to dynamic modeling of hair with a multi-camera setting; improvement of orientation filtering in the scale space, particularly for the areas in which no reliable orientation information is available; and the development of simple user interactive tools for enhancement and improvement of the modeled hair style.

Acknowledgements

We would like to thank Florence Bertails for some rendering results, Steve Marschner for his original rendering code, and Maxime Lhuillier, Tsang Kin Ting, and Wang Jingdong for various help. The work was done when Yichen Wei was a visiting student at Microsoft Research Asia. The work was supported by Hong Kong RGC Grant HKUST6182/04E.

References

BERTAILS, F., MÉNIER, C., AND CANI, M.-P. 2005. A practical self-shadowing algorithm for interactive hair animation. In *Proc. Graphics Interface*.

CHANG, J., JIN, J., AND YU, Y. 2002. A practical model for hair mutual interactions. In *Proc. of SIGGRAPH conference*, ACM.

FAUGERAS, O., AND LUONG, Q. 2001. *The geometry of Multiple Images*. The MIT Press, Cambridge.

FAUGERAS, O. 1999. *Three-Dimensional Computer Vision, a Geometric Viewpoint*. The MIT Press, Cambridge.

GELDER, A. V., AND WILHELMS, J. 1997. An interactive fur modeling technique. In *Proc. Graphics Interface*, 181–188.

GRABLI, S., SILLION, F., MARSCHNER, S., AND LENGYEL, J. E. 2002. Image-based hair capture by inverse lighting. In *Proc. Graphics Interface*, 51–58.

HARTLEY, R., AND ZISSERMAN, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

KIM, T.-Y., AND NEUMANN, U. 2002. Interactive multi-resolution hair modeling and editing. In *Proc. of SIGGRAPH conference*, ACM.

KUTULAKOS, K., AND SEITZ, S. 1999. A theory of shape by space carving. In *IEEE ICCV*, 307–314.

LAURENTINI, A. 1994. The visual hull concept for sihouette based image understanding. IEEE *Transactions on Pattern Analysis and Machine Intelligence 2*, 16, 150–162.

LHUILLIER, M., AND QUAN, L. 2005. A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE *Trans. on Pattern Analysis and Machine Intelligence vol 27*, 3, pages 418–433.

MARSCHNER, S., JENSEN, H., CAMMARANO, M., WORLEY, S., AND HANRAHAN, P. 2003. Light scattering from human hair fibers. *ACM Trans. on Graphics 3*, 22, 780–791.

MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S. J., AND MCMILLAN, L. 2000. Image-based visual hulls. In *Proc. of SIGGRAPH conference*, ACM.

NAKAJIMA, M., MING, K., AND H.TAKASHI. 1998. Generation of 3d hair model from multiple pictures. *IEEE Computer Graphics and Application 12*, 169–183.

PARIS, S., BRICENO, H. M., AND SILLION, F. X. 2004. Capture of hair geometry from multiple images. In *Proc. of SIGGRAPH conference*, ACM, 712–719.

QUAN, L., AND KANADE, T. 1997. Affine structure from line correspondences with uncalibrated affine cameras. IEEE *Transactions on Pattern Analysis and Machine Intelligence 19*, 8, 834–845.

SZELISKI, R. 1993. Rapid octree construction from image sequences. *Computer Vision, Graphics and Image Processing 1*, 23–32.

WARD, K., AND LIN, M. 2003. Adaptive grouping and subdivision for simulating hair dynamics. In *Proc. of Pacific Graphics*.