NON-PHOTOREALISTIC RENDERING FROM MULTIPLE IMAGES

Alberto Bartesaghi, Guillermo Sapiro

University of Minnesota
Electrical and Computer Engineering Department
Minneapolis, MN 55455

Tom Malzbender, Dan Gelb

Hewlett-Packard Laboratories
Media and Mobile Systems Lab
Palo Alto, CA 94304

ABSTRACT

A new paradigm for automatic non-photorealistic rendering (NPR) is introduced in this paper. Existing NPR approaches can be categorized in two groups depending on the type of input they use: image based and object based. Using multiple images as input to the NPR scheme, we propose a novel hybrid model that simultaneously uses information from the image and object domains. The benefit not only comes from combining the features of each approach, but most important, it minimizes the need for manual or user assisted tasks in extracting scene features and geometry, as employed in virtually all state-of-the-art NPR approaches. We describe a particular implementation of such an hybrid system and present a number of automatically generated pen-and-ink style drawings. This work then shows how to use and extend well developed techniques in computer vision to address fundamental problems in image representation and rendering.

1. INTRODUCTION

A simple schematic drawing obtained with NPR techniques emphasizes high level or salient perceptual features and at the same time effectively communicates shape and geometry, thus the interest in producing computer generated NPR. In particular, stroke-based drawings can be synthesized by adequately placing individual strokes over the image. It is the combination of two basic illustration principles: *density* and *orientation* of strokes that conveys the desired appearance to the drawings. Varying the stroke density we can represent different textures and lighting conditions. Stroke direction carries the shape information, and orientations that follow principal directions of objects in the scene are known to be appropriate in representing shape [1]. NPR synthesis has then two stages: 1) *Feature and geometry extraction stage*: determine the density-orientation combination of strokes that adequately represents the underlying scene. 2) *Rendering stage*: produce a configuration of strokes that represents the target density-orientation combination.

1.1. Previous Work

NPR algorithms can be categorized according to the type of input they use: image based NPR use 2D photographs as input [2, 3]. As any type of image can be used, there is no restriction in the complexity of models that can be rendered including real scenes, landscapes, etc. On the down side, no geometric information can be inferred from the image which is necessary to convey 3D shape to the drawings. Object based NPR use 3D models as input [4, 5, 6]. These are limited to work on computer generated scenes, and normally no other information but the geometry is available.

NPR algorithms can also be classified according to the degree of user intervention they require, ranging from interactive to fully automatic. The one we propose here belongs to the small yet very important class of fully automatic algorithms. Note that this *automation degree* refers only to the feature and geometry extraction stage, and this is where existing NPR algorithms perform poorly. Interactive image based systems were proposed in [2] and [3]. In [2] the user manually selects a desired *stroke texture* for the different regions in an image, the stroke direction and density being the key elements in defining a texture. In [3] target tone values are obtained from the input image, and interaction techniques are provided for the user to manually edit a direction field that specifies the orientation of stokes. In object based algorithms [6, 5] target tone values are obtained using standard lighting computations and principal directions can easily be computed on the 3D model. These algorithms have then a higher degree of automation, however, unless material or color properties are attached to the mesh we may fail to reproduce important surface features (eyes in faces is one such example).

For the rendering stage two class of methods have been used: the first draws individual strokes in the specified direction until the target density is reached [2, 4, 6, 7]. The second uses pre-defined textures (stroke arrangements representing different tonalities) and a texture synthesis algorithm reproduces those patterns throughout the image [5]. Besides reproducing the essential *tone-orientation* information, salient scene features can also be used to further en-

0-7803-8554-3/04/$20.00 ©2004 IEEE.

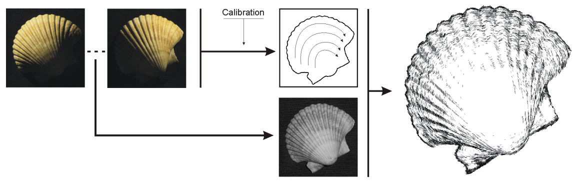

Fig. 1. NPR from multiple images. From the image set we extract principal directions. The rendering uses these orientations and tone values from an image of the set.

hance the renderings. Representing silhouettes, creases, edges, etc. can significantly enhance visual comprehension, as done in [6, 5, 8]. The relation of NPR to halftoning becomes evident, in the sense that both operate on an analog principle: the local density of primitive elements is proportional to the local grayness of the target image. Specifically, the gridless halftoning technique [9] is partially motivated to mimic graphic artwork similar to NPR techniques, and relies on analog image formation principles. By exploring this relation we think both techniques can benefit from one another.

1.2. Overview and Contributions

Using multiple images as input to the NPR scheme, we propose a novel hybrid model that simultaneously uses information from the image and object domains (see Figure 1), this having a number of advantages: 1) target illumination values are obtained directly from image gray levels (greatly reducing geometric modelling tasks), 2) we can *automatically* compute principal directions using multiple image geometry, 3) salient scene features to enhance renderings can easily be extracted from input images using standard image processing techniques, and 4) general scenes can be rendered (not only computer generated objects). Most important, the combination of all these features allows a degree of automatization not achieved by any existing image or object based NPR algorithm. As examples of our hybrid approach, we use input from two multiple-imaging systems: binocular stereo and a multiple-light photometric stereo.

Our implementation of such a system has the following steps: 1) get surface normals and principal directions directly from images and calibration data without intermediate 3D reconstruction (Section 2), 2) use input images as target tone values, and 3) generate the drawings using texture synthesis rendering (Section 3). We present a number of automatically generated examples in Section 4.

Tone values and principal directions may be obtained in many ways from multiple-image systems using well established computer vision techniques. Likewise, the rendering stage can use any existing non-photorealistic depiction or halftoning technique. What is novel and important is the concept of combining both sources of information (2D and 3D) as input for NPR, it's main implications being the increased degree of automation and the possibility to put together the advantages of each approach. This work also shows that those aspects of 3D geometry relevant to NPR can be extracted from multiple-imaging systems without going to the trouble of generating the 3D model.

2. GEOMETRY FROM MULTIPLE IMAGES

The geometry relevant to NPR are principal directions of objects in the scene, consequently, any multiple-image system that allows computation of principal directions can be used as input to our system. Actually, since we get principal directions from the normals, any imaging system allowing computation of normals can also be used. In the binocular stereo setting two photos of the same object are taken from different positions in space. Having a calibrated system, the usual approach is to build the 3D reconstruction of the surface from which normals and principal curvatures will then be derived. Instead, we follow the approach in [10] and use the images directly to compute surface normals without reconstructing the surface. In the photometric stereo setting the camera is held at a fixed position and pictures are taken as we change lighting conditions on the scene. We use the acquisition devices in [11] that allow automatic acquisition of multiple images, each illuminated with an individual strobe light source. In this case, from the intensity images and light source directions we directly obtain normals using a least squares fitting procedure [12].

Principal directions of a surface $\mathcal{S}$ are given by the eigenvectors of its second fundamental form $\mathrm{II}_{\mathcal{S}}$. Suppose the surface of an object in the scene can be represented as the graph of a function f. In this special case and considering the proper parametrization, $\mathrm{II}_{\mathcal{S}}$ can be expressed in terms of the already computed normals and their derivatives. Solving the eigenvalue problem we then get the principal directions. As pointed out in [6], stroke orientation should follow the curvature of the overall shape, whereas fine scene details are expressed through tonal variation. Accordingly, we compute principal directions in a lower resolution grid, smooth the resulting direction field, and then resample it on the original grid. Note that we are interested in smoothing *orientations* (invariant to π rotations), not *directions* (invariant to 2π rotations). For doing so, an adaptation procedure is first applied [13], to then run an anisotropic vector diffusion technique [14] that denoises directions preserving discontinuities.

3. RENDERING

The rendering is done using a priority-based texture synthesis algorithm [15]. We first build a basic set of textures using the *Tonal Art Maps* (TAMs) introduced in [5]. Since

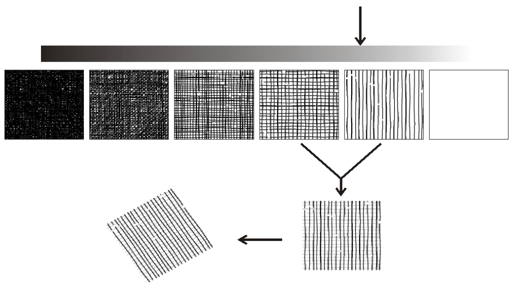

Fig. 2. Pattern formation: first blend two adjacent images from the TAM, then rotate the blended image.

TAMs only represent a discrete number of tones, blending between contiguous images is needed to represent intermediate tone values. For this reason, TAMs are specifically designed to have spatial coherence so the blending between adjacent textures can be done without artifacts. For each pixel in the synthesized image we have a target tone and orientation values, t and θ, respectively. The texture synthesis will use patterns representing these (t, θ) combinations constructed as follows: first blend the two adjacent discrete tone images $\lfloor t \rfloor$ and $\lceil t \rceil$ from the TAM, and then rotate the blended image an angle θ, see Figure 2.

4. DATA SETS AND EXAMPLES

A number of examples are presented in Figure 3. From binocular stereo pairs of real faces we generated (a) and (b). Note how principal directions guide the orientation of strokes so we get a compelling idea of shape and how features like the eyes are accurately reproduced (this could not be done simultaneously with single 2D or 3D existing techniques without significant user intervention). Examples from photometric stereo images are shown in (c) and (d). (a), (b) and (c) were rendered using the algorithm in the previous section, (d) was generated with the INSPIRE rendering system [7] that uses individual placement of strokes.

5. SUMMARY AND DISCUSSION

We presented an automatic system for the generation of non-photorealistic illustrations using multiple images as input. The novelty is in the simultaneous use of image and object space features to guide the rendering algorithm, being this that allows full automation of the synthesis process.

We would like to point out potential weaknesses for systems of this kind. First, as it is well know in stereo vision, we may get poor stereo estimates in some image areas. This will give unreliable normal estimates and consequently inaccurate orientations. Note that this will be improved considering more images and not just two as for the results shown in this paper. Second, and this is a general problem with NPR techniques, when computing principal directions we may get umbilical surface points where principal directions are not defined. If these are isolated points, either the smoothing can take care of these values or more sophisticated filling-in procedures can be applied like in [6].

To illustrate the ideas, we have used in this paper single objects, as commonly done in NPR applications. On the other hand, the combination of 2D and 3D NPR as here introduced provides the necessary tools for dealing with full scenes, see Figure 3 (e). This is the subject of current research. Note that in this case, the image+geometry information can be obtained from multiple images or from scanners that simultaneously output range and color information. How to deal with color images and its relation to color halftoning will also be addressed. Extending this work to NPR of video data and for real time applications (where frame-to-frame coherence issues arise) is the subject of current efforts as well.

Acknowledgments

This work is is partially supported by a grant from the Office of Naval Research, the Presidential Early Career Award for Scientists and Engineers (PECASE), and a National Science Foundation CAREER Award. We thank M. Nguyen and B. Chen for providing us the INSPIRE rendering system [7].

6. REFERENCES

[1] A. Girshick, V. Interrante, S. Haker, and T. Lemoine, "Line direction matters: an argument for the use of principal directions in 3d line drawings," *First International Symposium on NPR*, pp. 43–52, 2000.

[2] M. Salisburry, S. Anderson, R. Barzel, and D. Salesin, "Interactive pen-and-ink illustration," Tech. Rep., University of Washington, April 1994.

[3] M. Salisbury, M. Wong, J. Hughes, and D. Salesin, "Orientable textures for image-based pen-and-ink illustration," in *Proceedings of the 24th annual conference on CGIT*, 1997.

[4] G. Winkenbach and D. Salesin, "Rendering parametric surfaces in pen and ink," Tech. Rep., University of Washington, May 1996.

[5] E. Praun, H. Hoppe, M. Webb, and A. Finkelstein, "Real-time hatching," in *Proceedings of the 28th annual conference on CGIT*. 2001, p. 581, ACM Press.

[6] A. Hertzmann and D. Zorin, "Illustrating smooth surfaces," in *Proceedings of the 27th annual conference on CGIT*. 2000, pp. 517–526, ACM.

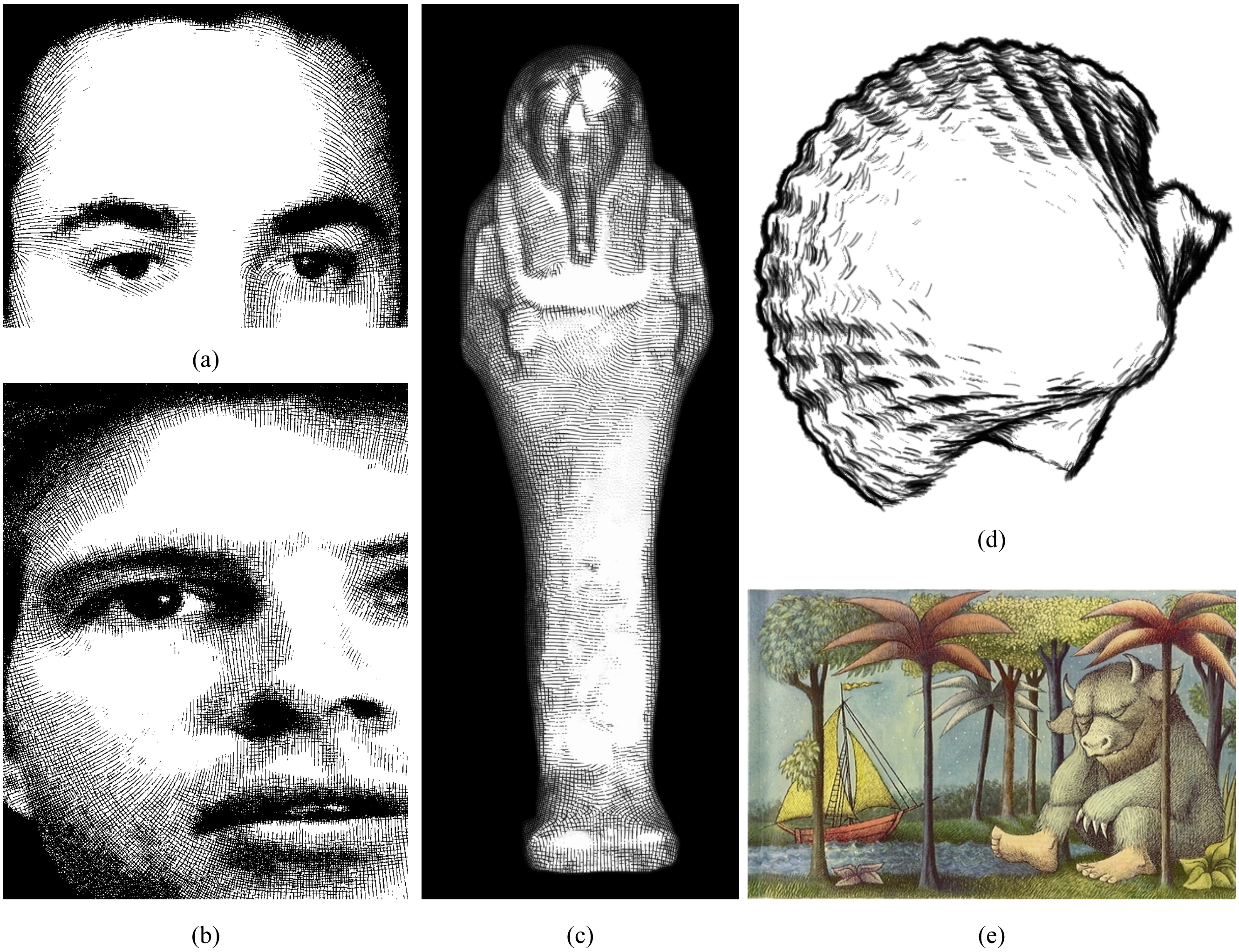

Fig. 3. (a,b) NPR from binocular stereo, data taken from `http://serdis.dis.ulpgc.es/~jsanchez`, and the RobotVis project at INRIA `http://www.inria.fr/`. (c,d) NPR from photometric stereo, data acquired with [11]. The shell image was produced using the INSPIRE rendering system [7]. (e) Illustration from "Where the wild things are" by M. Sendak. Our approach brings the necessary tools to produce this type of full scene NPR. Note that NPR from multiple images can be combined with realistic rendering, where portions of the image can be pasted in the rendered scene, thereby combining different rendering techniques. Previous single 2D or 3D techniques can not produce this.

[7] M. Nguyen, H. Xu, X. Yuan, and B. Chen, "Inspire: An interactive image-assisted npr system," in *SIGGRAPH*, Alberta, Canada, October 8-10 2001.

[8] T. Saito and T. Takahashi, "Comprehensible rendering of 3-d shapes," *Computer Graphics*, vol. 24, no. 4, pp. 197–206, August 1990.

[9] Y. Pnueli and A. Bruckstein, "Gridless halftoning: a reincarnation of the old method," *Graphical models and image processing*, vol. 58, no. 1, pp. 38–64, 1996.

[10] F. Devernay and O. Faugeras, "Computing differential properties of 3-d shapes from stereotopic images without 3-d models," Tech. Rep. 2304, INRIA, 1994.

[11] T. Malzbender, D. Gelb, and H. Wolters, "Polynomial texture maps," in *SIGGRAPH*, 2001, pp. 519–528.

[12] Robert J. Woodham, "Gradient and curvature from the photometric-stereo method, including local confidence estimation," *Journal of the Optical Society of America A*, pp. 3050–3068, November 1994.

[13] P. Perona, "Orientation diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 457–467, 1998.

[14] B. Tang, G. Sapiro, and V. Caselles, "Direction diffusion," in *ICCV (2)*, 1999, pp. 1245–1252.

[15] K. Toyama A. Criminisi, P. Prez, "Object removal by exemplar-based inpainting," in *CVPR*, June 2003.