Stylized depiction of images based on depth perception

Jorge Lopez-Moreno
Universidad de Zaragoza

Jorge Jimenez
Universidad de Zaragoza

Sunil Hadap
Adobe Systems, Inc

Erik Reinhard
Bristol University

Ken Anjyo
OLM Digital, Inc

Diego Gutierrez
Universidad de Zaragoza

Figure 1: *Different stylizations based on depth perception. Left: original image,* Vanitas *by Pieter Claesz (1630), oil on canvas. Middle: our* dynamic lines*, after relighting the scene (the window has been manually painted to motivate the new lighting scheme). Right: Color relighting imitating the chiaroscuro technique used by old masters like Caravaggio.*

Abstract

Recent works in image editing are opening up new possibilities to manipulate and enhance input images. Within this context, we leverage well-known characteristics of human perception along with a simple depth approximation algorithm to creatively relight images for the purpose of generating non-photorealistic renditions that would be difficult to achieve with existing methods. Our real-time implementation on graphics hardware allows the user to efficiently explore artistic possibilities for each image. We show results produced with four different styles proving the versatility of our approach, and validate our assumptions and simplifications by means of a user study.

CR Categories: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

Keywords: non-photorealistic rendering, relighting, image processing, human visual system

1 Introduction

Whether the goal is to convey a specific mood, to highlight certain features or simply to explore artistic approaches, non-photorealistic rendering (NPR) provides an interesting and useful set of techniques to produce computer-assisted stylizations. Most of those techniques either leverage 3D information from a model, work entirely in 2D image space, or use a mixed approach (probably by means of a Z- or G-buffer) [Durand 2002]. We are interested in exploring new possibilities for stylized depiction using a single image as input, while escaping traditional limitations of a purely 2D approach. For instance, the design of lighting schemes is crucial to communicate a scene's mood or emotion, for which depth information is required.

Our key observation is the fact that a single photograph or painting has richer information than we might expect. In particular, we ask ourselves what layers of information present in an image may have been usually overlooked by stylized depiction techniques? And what would the simplest way to access that "hidden" information be, in a way that allows dramatic manipulation of the look of an image?

This paper presents a set of stylization techniques that deal with a single photograph as input. It is well known that, when it comes to stylized depiction, human perception is able to build complex shapes with very limited information, effectively filling in missing detail whenever necessary, as illustrated in Figure 2 (left). The power of suggestion and the influence of light and shadows in controlling the emotional expressiveness of a scene has also been extensively studied in photography and cinematography [Kahrs et al. 1996; Alton 1945]: for instance, carefully placed shadows can turn a bright and cheerful scene into something dark and mysterious, as in Figure 2 (right).

With this in mind, we propose a new class of methods for stylized depiction of images based on approximating significant depth information at local and global levels. We aim to keep the original objects recognizable while conveying a new mood to the scene. While the correct recovery of depth would be desirable, this is still an unsolved problem. Instead, we show that a simple methodology suffices to stylize 3D features of an image, showing a variety of 3D lighting and shading possibilities beyond traditional 2D methods, without the need for explicit 3D information as input. An additional advantage of our approach is that it can be mapped onto the GPU, thus allowing for real-time interaction.

Within this context, we show stylized depictions ranging from simulating the *chiaroscuro* technique of the old masters like Caravaggio [Civardi 2006] to techniques similar to those used in comics. In recent years, both the movie industry (Sin City, A Scanner Darkly, Renaissance etc.) and the photography community (more than 4000 groups related to comic art on Flickr) have explored this medium. The goal of obtaining comic-like versions of photographs has even motivated the creation of applications such as Comic Life[1].

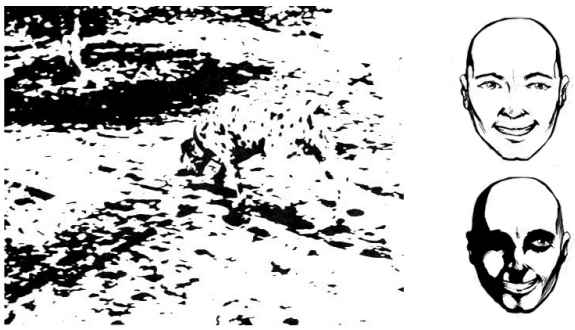

Figure 2: *Left: The classic image of "The Dog Picture", well known in vision research as example of emergence: even in the absence of complete information, the shape of a dog is clearly visible to most observers (Original image attributed to R. C. James [Marr 1982]). Right: Example of dramatically altering the mood of an image just by adding shadows.*

2 Previous Work

Our work deals with artistic, stylized depictions of images, and thus falls under the NPR category. This field has produced techniques to simulate artistic media, create meaningful abstractions or simply to allow the user to create novel imagery [Strothotte and Schlechtweg 2002; Gooch and Gooch 2001]. In essence, the goal of several schools of artistic abstraction is to achieve a depiction of a realistic image where the object is still recognizable but where the artist departs from the accurate representation of reality. In this departure, the object of depiction usually changes: a certain mood is added or emphasized, unnecessary information is removed and often a particular visual language is used.

In this paper, we aim to explore what new possibilities can be made available by adding knowledge about how the human visual system (HVS) interprets visual information. It is therefore similar in spirit to the work of DeCarlo and Santella [DeCarlo and Santella 2002] and Gooch et al. [Gooch et al. 2004]. DeCarlo and Santella propose a stylization system driven by both eye-tracking data and a model of human perception, which guide the final stylized abstraction of the image. Their model of visual perception correlates how interesting an area in the image appears to be with fixation duration, and predicts detail visibility within fixations based on contrast, spatial frequency and angular distance from the center of the field of view. Although it requires the (probably cumbersome) use of an eye-tracker, as well as specific per-user analysis of each image to be processed, the work nevertheless shows the validity of combining perception with NPR techniques, producing excellent results.

Instead, we apply well-established, general rules of visual perception to our model, thus freeing the process from the use of external hardware and individual image analysis. The goals of both works also differ from ours: whilst DeCarlo and Santella aim at providing meaningful abstraction of the input images, we are predominantly interested in investigating artistic possibilities.

Gooch and colleagues [Gooch et al. 2004] multiply a layer of thresholded image luminances with a layer obtained from a model of brightness perception. The system shows excellent results for facial illustrations. It is noted that in their approach some visual details may be difficult (or even impossible) to recover. Although in the context of facial stylization this counts as a benefit, it might not be desirable for more general imagery.

Depth information has previously been used to aid the generation of novel renditions. For instance, ink engravings can be simulated by estimating the 3D surface of an object in the image, and using that to guide strokes of ink [Ostromoukhov 1999]. This method is capable of producing high-quality results, although it requires the user to individually deform 3D patches, leading to a considerable amount of interaction. The algorithms proposed by Oh et al. [Oh et al. 2001] cover a wide range of image scenarios with specific solutions to extract 3D data for each one, but also come at the expense of considerable manual input. Okabe and colleagues [Okabe et al. 2006] present an interactive technique to estimate a normal map for relighting, whereas in [Yan et al. 2008], painterly art maps (PAMs) are generated for NPR purposes. While both works show impressive results, they again require intensive, skilled user input, a restriction we lift in our system.

In their work, Raskar and colleagues [Raskar et al. 2004] convey shape features of objects by taking a series of photographs with a multi-flash camera strategically placed to cast shadows at depth discontinuities. Akers et al. [Akers et al. 2003] take advantage of relighting to highlight shape and features by combining several images with spatially-varying light mattes, while in [Rusinkiewicz et al. 2006] details are enhanced in 3D models via exaggerated shading. In contrast, our approach operates on single off-the-shelf images, allows for new, artistic lighting schemes, and requires at most a user-defined mask to segment objects, for which several sophisticated tools exist [Li et al. 2004; Rother et al. 2004].

Finally, the 2.5D approach has been explored in the context of video stylization [Snavely et al. 2006], aiding the production of hatching and painterly effects. This method, however, requires the specific calibrated capture of the 2.5D video material to be processed, which is still either cumbersome or expensive. We show that 2.5D approximations suitable for NPR can be obtained from off-the-shelf images by applying current theories about the perception of shape.

3 Perceptual Background

At the heart of our algorithm, which will be described in the next section, lies the extraction of *approximate* depth information from the input image. Since we do not have any additional information other than pixel values, we obviously cannot recover depth accurately, and therefore the result will potentially contain large errors. However, given that we are interested in stylized depictions of images, we will show that we do not require physical accuracy, but only approximate values which yield pleasing, plausible results. Our depth approximation algorithm leverages some well-known characteristics of the human visual system. Although the inner workings of human depth perception are not yet fully understood, there exist sufficient indicators of some of its idiosyncracies that enable us to approximate a reasonable depth map for our purposes. In particular we rely on the following observations:

1. Belhumeur et al. [Belhumeur et al. 1999] showed that for unknown Lambertian objects, our visual system is not sensitive to scale transformations along the view axis.This is known as the *bas-relief ambiguity*, and due to this implicit ambiguity

[1] http://plasq.com/comiclife-win

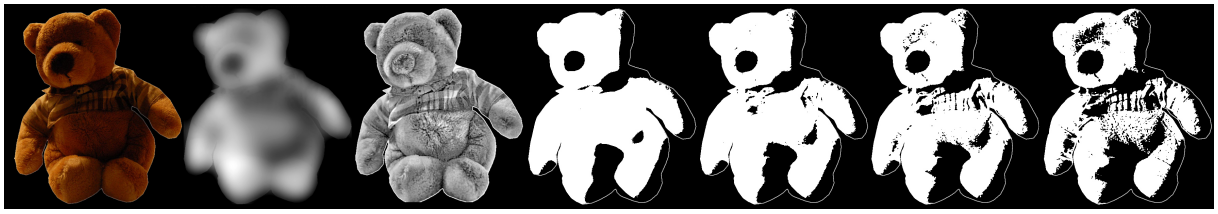

Figure 3: *Different combinations of the detail and base layer yield different depictions (here shown for the halftoning technique). From left to right: original image, base and detail layers, plus different depictions with a fixed $F_b = 1.0$ and increasing F_d from 0 to 1 in 0.25 increments.*

large scale errors along the view axis such as those produced in many single view surface reconstruction methods tend to go unnoticed.

2. Human vision tends to reconstruct shapes and percepts from limited information, for instance filling in gaps as shown in Figure 2, and is thought to analyse scenes as a whole rather than as a set of unconnected features [Loffler 2008; Elder and Zucker 1996].

3. Causal relationships between shading and light sources are difficult to detect accurately [Ostrovsky et al. 2005]. The visual system does not appear to verify the global consistency of light distribution in a scene [Langer and Zucker 1997]. Directional relationships tend to be observed less accurately than radiometric and spectral relationships.

4. There is evidence that human vision assumes that the angle between the viewing direction and the light direction is 20-30 degrees above the view direction [O'Shea et al. 2008].

5. In general, humans tend to perceive objects as globally convex [Langer and Bülthoff 2000].

In the next section we describe our algorithm, while, in Section 6 we will show the results of a user test validating our assumptions.

4 Algorithm

We rely on prior knowledge about perception, summarized above, to justify the main assumptions of our depth approximation algorithm. In particular, the bas-relief ambiguity (Observation 1) implies that any shearing in the recovered depth will be masked by the fact that we will not deviate from the original viewpoint in the input image [Koenderink et al. 2001]; in other words, we assume a fixed camera. The second and third observations suggest that an NPR context should be more forgiving with inaccurate depth input than a photorealistic approach, for instance by allowing the user more freedom to place new light sources to achieve a desired look, as we will see. Finally, the combination of the first, fourth and last observations allows us to infer approximate depth based on the dark-is-deep paradigm, an approach used before in the context of image-based material editing [Khan et al. 2006] and simulation of caustics [Gutierrez et al. 2008].

The outline of the process is as follows: first the user can select any object (or objects) in the image that should be treated separately from the rest. Usually the selection of a foreground and a background suffices, although this step may not be necessary if the image is to be manipulated as a whole. We assume that such selection is accomplished by specifying a set of masks using any existing tool [Li et al. 2004; Rother et al. 2004].

As stated before, the key process is the extraction of approximate depth information for the selected areas of the images. Accurate extraction of such information is obviously an ill-posed problem, studied for decades in the vision community [Durou et al. 2008], but for which a general-purpose solution has not been found. However, we will show how a very simple technique, which would perform poorly in other contexts, can actually yield excellent results for stylized depiction of images. In the last step of the process, the user can specify new lights as necessary (for which object visibility will be computed), and choose from a variety of available styles.

4.1 Depth Recovery

Our goal is to devise a simple depth recovery algorithm which works well in an NPR context and offers the user real-time control for stylized depiction. We aim to approximate the main salient features without requiring a full and accurate depth reconstruction. We take a two-layer approach, following the intuition that objects can be seen as made up of large features (low frequency) defining its overall shape, plus small features (high frequency) for the details. This approach has been successfully used before in several image editing contexts [Bae et al. 2006; Mould and Grant 2008; Rusinkiewicz et al. 2006], and has recently been used to extract relief as a height function from unknown base surfaces [Zatzarinni et al. 2009]. We begin by computing luminance values on the basis of the (sRGB) pixel input using $L(x,y) = 0.212R(x,y) + 0.715G(x,y) + 0.072B(x,y)$ [I.T.U. 1990]. Then we decompose the input object in the image into a base layer $B(x,y)$ for the overall shape as well as a detail layer $D(x,y)$ [Bae et al. 2006], by means of a bilateral filter [Tomasi and Manduchi 1998]. Additionally, as the methods based on the dark-is-deep assumption tend to produce depth maps biased towards the direction of the light, we smooth this effect by filtering $B(x,y)$ with a reshaping function [Khan et al. 2006] which enforces its convexity, producing an inflation analogous to those achievable by techniques like *Lumo* [Johnston 2002].

The detail layer D can be seen as a bump map for the base layer B. We decouple control over the influence of each layer and allow the user to set their influence in the final image as follows:

$$Z(x,y) = F_b B(x,y) + F_d D(x,y) \quad (1)$$

where $Z(x,y)$ is interpreted as the final, approximate depth, and F_b and F_d are user-defined weighting factors to control the presence of large and small features in the final image respectively, both independent and $\in [0,1]$. Figure 3 shows the results of different combinations of the base and detail layer of the teddy bear image, using the halftoning technique described in Section 5. This depth Z is stored in a texture in our GPU implementation (lower values meaning pixels further away from the camera). Figure 4 shows 3D renderings of the recovered depth for an input image; it can be seen how depth inaccuracies are more easily noticed if the viewpoint changes, while they remain relatively hidden otherwise.

The depth map Z serves as input to the relighting algorithm. Although a normal map could be derived from the depth map, it is not necessary for our purposes (except for the color relighting effect explained in Section 5).

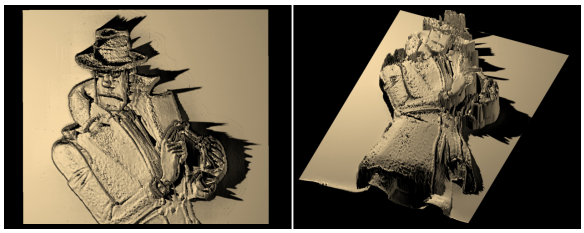

Figure 4: *Recovered depth from a given image. Errors remain mostly unnoticed from the original viewpoint (left), but become more obvious if it changes (right). Light and shadows have been added for visualization purposes.*

4.2 Computing Visibility for New Light Sources

The user can now adjust the lighting of the scene by defining point or directional light sources, to obtain a specific depiction or mood of the image. In the following, we assume a point light source at $\mathbf{p} = (p_x, p_y, p_z)^T$. There are no restrictions on where this light source can be placed.

Visibility is then computed on the GPU (in a similar fashion as other techniques such as parallax mapping [Tatarchuk 2006]): for each pixel in the framebuffer $\mathbf{q} = (x, y, z(x, y))^T$ belonging to an object we wish to relight, the shader performs a visibility test for the light (see Figure 5), by casting a ray towards its position. The pixels visited between $\mathbf{q}$ and $\mathbf{p}$ are given by Bresenham's line algorithm. The z-coordinate of the ray is updated at each step. Visibility is determined by querying the corresponding texels on the depth map. This information will be passed along to the specific NPR stylization techniques (see Section 5). Once a pixel visibility has been established, we can apply different NPR techniques to produce the desired stylized depiction of the image.

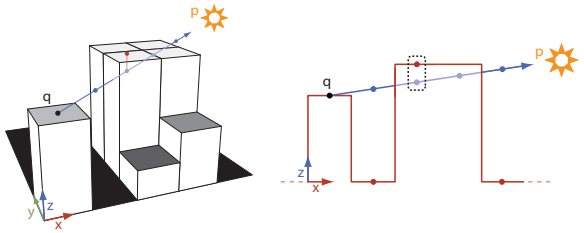

Figure 5: *3D and lateral views of the visibility computations for each texel.*

Figure 6: *From left to right: Input image. Output yielded by halftoning as described in [Mould and Grant 2008] (both images courtesy of D. Mould). Result lit by a close point light. Another result lit by a directional light.*

5 Results

We show a variety of examples which are currently implemented in our system. In each case, the defining difference over existing NPR work is the ability to relight the original image on the basis of the recovered 2.5D depth information. This adds versatility and artistic freedom. The different effects can be combined in layers for more complex looks, as some of our results show.

Halftoning: By simply mapping pixels visible from a light source to white and coloring all other pixels black, a halftoned rendition of the image is achieved. Figure 6 shows two examples of new relighting from an original input. Starting from a single image, we first create a halftoned version similar to what can be achieved with other systems (we use the implementation described in [Mould and Grant 2008], where the authors present a method based on segmentation from energy minimization). The remaining sequence of images in this figure shows the application of two novel lighting schemes that leverage the recovered depth information, thereby extending the capabilities of previous approaches. In the first one, a point light source has been placed at $(165, 240, 450)$ (in pixel units), whereas the second is lit by a directional light in the x direction. The weighting between detail and base layers is $(F_b, F_d) = (1.0, 0.9)$ for both images.

Multitoning: The spatial modulation of more than two tones (such as the black and white used in halftoning, plus several shades of gray) is known as multitoning or multilevel halftoning. In our implementation the user sets the position of a light source, after which a set of new lights with random positions located nearby the original is automatically created (the number of new lights is set by the user). This approach creates visually appealing renditions without having to place all light sources manually. Visibility is then computed separately for each light, and the results are combined in a single output by setting the value of each pixel in the final image to the average of the corresponding pixels in each layer. Results are shown in the second and sixth images in Figure 11 (in reading order) and the middle image of Figure 12 for three different input images.

Dynamic Lines: When sketching, an artist may draw lines towards the light source to add a more dynamic look to the scene. We can emulate a similar technique just by direct manipulation of the depth map. We randomly select a set of object pixels; the probability of choosing a specific pixel is set to be inversely proportional to the Euclidean distance to the position of the considered light source. The depth values of the selected pixels are altered, effectively changing the results of the visibility computations in the image and casting shadows which are perceived as lines. The third and ninth image in Figure 11 show final examples using this technique.

Color relighting: For each pixel belonging to the object, we compute a normalized surface normal $\vec{n}(x, y)$ from the gradient field $\nabla z(x, y)$ [Khan et al. 2006]:

$$\vec{g}_x(x, y) = [1, 0, \nabla_x z(x, y)]^T \quad (2)$$

$$\vec{g}_y(x, y) = [0, 1, \nabla_y z(x, y)]^T \quad (3)$$

$$\vec{n}(x, y) = \vec{g}_x \times \vec{g}_y / \|\vec{g}_x \times \vec{g}_y\| \quad (4)$$

Using this normal map as well as the 3D position of a light source, it is straightforward to modify pixel luminances or shading as function of the angle between the normals and the lights. Figures 11, 12 and 13 show examples with Gouraud shading. More sophisticated shaders could be easily incorporated.

6 Evaluation

In order to test our algorithm and the assumptions it relies on, we devised a psychophysical experiment to objectively measure how

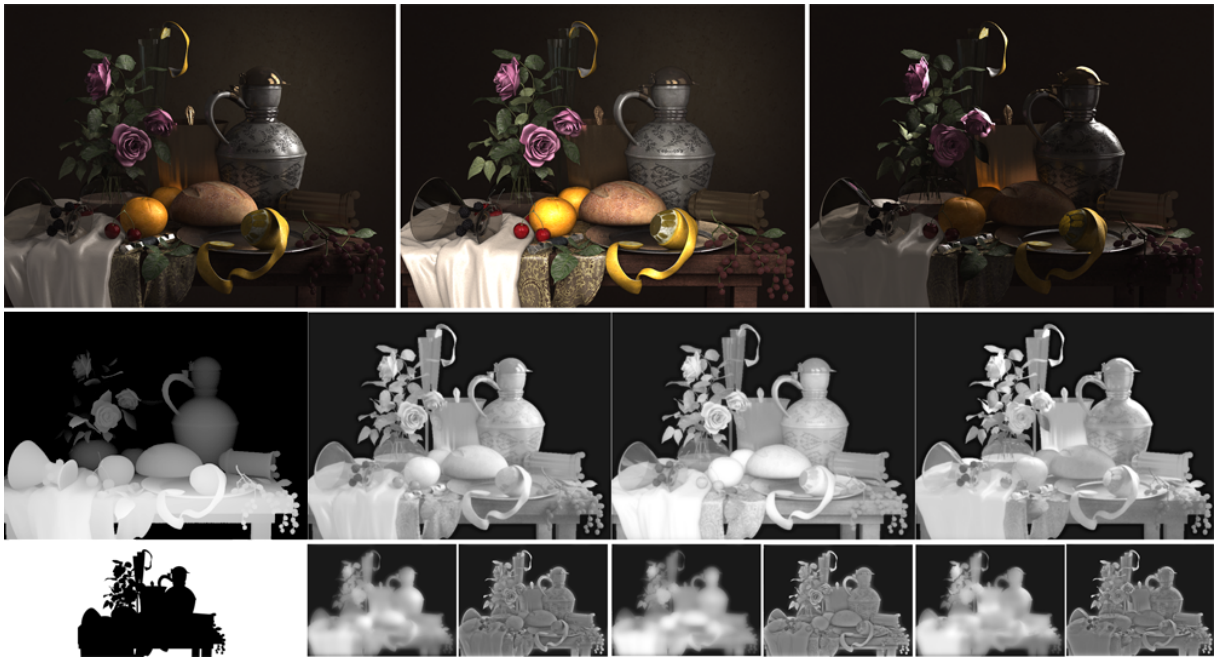

Figure 7: *First row: The three rendered images used as input in our test, lit by the original, frontal and back illumination schemes respectively. Second row: Ground truth depth map obtained from the 3D information of the scene (bumpmaps not included), plus approximate depths recovered for each of the input images. Third row: alpha mask, plus the base and detail layers of each image, used to obtain the corresponding depth maps.*

inaccurate the recovered depth is, compared to how well these inaccuracies work in an NPR context. The test is designed as follows: we take a rendered image of a 3D scene of sufficient diversity, having both complex and simple shapes, and a wide range of materials including transparent glass. Since it is a synthetic scene, its depth information is accurate and known, and we can use it as ground-truth. We then generate two additional depictions of the same scene, changing the lighting conditions. The original image has the main light falling in front of the objects at an angle from right-above; we thus create two very different settings, where light comes a) from the camera position (creating a very flat appearance) and b) from behind the objects. Together, the three lighting schemes (which we call original, front and back) plus the variety of shapes and materials in the scene provide an ample set of conditions in which to test our algorithm. Figure 7, top, shows the three resulting images.

We then compare the ground-truth depth map of the 3D scene with each of the approximate depths recovered using our image-based algorithm (with $F_b = 1.0$ and $F_d = 0.3$ according to Equation 1). Figure 7 (middle and bottom rows) shows the four depth maps, the alpha mask used to define foreground and background, and the base and detail layers for each approximate depth map. Note that the ground-truth depth is the same for the three images, whereas our approximated depth is different since it depends on pixel values.

Table 1 shows the results of the L_2 metric and correlation coefficient: our algorithm cannot recover precise depth information from just a single image, but the correlation with the ground truth is extremely high. Additionally, we also compare with a gray-scale version of the Lena image and with gray-level random noise (with intensity levels normalized to those of the 3D scene render), in both cases interpreting gray levels as depth information; both metrics yield much larger errors and very low, negative correlation. These results suggest that our simple depth extraction method approximates the actual depth of the scene well (from the same point of view, since we are dealing with static images). The question we ask ourselves now is, is this good enough for our purposes? In other words, is the error obtained low enough to achieve our intended stylized depictions of the input image, without a human observer perceiving inconsistencies in the results?

Input image	L_2	$Corr$
Original	100.16	0.93
Front	120.47	0.952
Back	121.66	0.925
Lena	383.92	-0.138
Random noise	524.74	-0.00075

Table 1: *Results of the L_2 metric and correlation coefficient comparing the ground-truth depth of the 3D scene with the approximate depth extracted from each input image, plus a gray-scale version of the Lena image and gray-level random noise (interpreting gray levels as depth).*

One of the main advantages of our approach over other image-based stylization techniques is the possibility of adding new light sources. We thus explore that dimension as well in our test: for each of the three input images, we create two new lighting schemes, one with slight variations over the original scheme, and one with more dramatic changes. Finally, for each of the six resulting images, we create halftoning, multitoning and color relighting depictions, thus yielding a total of eighteen images.

Given that the ultimate goal of our test is to gain some insight into how well our recovered depth performs compared to real depth information, for each of the eighteen stimuli we create one version using real depth and another using recovered depth. We follow a two-alternative forced choice (2AFC) scheme showing images side-by-side, and for each pair we ask the participants to select the one that looks better from an artistic point of view. A gender-balanced

set of sixteen subjects (ages from 21 to 39 years old) with normal or corrected-to-normal vision participated in the experiment. All subjects were unaware of the purpose of the study, and had different areas of knowledge and/or artistic backgrounds. The test was performed through a web site, in random order, and there was no time limit to complete the task (although most of the users reported having completed it in less than five minutes). Figure 8 shows some examples of the stimuli, comparing the results using real and approximate depth, for the three stylized depictions[2].

Figure 8: *Examples of the stimuli used in our user test, for the halftoning (top row), multitoning (middle row) and color relighting styles (bottom row).*

Figure 9 summarizes the results of our test, for the three styles (halftoning, multitoning and color relighting) and two light variations (similar, different). The bars show the percentage of participants that chose the depiction using approximate depth over the one generated with real depth. Despite the relatively large errors in the approximate depth (as the metrics from Table 1 indicate), the results lie very closely around the 50-percent mark. This indicates that, despite the sometimes obvious differences in the depictions due to the different depths employed (see for instance the two multitoning images in Figure 8), there is no significant difference in the participants' choices when judging the resulting artistic stylizations.

7 Discussion

We have shown results with a varied number of styles, all of which have been implemented on the GPU for real-time interaction and feedback, including relighting[3]. Our simple depth approximation model works sufficiently well for our purposes, while allowing for real-time interaction, which more complex algorithms may not achieve. On a GeForce GTX295, and for a 512×512 image and a single light source, we achieve from 110 to 440 frames per second. Performance decays with the number of lights: in our tests, real-time operation can be maintained with up to 5 light sources on average.

[2]Please refer to the supplementary material for the complete series.

[3]Please refer to the video.

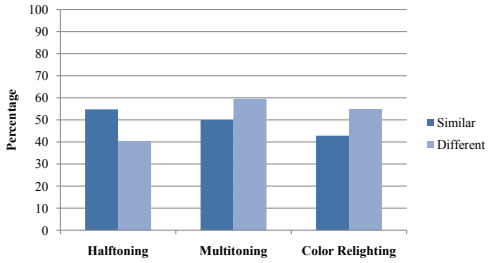

Figure 9: *Percentage of participants that chose the depiction using approximate depth over the one generated with real depth, for the three styles (halftoning, multitoning and color relighting) and two light variations (similar, different) per input image.*

Our approach has several limitations. If the convexity assumption is violated, the depth interpretation of our method will yield results which will be the opposite to what the user would expect them to be. For instance, the black nose of the bear toy in Figure 3 will be taken as an intrusive region, whereas in reality is protusive; thus, it cannot cast shadows and relighting may look wrong in that area. For small features such as the toy's nose it usually goes unnoticed, but if the object is not *globally* convex the results may not be plausible. It also assumes relatively Lambertian surface behavior: while highlights could be removed through thresholding or hallucination techniques, our assumptions on the perception of depth are broken in the case of highly refractive or reflective objects. In the latter case, shape-from-reflection techniques could be investigated. Also, since we do not attempt to remove the original shading from the image, our technique could potentially show artifacts if new lights are placed in the same direction of existing shadows (see Figure 10). However, our results confirm that quite large shading inaccuracies tend to go unnoticed in a NPR context. Finally, since we recover only depth information from camera-facing object pixels, completely accurate shadows cannot be produced.

Figure 10: *Artifacts due to original shadows in the image. Left: Detail of the original image depicted in Figure 11. Right: Relighting with a light source at* $(510, 520, 740)$ *wrongly illuminates the shadowed areas.*

Our method could potentially be used for video processing, for which temporal coherence should be considered. For the dynamic lines stylization technique proposed here, this could be specially tricky since it would most likely require tracking features at pixel level. Video segmentation is also a difficult task that would be necessary to address (although as some of the images in this paper show, compelling results can also be achieved in certain cases by processing the image as a whole). Finally, we expect that advances in the fields of perception and shape-from-shading will provide more exciting new grounds for artistic depiction of images and video.

8 Conclusions

We have presented a new methodology to develop NPR techniques based on the recovery of information about the depth from input images. Relying on known characteristics of human visual perception, our work offers more flexibility and artistic freedom than previous approaches, including the possibility of extreme relighting of the original image. Accurate extraction of depth information from a single image is still an open, ill-posed problem for which no solution exists. In this work we have shown that while our recovered depth is not accurate enough for certain applications, non-photorealistic stylization of images provides a much more forgiving ground, masking possible inconsistencies and leaving the abstraction process unhampered.

The fact that the algorithm also works well with a painted image (*Vanitas*) is quite interesting: a human artist painting the scene performs inaccurate depth recovery and very coarse lighting estimation, and the perceptual assumptions made by our algorithm seem to correlate well with the human artistic process. Future work to develop a system that mimics this process more closely can give us valuable insight and become a very powerful NPR tool.

Our 2.5D interpretation of objects in images yields an appropriate basis for appealing visual effects. We have shown several applications for this approach, such as halftoning, multitoning, dynamic lines and color relighting, but many more effects could be devised. For instance an interesting line of future work would be to incorporate local control over the stylization process as shown by Todo and colleagues [Todo et al. 2007]. We believe that the combination of our approach with other techniques such as gradient painting [McCann and Pollard 2008] or depth painting [Kang 1998] could open a wide range of possibilities in the field of image and video processing, and expect to see increasing future work on this subject.

9 Acknowledgments

The authors would like to thank Stephen Withers for the 3D still scene used in figure 7, and Ignacio Echevarria for the different renderings. Also the following Flickr users should be credited for the images used in our examples: JF Sebastian, gmeurope, 844steamtrain, sunset_chaser and km6xo. This research was partially funded by a generous gift from Adobe Systems Inc, the Gobierno de Aragón (projects OTRI 2009/0411 and CTPP05/09) and the Spanish Ministry of Science and Technology (TIN2007-63025).

References

AKERS, D., LOSASSO, F., KLINGNER, J., AGRAWALA, M., RICK, J., AND HANRAHAN, P. 2003. Conveying shape and features with image-based relighting. In *VIS '03: Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, IEEE Computer Society, Washington, DC, USA, 46.

ALTON, J. 1945. *Painting with Light*. Berkeley: University of California Press.

BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph. 25*, 3, 637–645.

BELHUMEUR, P. N., KRIEGMAN, D. J., AND YUILLE, A. L. 1999. The bas-relief ambiguity. *Int. J. Comput. Vision 35*, 1, 33–44.

CIVARDI, G. 2006. *Drawing Light and Shade: Understanding Chiaroscuro (The Art of Drawing)*. Search Press.

DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. *ACM Trans. Graph. 21*, 3.

DURAND, F. 2002. An invitation to discuss computer depiction. In *NPAR '02: Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, ACM, New York, NY, USA, 111–124.

DUROU, J.-D., FALCONE, M., AND SAGONA, M. 2008. Numerical methods for shape-from-shading: A new survey with benchmarks. *Comput. Vis. Image Underst. 109*, 1, 22–43.

ELDER, J. H., AND ZUCKER, S. W. 1996. Computing contour closure. In *In Proc. 4th European Conference on Computer Vision*, 399–412.

GOOCH, B., AND GOOCH, A. 2001. *Non-Photorealistic Rendering*.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: creation and psychophysical evaluation. *ACM Trans. Graph. 23*, 1, 27–44.

GUTIERREZ, D., LOPEZ-MORENO, J., FANDOS, J., SERON, F. J., SANCHEZ, M. P., AND REINHARD, E. 2008. Depicting procedural caustics in single images. *ACM Trans. Graph. 27*, 5, 1–9.

I.T.U. 1990. *Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange*. Geneva, ch. ITU-R Recommendation BT.709, Formerly CCIR Rec. 709.

JOHNSTON, S. F. 2002. Lumo: illumination for cel animation. In *NPAR '02: Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, ACM, New York, NY, USA, 45–ff.

KAHRS, J., CALAHAN, S., CARSON, D., AND POSTER, S. 1996. Pixel cinematography: A lighting approach for computer graphics. In *ACM SIGGRAPH Course Notes*, 433–442.

KANG, S. B., 1998. Depth painting for image-based rendering applications. U.S. Patent no. 6,417,850, granted July 9, 200.

KHAN, E. A., REINHARD, E., FLEMING, R., AND BÜLTHOFF, H. 2006. Image-based material editing. *ACM Transactions on Graphics (SIGGRAPH 2006) 25*, 3, 654–663.

KOENDERINK, J., DOORN, A. V., KAPPERS, A., AND TODD, J. 2001. Ambiguity and the mental eye in pictorial relief. *Perception 30*, 4, 431–448.

LANGER, M., AND BÜLTHOFF, H. H. 2000. Depth discrimination from shading under diffuse lighting. *Perception 29*, 6, 649–660.

LANGER, M., AND ZUCKER, S. 1997. Casting light on illumination: A computational model and dimensional analysis of sources. *Computer Vision and Image Understanding 65*, 322–335.

LI, Y., SUN, J., TANG, C.-K., AND SHUM, H.-K. 2004. Lazy snapping. In *Siggraph*, ACM, Los Angeles, California, 303–308.

LOFFLER, G. 2008. Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision research* (May).

MARR, D. 1982. *Vision*. W. H. Freeman and Company, New York.

MCCANN, J., AND POLLARD, N. S. 2008. Real-time gradient-domain painting. *ACM Trans. Graph. 27*, 3, 1–7.

MOULD, D., AND GRANT, K. 2008. Stylized black and white images from photographs. In *NPAR '08: Proceedings of the 6th*

international symposium on Non-photorealistic animation and rendering, ACM, New York, NY, USA, 49–58.

OH, B. M., CHEN, M., DORSEY, J., AND DURAND, F. 2001. Image-based modeling and photo editing. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 433–442.

OKABE, M., ZENG, G., MATSUSHITA, Y., IGARASHI, T., QUAN, L., AND YEUNG SHUM, H. 2006. Single-view relighting with normal map painting. In *In Proceedings of Pacific Graphics 2006*, 27–34.

O'SHEA, J. P., BANKS, M. S., AND AGRAWALA, M. 2008. The assumed light direction for perceiving shape from shading. In *ACM Applied Perception in Graphics and Visualization (APGV)*, 135–142.

OSTROMOUKHOV, V. 1999. Digital Facial Engraving. ACM Press/ACM SIGGRAPH, New York, 417–424.

OSTROVSKY, Y., CAVANAGH, P., AND SINHA, P. 2005. Perceiving illumination inconsistencies in scenes. *Perception 34*, 1301–1314.

RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, ACM, New York, NY, USA, 679–688.

ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. GrabCut: Interactive foreground extraction using iterated graph cuts. In *Siggraph*, ACM, Los Angeles, California, 309–314.

RUSINKIEWICZ, S., BURNS, M., AND DECARLO, D. 2006. Exaggerated shading for depicting shape and detail. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, ACM, New York, NY, USA, 1199–1205.

SNAVELY, N., ZITNICK, C. L., KANG, S. B., AND COHEN, M. 2006. Stylizing 2.5-d video. In *NPAR '06: Proceedings of the 4th international symposium on Non-photorealistic animation and rendering*, ACM, New York, NY, USA, 63–69.

STROTHOTTE, T., AND SCHLECHTWEG, S. 2002. *Non-Photorealistic Computer Graphics*.

TATARCHUK, N. 2006. *Shader X5*. Charles River Media, ch. Practical Parallax Occlusion Mapping, 75–105.

TODO, H., ANJYO, K., BAXTER, W., AND IGARASHI, T. 2007. Locally controllable stylized shading. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, ACM, New York, NY, USA, 17.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision*, 836–846.

YAN, C.-R., CHI, M.-T., LEE, T.-Y., AND LIN, W.-C. 2008. Stylized rendering using samples of a painted image. *IEEE Transactions on Visualization and Computer Graphics 14*, 2, 468–480.

ZATZARINNI, R., TAL, A., AND SHAMIR, A. 2009. Relief analysis and extraction. *ACM Transactions on Graphics, (Proceedings of SIGGRAPH ASIA 2009) 28*, 5.

Figure 11: *Stylized results achieved with our method. Top row, left: Original input image. Top row, right: Multitoned depiction with two point light sources at (506,276,1200) and (483,296,900), and using* $(F_b, F_d) = (0.5, 0.8)$. *Second row, left: Multitoned image with two layers of dynamic lines added, generated from the same light at (500,275,1000). Second row, right: Result of multiplying color relighting with the multitoned version. Third row, from left to right: Mask with foreground objects (window painted manually for artistic effect and motivate subsequent relighting), multitone depiction of* Vanitas*, and result of multiplying two layers of color relighting and five layers of dynamic lines (please refer to the supplementary material to see the individual layers). Fourth row, from left to right: Original input image,* Dynamic lines *version placing a light source at both headlights, and a multilayer combination similar to* Vanitas *figure above.*

Figure 12: *Application of our method to a very diffusely lit image. In this example we aim to obtain different moods by changing the light environment and the degree of stylization. Left: Original input image. Middle: A very stylized and dark version of the input by multitoned depiction with four point light sources at (140,400,300), (140,400,350), (140,400,400) and (140,400,900) and using* $(F_b, F_d) = (1.0, 0.2)$. *Right: Less stylized depiction obtained by combination of multitone and color relighting effects with lights at (134,530,290), (115,15,270), (315,695,350), (100,400,1000) and (589,325,325). No mask was used for these depictions.*

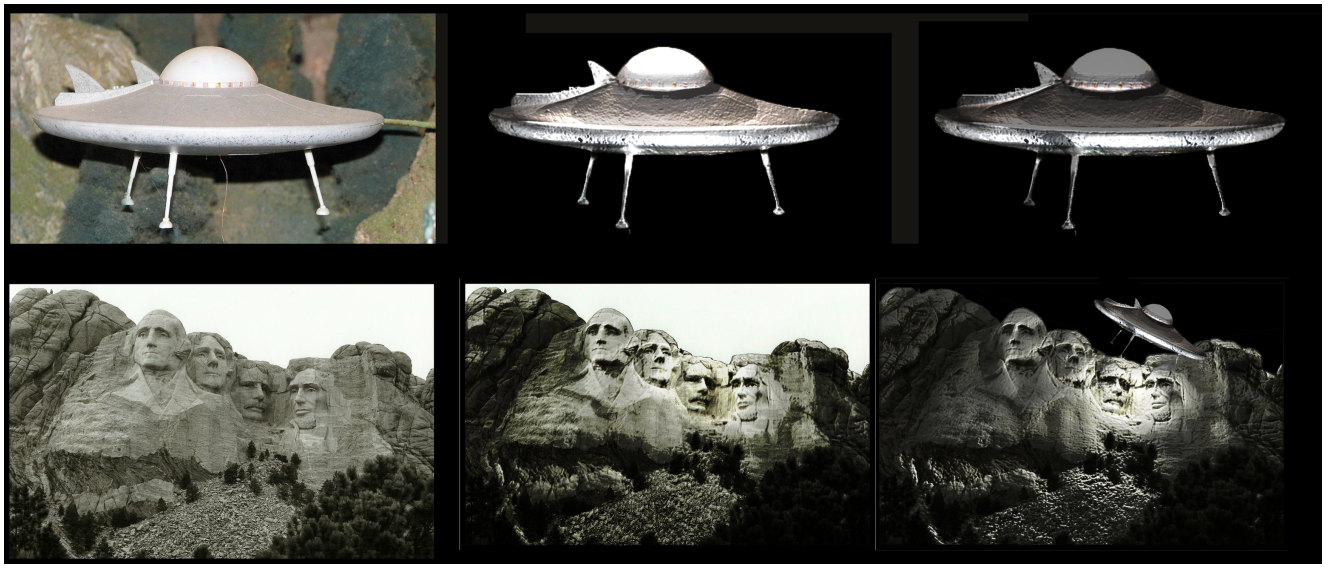

Figure 13: *Composition of results. Top row, left: Original input image. Top row, middle: Color relighting with five point light sources:two from above at* $x = 480, y = 520, z = (500, 250)$ *and three surrounding the disk at* $x = (50, 550, 100), y = 400, z = 1000$, *and using* $(F_b, F_d) = (1.0, 0.1)$. *Top row, Right: result of multiplying a shadow layer created by a light source at (580,0,500) and the relighted image (middle). Second row, from left to right: Original input image, stylized depiction by combination of color relighting and halftone, and result of compositing the relighted UFO from top row and a new relit version of the input image.*