INTERACTIVE NON-PHOTOREALISTIC VIDEO SYNTHESIS FOR ARTISTIC USER EXPERIENCE ON MOBILE DEVICES

Javier Villegas and Angus Graeme Forbes

Creative Coding Lab
School of Information: Science, Technology, and Arts
University of Arizona

ABSTRACT

This paper presents analysis/synthesis strategies for generating abstract, creative representations via the camera input on a mobile device. Mobile devices are well suited for interactive video processing since they are simultaneously capable of image capture, display, and manipulation. Analysis/synthesis methods are particularly powerful in interactive arts projects as they enable even drastic manipulations of the input image while still maintaining fundamental aspects of its original identity. Moreover, by using abstract synthesis elements (i.e., coherent elements larger than single pixels), we are able to directly interact with the image and to manipulate its final output. We describe some of the exciting capabilities of video processing and interaction on mobile devices and introduce a series of mobile applications that use analysis/synthesis techniques.

1. INTRODUCTION

Nearly all modern smart phones and tablets are capable of handling multimedia data, including video input and output. Moreover, most mobile devices include dedicated GPUs that are powerful enough to allow real-time manipulation of video. These devices are also equipped with various sensors that enable many interaction possibilities. This paper presents an investigation of analysis/synthesis (hereafter, A/S) techniques that take advantage of these capabilities.

1.1. Analysis/synthesis techniques on video signals

In A/S techniques, the original signal is recreated using the parameters that result from an analysis process. This strategy provides a high level of manipulability in the input signal without altering its fundamental identity. This is the main reason why A/S techniques are extensively used in creative audio processing (e.g., Vocoder, Autotune). But, although A/S approaches are also used in many image and video applications like compression and coding, few examples of creative manipulation of video signals exist.

An area of research closely related to our work in A/S techniques is called "non-photorealistic rendering" (or NPR), which aims to creatively interpret the raw data from realistic images and videos. NPR is primarily concerned with automatically recreating the look of different styles of hand-made paintings. In fact, most of the image-based NPR algorithms are A/S processes, in which an input image is used to calculate the position, color, orientation or texture of the synthesis elements [1]. In a seminal paper, Haeberli demonstrates different alternatives for abstract representation of natural and synthetic images [2]. He explores the use of different primitives, such as brush strokes, and successfully mixes together automatic and semi-automatic techniques. Stylization effects, including "mosaic," "pixelate," and "cubism," are now part of standard digital photo-manipulation tools. These effects can be interpreted as A/S processes (although they are usually not referred to as such).

1.2. Art on mobile devices

Many artists have taken advantage of the various possibilities of mobile devices and have recognized their potential to engage audiences in novel ways. This potential includes both the increasing computational power of current hardware and also the distribution model of the applications that can be easily downloaded to run on them [3]. Recent work by the Creative Coding Lab at the University of Arizona explores various methods to creatively manipulate video, mostly based on A/S approaches, and implements these approaches on mobile devices. In our work, each video frame is analyzed and then recreated after incorporating the creative manipulations generated by user input. For programming of the mobile devices we utilize the *Aluminum* framework, a powerful multimedia codebase developed at the Creative Coding Lab that gives developers access to current, low-level features of OpenGL and OpenGL ES [4]. Specifically, the framework provides access to the GPU capabilities of modern iOS mobile devices while simplifying most of the Objective-C configuration code that iOS devices require.

The power of A/S strategies is based on their ability to distort the input image parameters in various ways before reconstructing an output. In the next section, we illustrate some of the parameters that can potentially be manipulated before this reconstruction.

2. NON-PHOTOREALISTIC VIDEO ELEMENTS

Humans give meaning to images by grouping elements within them and then associating them with previous knowledge. There are many different ways in which picture elements can be grouped [5], and these different grouping principles can be exploited to create *abstract mirrors*, or non-photorealistic representations of a raw video capture.

2.1. Regions

A *region* can be defined as a group of connected pixels that share a similar color, luminance, or texture. Figure 1 shows different ways of re-drawing regions that were detected in an analysis stage. Manipulations can then occur before the reconstruction of the image.

Fig. 1. Different ways to present identified regions: (Left) replacing regions with ellipses; (Right) re-drawing the regions after modifying their Fourier descriptors.

2.2. Edges and lines

The identification of contours is fundamental to object perception [5]. A completely new family of examples can be created if, in the analysis stage, parametric curves are fitted to the edge of the map on the input image. An example is shown in Figure 2.

2.3. Orientation

The human visual system can also separate regions based on orientation [6]. A direct mapping from gray value to line orientation produces the results shown in Figure 4.

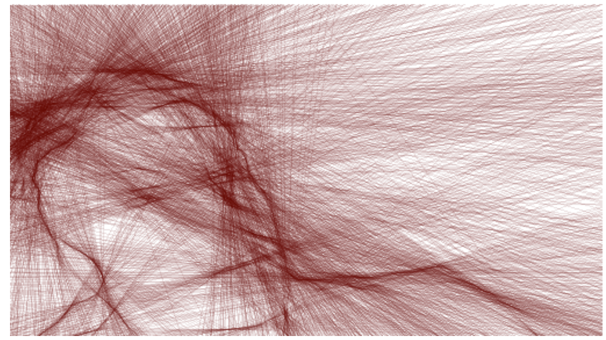

Fig. 2. Edges of the image recreated with straight lines.

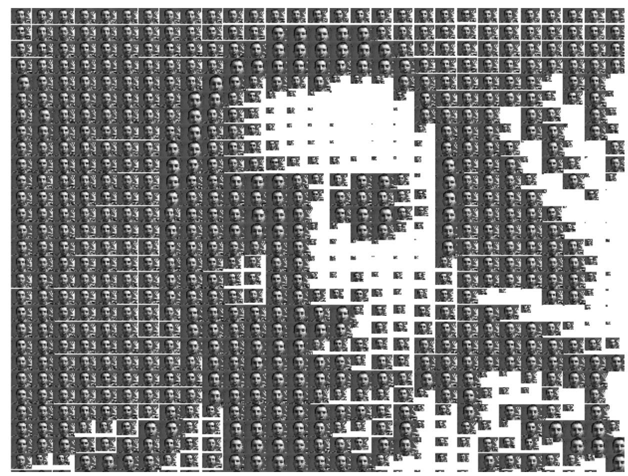

Fig. 3. The gray-scale levels are mapped to the size of the synthesis element.

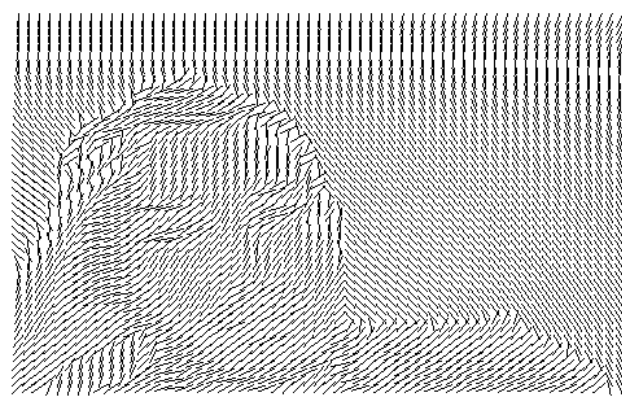

Fig. 4. Gray-levels are mapped to line orientation.

2.4. Size

Objects can also be grouped into different regions by size. Areas represented with larger objects will be perceived as darker than areas represented with smaller objects (assuming dark objects on white background). Figure 3 shows a visual experiment using size as a mapping parameter. A gray-scaled image is used as input and also as the synthesis object.

2.5. Density

We perceptually separate regions that have similar element density [5]. Also, a region can be perceived as darker than others if there is more object density in that region. This is the basic concept behind many dithering techniques [1]. In Figure 5 the dot density is changed to recreate the gray levels of the input.

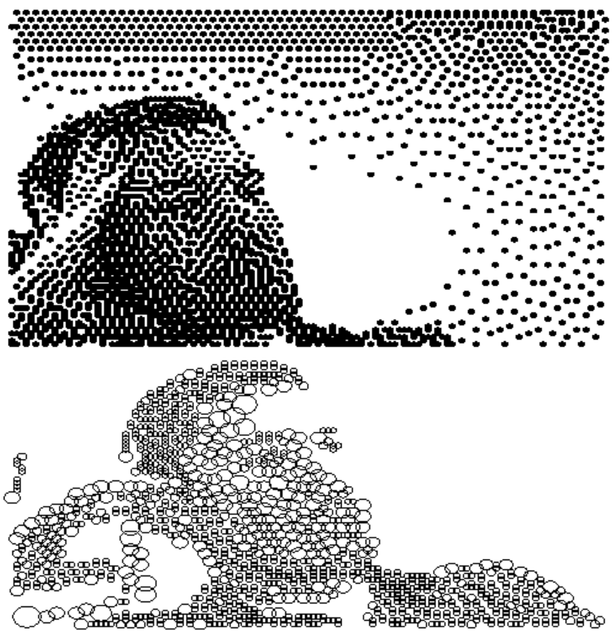

Fig. 5. Image represented by manipulating the density of objects.

2.6. Connectedness

Figure 6 shows different ways of using the output of a dithering algorithm to join the resultant black points with lines following different paths. Connectedness is another strong grouping principle [5]. Small differences in the way that points are connected can drastically change the look of the result.

3. REAL-TIME MOBILE IMPLEMENTATIONS OF NON-PHOTOREALISTIC VIDEO SYNTHESIS

In this section we show how we use some of the grouping principles described above to implement interactive abstract mirrors that work in real-time on current mobile devices.

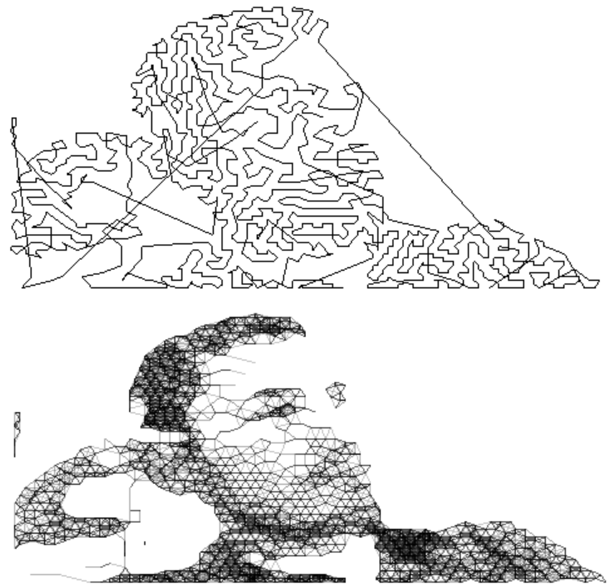

Fig. 6. Two images showing different ways to connect dots.

Specifically, we present three different applications that we have implemented using the *Aluminum* framework [4].

3.1. The *Fluid Automata* application

Fluid Automata is an interactive fluid simulation and vector field visualization application based on cellular automata systems [7]. A flow of energy is distributed through the image using local rules that are inspired by the physics of fluids. The system is illustrated on Figure 7. Different alternatives for interaction with mobile devices have been explored using the multitouch capabilities of mobile devices. For instance, energy can be added to the system by tapping the screen or moving a finger across the screen to change the direction of vectors in the fluid system. Figure 8 shows one of the possible outputs of the system. *Fluid Automata* has been presented as an independent artistic piece [8] and also as a visual accompaniment for dynamic music compositions [9].

3.2. The *Angle Shift* application

Figure 4 illustrates how orientation can be used a a discriminator for regions. *Angle Shift* is an interactive application that combines orientation, color and motion. The synthesis elements are thin lines that change position, direction, and color according to the current camera input. The new state of each line is updated from the previous one in the CPU using the camera image and the interaction information (touchscreen) as inputs. The updated information is then sent to the GPU for drawing via modern OpenGL commands. Figure 9 shows a block diagram of the system and

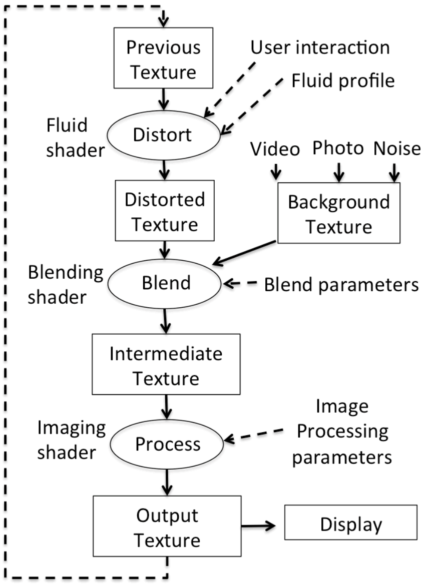

Fig. 7. Steps of the *Fluid Automata* application.

Fig. 8. The *Fluid Automata* application running on an iPad.

Figure 10 show an example of a typical output of the application.

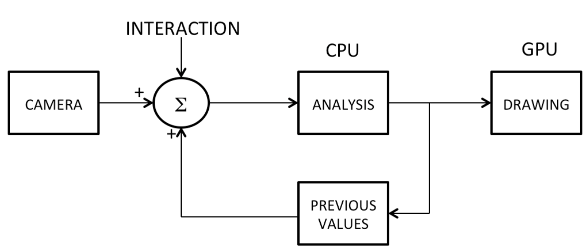

Fig. 9. Block diagram of the *Angle Shift* application.

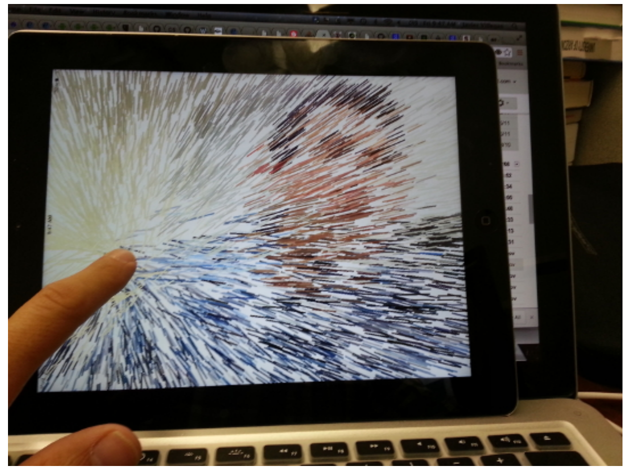

Fig. 10. The *Angle Shift* application running on an iPad.

3.3. The *Meshflow* application

In the *Meshflow* application, a set of points evolves slowly before finally morphing into the image of the camera input. The motion is continuous and the position of each point directly evolves from its position in the previous frame [10]. In *Meshflow*, the nodes on the grid are attracted to the darkest areas of the image, but the grid structure is kept. The motion of the nodes is constrained by physics laws that describe an attraction between neighbor nodes and a drag force. Figure 11 shows the block diagram of the application.

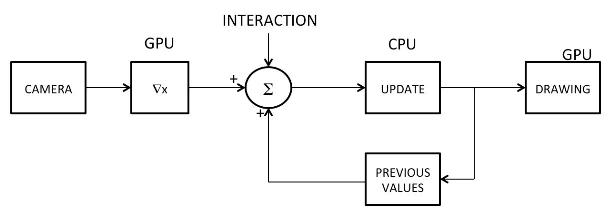

Fig. 11. Block diagram of the Meshflow application

The gradient of the image is calculated using the Sobel operator in the GPU. This information is then encoded into two 8-bit channels and then read back into the CPU. In the CPU, the previous values of the nodes of the grid are updated using the gradient information and using the coordinates of the interaction in the touchscreen. The geometry information is then sent back again to the GPU for final rendering to the screen. Figure 12 shows a screen shot of the application running in an iPad.

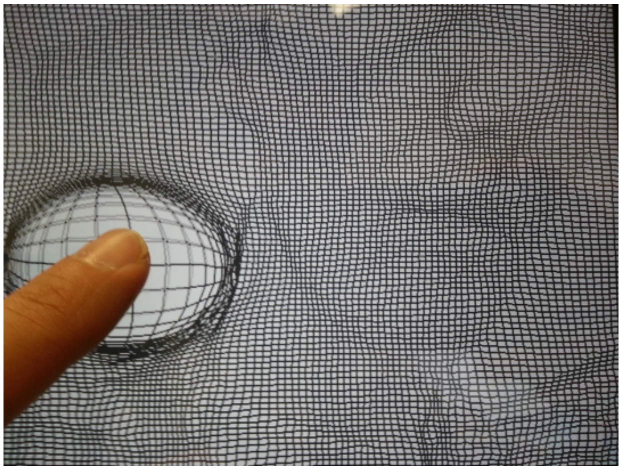

Fig. 12. The *Meshflow* application running on an iPad.

4. SUMMARY AND FUTURE WORK

We believe that A/S strategies are an effective means to generate creative mobile applications that function as abstract mirrors. Using A/S techniques, an input video frame is reduced to a set of parameters, then those parameters are manipulated, and finally an abstract representation is created with the modified parameters. In each of the examples presented here, the input information is not used to recreate the image from scratch, but instead as a way to update the previous state of the synthesis elements. This guarantees that the synthesis elements evolve smoothly on the output image. Strong changes in object parameters (particularly position) are the factors that have perhaps the strongest impact in the perceived quality of an abstract animation. We are exploring ways to measure the quality, perceived naturalness, and smoothness of the transitions, a surprisingly difficult task. Some potential evaluations we are exploring include: the displacement distribution of the synthesis objects; the minimum jerk rule (stipulating that a a motion path between two points is maximally smooth when it minimizes the mean square of the jerk, i.e., the derivative of the acceleration) [11]; and the spatial dispersion of the direction of motion. Part of our present research involves finding correlations between these quantitative measurements and the qualitative evaluations gathered via user studies.

In each of the three applications we presented, users interact via the touchscreen. Future versions may also include other sensors common to mobile devices (e.g., accelerometers, gyroscopes, audio inputs) to manipulate various parameters before the output synthesis. Lastly, in future versions we would like to incorporate the concept of "cross-synthesis," that is, mixing the parameters of two images mixed *before* the synthesis stage. One application of cross-synthesis is the creation of what we call "double meaning" images, where a single image has both a local and a global narrative occurring simultaneously. Future work will extend our previous research on double meaning images [12] and explore its application to mobile devices.

5. REFERENCES

[1] Thomas Strothotte and Stefan Schlechtweg, *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*, Morgan Kaufmann, 2002.

[2] Paul Haeberli, "Paint by numbers: abstract image representations," *SIGGRAPH Comput. Graph.*, vol. 24, pp. 207–214, September 1990.

[3] Scott Snibbe, "Snibbe studio apps," `http://www.snibbestudio.com/apps/`.

[4] Angus Graeme Forbes, "Aluminum," `https://github.com/CreativeCodingLab/aluminum`.

[5] C. Ware, *Information visualization: perception for design*, vol. 22, Morgan Kaufmann, 2004.

[6] Jacques Bertin, *Semiology of graphics: diagrams, networks, maps*, University of Wisconsin press, 1983.

[7] Martin Gardner, "Mathematical games: The fantastic combinations of john conways new solitaire game life," *Scientific American*, vol. 223, no. 4, pp. 120–123, 1970.

[8] Angus Graeme Forbes, Tobias Höllerer, and George Legrady, "Generative fluid profiles for interactive media arts projects," in *Proceedings of the Symposium on Computational Aesthetics*. ACM, 2013, pp. 37–43.

[9] Angus Graeme Forbes and Kiyomitsu Odai, "Iterative synaesthetic composing with multimedia signals," in *Proceedings of the International Computer Music Conference (ICMC)*, Ljubjiana, Slovenia, September 2012, pp. 573–578.

[10] Javier Villegas, "Meshflow: A grid warping mirror," 2010, `https://vimeo.com/21653145`.

[11] P. Viviani and T. Flash, "Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 1, pp. 32, 1995.

[12] Javier Villegas and Angus Graeme Forbes, "Double-meaning: Interactive animations with simultaneous global and local narrative," in *Proceedings of of re-new 2013 media art conference and festival*. Re new, 2013, pp. 300–304.