UNIVERSITY OF CALIFORNIA
Santa Barbara

An Analysis-Synthesis Approach to the Creative Processing of Video Signals

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Media Arts & Technology

by

Javier Villegas Plazas

Committee in Charge:

George Legrady, Chair

Curtis Roads

Matthew Turk

Lisa Jevbratt

September 2012

The Dissertation of
Javier Villegas Plazas is approved:

Curtis Roads

Matthew Turk

Lisa Jevbratt

George Legrady, Committee Chairperson

June 2012

An Analysis-Synthesis Approach to the Creative Processing of Video Signals

Acknowledgements

I would like to thank my wife for drastically changing the path of her life on order to join me in this adventure, giving me the happiest days of my life. Thanks to my sisters, my dad and my in-laws for all their support during these years. Many thanks to the Media Arts and Technology Program and the Office of the Regents for their generosity with the Central Fellowship. I would also like to thank all the current and former MAT faculty, especially the ones with which I have had the opportunity to work: George Legrady, Lisa Jevbratt, Jerry Gibson and Stephen Pope. An additional big thanks to Curtis Roads, Larry Zins, Matthew Turk, Angus Forbes and Theodore Kim for their help in various stages of the project. In general, thanks to all the MAT community whose diversity and creativity has been motivating and inspirational.

Curriculum Vitæ
Javier Villegas Plazas

Education

2000	Bachelor in Electrical Engineering: Javeriana University, Bogotá Colombia.
2002	Master of Science in Electrical and Computer Engineering, University of Los Andes, Bogotá Colombia.
2012	Doctor of Philosophy in Media Arts & Technology, University of California, Santa Barbara (expected)

Experience

2005-2007	Assistant Professor, Department of Electrical Engineering; Javeriana University, Bogotá Colombia.
2008-2011	Teaching Assistant, Media Arts & Technology Department, University of California, Santa Barbara.
2008, 2010	Project Engineer, "We Are Stardust," interactive installation. Pasadena and Vancouver.
2010	Video Processing Intern, Intel Corporation, Digital Home Group, Chandler Arizona

Awards and Honors

"Narrative Line," Short animation, 2^{nd} place in animation category, Swan Lake: Moving Image & Music Awards 2011, Mittweida Germany, 2011.

"Merit Award," Video Processing Intern, Intel corporation, Digital Home Group, Chandler Arizona, Summer 2010.

"The Fitting Dance," Experimental animation, Short animation division, Jury recommended work 2009 Japan media Arts Festival. Japan, 2009.

"Regents Special Fellowship," University of California, Santa Barbara, 2007.

Selected Publications

"A Foveal Architecture for Stereo Matching," With Restrepo, Proc IEEE ICIP 2002 Vol II, pp 521-524 Rochester NY, September 2002.

"A Coding Method for Visual Telephony Sequences," With Barcenas et al. Proceedings of the Auditory-Visual speech Processing 2005, Pg 87-92. Vancouver, British Columbia, CANADA. ISBN 1 876346 53 1.

"The We Are Stardust Installation," With legrady and Burbano. Proceedings of the seventeen ACM international conference on Multimedia. Pg: 1087-1090, ISBN:978-1-60558-608-3. Beijing, China, 2009.

"The Autonomous Duck: Exploring the Possibilities of a Markov Chain Model in Animation," Arts and technology, First International Conference, ArtsIT 2009, Yi-Lan, Taiwan, September 24-25, 2009, Revised Selected Papers, pg 272-278. Springer ICST 2010.

"Meshflow: A Grid-Warping Mirror," Advances in Computer Entertainment Technology (ACE), Taipei Taiwan, 2010.

"Real Time 2D to 3D Conversion: Technical and Visual Quality Requirements," With Caviedes, ICCE, las vegas, 2011.

"Generating Time-Coherent Animations from Video Data," With Legrady, to appear in volume 0101 of the Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering series.

Exhibitions

"We Are Stardust" in "Observe" exhibition with George Legrady, at the Art Center College of Design (October 10, 2008 and January 11, 2009) and in the CODE LIVE exhibition at the Vancouver Olympics 2010.

"The Fitting Dance," Short animation division, 13^{th} Japan media Arts Festival. Japan 2009.

"The Fitting Dance," Short animation, Byte Gallery International Exhibition summer 2011. Transylvania University. 2011.

"Triangles and Cats," Short animation, 6^{th} edition of the Streaming Festival. December 1 to 18 2011.

"Self-Portrait with Lines," "Self-Portrait with Ellipses," Exhibition of Mathematical Art, joint mathematics meetings 2012.

Abstract

An Analysis-Synthesis Approach to the Creative Processing of Video Signals

Javier Villegas Plazas

Analysis and synthesis (A/S) approaches are common in multimedia signal processing. Reducing a signal to its constituent parts is relevant because once the signal is analyzed, the set of extracted parameters can be more convenient for storage, finite length representation, transmission or manipulation than the signal in its original form. The community of audio artists have adopted A/S techniques as an important part of their arsenal of technological tools for the creative manipulation of sounds. Parameters can be altered before the synthesis, generating sophisticated sound effects while maintaining the identity of the original signal. Tools such as the tracking phase vocoder allow musicians to manipulate pitch and duration independently or to combine two signals in exotic ways (cross-synthesis). However, a similar emphasis has yet to be pursued for the creative manipulation of video signals. This dissertation describes how A/S approaches can be used to obtain non-photorealistic representations from live video. In this research, new algorithms have been designed to overcome problems like temporal coherence and to extend concepts like the cross-synthesis of two video signals. The creative and narrative possibilities of the developed techniques are used to compose complete pieces divided between short animations and real time installations. Details on

the design and implementation of those pieces are included in this document. The pieces, together with the different examples used throughout the dissertation illustrate the versatility and control that can be obtained with A/S approaches and confirm that these sets of techniques can be considered for the creative manipulation of video data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Description

The definition of analysis in the Random House unabridged American dictionary is: “The separating of any material or abstract entity into its constituent elements (opposed to synthesis).” In the same dictionary, synthesis is defined as: “The combining of the constituent elements of separate material or abstract entities into a single or unified entity (opposed to analysis).” As these definitions show, in a general context, analysis and synthesis are considered to be opposite processes. Analysis is understood as breaking something into its parts, and synthesis is defined as putting parts together to form a whole. “Analysis/Synthesis” (A/S) refers to the combination of these two processes. An entity is built again using the parts resulting from the analysis stage. A complete analysis process that is followed by a synthesis process, should leave the input unaffected if the parts are not altered in between the two processes.

These standard definitions fall short in a signal processing context. In signal processing, analysis and synthesis are closely related to the problem of signal representation. The set of "parts" that conform a signal varies according to the chosen representation. Even more, the analysis stage sometimes implies the adoption of a particular model of production as well as the extraction of its parameters. (e.g., the source-filter model of speech generation). To emphasize the fact that the synthesis is made using features acquired through the analysis of an existing signal, the chain of processes is sometimes called "Analysis/Resynthesis."

Sometimes the representation used is not adequate for the type of signal analyzed. In those cases, the reconstruction can only be approximated and artifacts will appear. Artifacts will also be present if the synthesis uses only a subset of the parts that were obtained during the analysis stage to recreate the signal. A/S techniques are the base of lossy compression techniques. In lossy compression, some information that is not considered perceptually relevant is discarded before the synthesis.

Besides compression applications, in the audio domain A/S approaches have also been used very successfully for creative manipulation of sounds. Parameters can be altered before the synthesis, generating sophisticated sound effects while maintaining the identity of the original signal. Artifacts from imperfect reconstructions can be used as a tool for expression. Different A/S algorithms have been developed to allow musicians to do refined manipulations: changing independently the pitch and duration [23], controlling the amount of noisy and periodical components in a sound [59], separating the instrument response from the excitation, or combining two signals in exotic ways (cross-synthesis) [56]. However, despite the evident creative and narrative possibilities of video signals and the accessibility of low cost tools for producing and sharing videos, little attention has been given to the use of A/S approaches for the creative manipulation of video sequences. A/S techniques are also extensively used in videos for compression, for the automation of the process of rotoscoping, mosaicing, stylization and many other uses. But little exploration has been done regarding the possibilities of the manipulation of A/S parameters as a creative tool.

This thesis will explore the use of A/S techniques for the creative manipulation of video signals. Video signals are sequences of still images, usually with strong information correlation between consecutive frames. When manipulations are done independently on each of the frames, the temporal coherence of the video is lost. One of the principal problems addressed in this document (chapter 4) is how to recreate a video after manipulations with A/S strategies while keeping the temporal coherence of the output.

If the constituent elements that recreate an image are chosen not to be just abstract shapes but figurative objects with identifiable form and behavior, the reconstructed image will have information at local and global levels. This ambiguity can be considered as an equivalent of cross-synthesis on the video domain and will be discussed in detail in chapter 5.

1.2 Scope and Limitations

In all the explorations presented in this work, some information is extracted from a real world image in order to build a non-photorealistic representation of it. In order to do that, no strong assumptions about the contents of the input image are made, and a search for specific high level information is never done. None of the algorithms presented in this document as part of the analysis stage look for specific concrete elements such as heads, facial features, joints, hands, etc. The only input used was an RGB image or at times a silhouette. In addition, this work does not include animations created from other sophisticated motion capture systems.

In this work, synthesis is always done assembling objects with a spatial structure bigger than the pixel. Pixel oriented transformations are not considered and neither are applications that recreate concrete entities associated with well determined inputs (like virtual glasses, hats, or face substitution). Although there is a full chapter dedicated to the problem of temporal coherence, and a significant relevant work on this topic has been done by the community of researchers of non-photorealistic animation, the algorithms described here are not intended as general solution to the problem of time coherent stylistic animations from video. For instance, they do not try to recreate hand drawn styles on moving images.

Different research questions will be addressed in this work:

1. What new aesthetical and expressive possibilities are favored by Analysis and Synthesis approaches?

2. How to create video outputs where the synthesis has temporal coherence?

3. What can be considered as the equivalence to cross-synthesis on the video domain?

4. What algorithms are needed to generate cross-synthesized videos?

5. What are the narrative possibilities of cross-synthesis on videos?

1.3 Contributions

The contributions of this work can be summarized as follows:

- **Mapping algorithms:** Different algorithms for recreating digital images with a collection of objects are presented. The possibilities of these mappings for animation and real-time installations are shown through examples. The basic techniques can be extended to new designs and new applications.

- **Temporal coherence:** Novel approaches to the problem of time coherent animation from video are presented, some of these approaches can be used on real-time installations. A quantitative measurement of smoothness versus quality of representation is defined. Although this particular evaluation may not be generalizable to other techniques, the introduction of a quantitative evaluation gives a new perspective on how to analyze these kinds of results.

- **The automatic generation of ambiguous images/videos:** The popular musical effect of cross-synthesis is extended to images and videos. An algorithmic approach for the creation of ambiguous images is presented and it is extended to ambiguous videos.

- **The use of cross-synthesis for parallel narratives:** Examples of animations and real-time installation showing the possibilities of simultaneous narratives when using cross-synthesis on videos are presented.

1.4 Document Outline

In the next chapter, an overview of creative uses of A/S approaches on media arts is given and some related problems on non-photorealistic animation are described. Chapter 3 will describe the general system and will illustrate it with examples of the different categories of image mappings. In chapter 4, three alternatives for the generation of time coherent video sequences are presented. The first one is based on the matching of similar objects through combinatorial optimization. The second and third alternatives use a particle system interpretation of the atomic synthesis elements to create coherent animations at interactive rates. Chapter 4 includes quantitative measurements for the comparison of the different techniques. Chapter 5 will show alternatives for the generation of cross-synthesis on video signals. Chapter 6 goes through examples that clarify the narrative possibilities of A/S methods for animators and media artists. Finally, in chapter 7 the conclusions and implications and presented.

1.5 Chapter Summary

The main point of this chapter was to show that, contrary to what happens in the music community, artists that use the moving image as a medium of expression are not exploiting the full possibilities of A/S techniques. New challenges and research questions arise when these approaches are used. The aim of this work is to give answers to those research questions, not only by designing and presenting algorithms, but also through the creation of pieces based on them.

Chapter 2

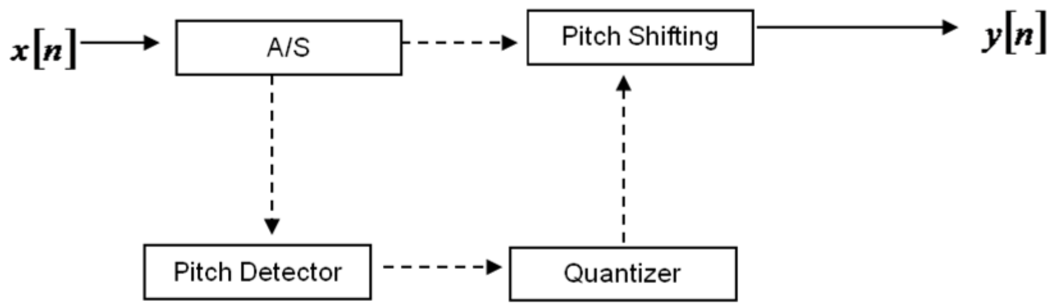
Review of A/S in Media Arts

2.1 A/S Techniques on Audio Signals

The extraordinary versatility that the A/S approaches provide, have been exploited mostly in the music domain. An important variety of algorithms has been developed and adopted by new generations of sound designers. These techniques include the phase vocoder [23]. This technique and some of its variations have been successfully implemented in multiple software packages and they are now standard components in the music industry. The phase vocoder allows different manipulations of the audio signal: independent pitch and duration modification, dispersion, robotization or whispering [74].

An interesting way to extend a pitch modification algorithm is to use a pitch detection module as input information for the pitch shifting stage. Natural sounds can be forced to be only within certain ranges (Figure 2.1). In this manner, the phase vocoder can be used for tone correction. This technology has been taken to extremes with strong quantization steps and it is now known as the "Cher Effect"' or "Autotune Effect." This unnatural tone jumping that was considered an undesired artifact in speech compression applications is now considered as a new vocal texture, and it is used extensively by many popular artists [27].

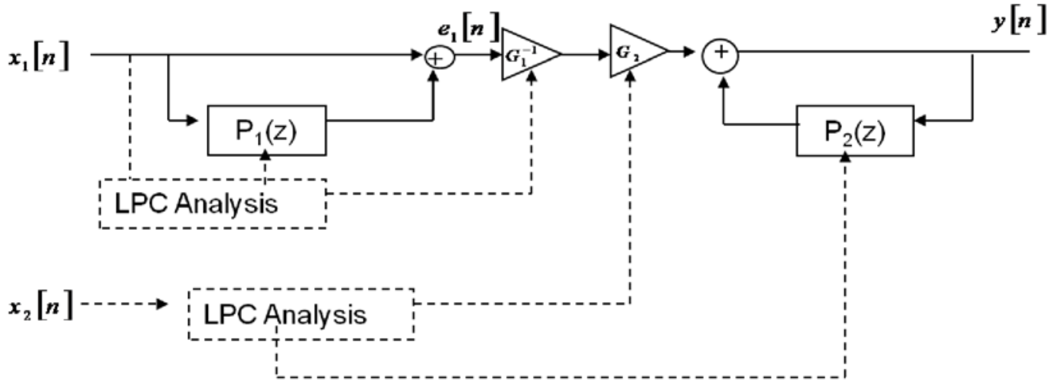
Figure 2.1: Pitch correction with a phase vocoder.



Another very classical A/S based audio effect known as cross-synthesis is the mixture of the analysis data of two signals. The output of the mixture process is a hybridized version of the two inputs. One way to create this effect is by using the LPC vocoder as a source-filter separator. The residual of one signal is used to excite the filter model of a second one [38]. The system is illustrated in Figure 2.2.

A more detailed historical review of A/S in audio can be found in [56]; the two previous approaches were mentioned here because they have special significance in the proposed application of A/S strategies over video signals.

Figure 2.2: Cross-synthesis with the LPC vocoder.



2.2 A/S on Images

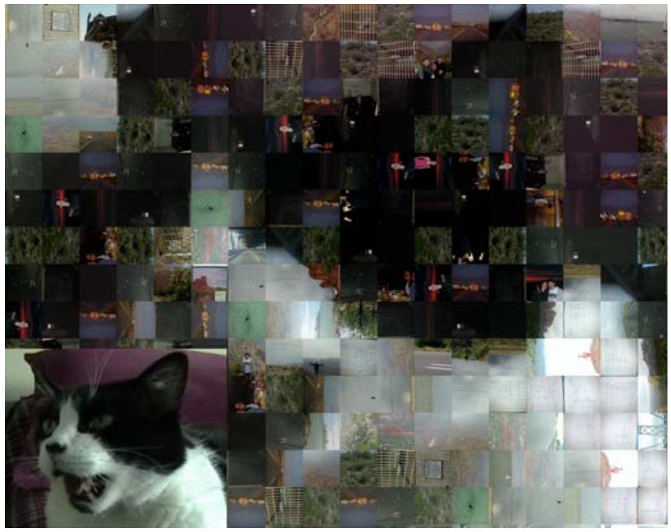
In 1966, Kenneth Knowlton and Leon Harmon, while working in Bell labs, recreated a nude picture using an array of small symbols [22]. This work was exhibited at one of the earliest digital art exhibitions and it became famous after it was printed in The New York Times. Similar approaches were followed by others, such as Waldemar Cordeiro in his work "Derivadas de Uma Imagem"(1969) [25]. In 1967 Charles Csuri and James Shaffer created the "Sine Curve Man" which was a line drawing image of the face of a man that was digitalized. After that, the vertical coordinate of the black points in the image was altered by adding a sinusoidal term of different amplitudes [19].

Also in the late 60's the Japanese Computer Technique Group (CGT) was well know for the use of computers to generate geometrical compositions based on portraits. Part of their works were presented at the visionary exhibition Cybernetic Serendipity (1968)[54].

As it was shown in section 2.1 for the audio case, the discretization of parameters perceived as being continuous can produce interesting creative results. Mosaics can be seen as the spacial quantization of an image. Perhaps, mosaics are the best known applications of A/S techniques on images. Mosaics are two-dimensional arrays of small pieces that conform to an image or pattern. Mosaics can be created after an analysis process where the characteristics of the small elements (e.g., size, color and orientation) can be calculated from an input image. Chuck Close is a remarkable portrait artist that uses an A/S method for his creations and explorations using a grid method and without the use of computers [26]. Close's work has inspired different attempts to make digital approximations of it [48, 1].

Ken Knowlton is still an active producer of computer assisted portraits [39]. His work has been inspirational for Robert Silvers, who created a technique (and a company) for building mosaics out of small pictures. This technique is now know as Photomosaics[61]. It is important to note that in most of Knowlton and Silvers work, the small constituent elements are not pure abstract shapes. The elements have an identity and there is usually a relation between the global and the local. Figure 2.3 was created using the free photomosaic application "AndreaMosaic" [21]. Vic Muniz is a Brazilian artist that also recreates digital images using different kinds of materials, from sugar or chocolate to garbage [51, 69]

Figure 2.3: A photomosaic created with the free software: “AndreaMosaic.”



Golan Levin has developed interesting and original portraits using A/S techniques. In “Segmentation and Symptom” (2000), he created Voronoi diagrams presumably from dithered versions of portraits [42]. The Voronoi mesh is denser in darker areas of the original picture. Previously, in his 1999 work “Floccular Portraits,” black and white filaments are attracted to the darker and brighter parts of an input image [41].

2.2.1 Non-Photorealistic Rendering

Non-photorealistic rendering (NPR) is primarily concerned with automatically recreating the look of different styles of hand made paintings. Most of the image-based NPR algorithms are A/S processes. An input image is used to calculate the position, color, orientation or texture of the synthesis elements [64]. In his seminal paper [31], Haeberli showed different alternatives for abstract representation of natural and synthetic images. He explores the use of different primitives as brush strokes, and mixed automatic and semi-automatic techniques. Stylization effects like mosaic, pixilation and cubism, are now part of standard digital picture manipulation tools [7]. These effects can be interpreted as A/S processes, although they are usually not referred to as such. Figure 2.4 shows different standard effects that can be easily obtained with these popular tools.

Figure 2.4: Different A/S effects that can be obtained in minutes with standard digital photograph editing tools.

2.3 A/S on Videos

There is a natural interest in extending the algorithms of image manipulations to moving sequences. Emulating the ASCII art movement, Vuk Cosic in 2007 created ASCII movies of classic films [18]. Video effects software packages often extend the standard two-dimensional set of image effects on each of the video frames, including those effects based on A/S techniques [53]. In addition, researchers in academy and industry working on non-photorealistic rendering have been looking for strategies to create stylistic animations from video. The problem of temporal coherence is still the main focus of research. As pointed out by Bénard, Bousseau and Thollot [9] in their excellent state of the art review on time coherent stylistic animations: "the issue of temporal coherence prevents the widespread adoption of research algorithms by the industry." Different alternatives to guarantee temporal coherence on non-photorealistic rendered animations have been explored. In 1996, Meier [50] used particles over 3D surfaces to keep track of the position and direction of brush strokes in consecutive frames of pure synthetic non-photorealistic scenes. In SIGGRAPH 1997, Litwinowicz presented a paper [44] where he describes a technique to create hand-painted-looking image animations from live action video. He used the edge map of an input image to constrain the length of the strokes, then he used optical flow to ensure coherence of the strokes between successive frames. A video created with those techniques was also presented at the same conference [43]. A similar approach oriented to interactive real-time applications, was created by Hertzmann and Perlin [34]. Bénard et al. had explored the use of dynamic textures and Gabor noise primitives to create time coherent stylizations [8, 10]. In Animosaics [62], Smith, Liu and Klein explored rules for the smooth motion of mosaic tiles in animated mosaics. They pursued not only coherence between individual elements, but cohesion in the movement of groups of tiles. A detailed review of techniques of non-photorealistic animation can be found on the papers from Bénard, Bousseau and Thollot [9], and Agrawal [2].

2.3.1 Musical videos

Commercial musical videos have sometimes been a fertile area for creative exploration. Some popular video clips have used A/S strategies with diverse levels of automation. The groundbreaking video for the song "Take on Me" [3] mixes the real world with hand drawing animations using rotoscoping. The Michel Gondry video for the white stripes song "Fell In Love With A Girl" recreates live action with LEGO® blocks [29]. Less popular videos that also use A/S strategies are: "Black Tambourine" by Beck (2004) which was created with animated ASCII art, David Fincher directed on 2005 the Nine Inch Nails video for the song "Only" in which the 3D information of the singer is recreated on a nail array. "Ankle Injuries" by Fujiya & Miyagi (2007) which features live footage recreated using colored dice as primitives. The 2009 Radiohead video "House of Cards" shows 3D points in space affected with particle dynamics, the video was created as an Open Source project and the 3D point data is available for download [6]. A similar strategy of altering the 3D points with dynamics and connections is used in the video "Fragile Tension" by Depeche Mode (2010). The 2010 video "Just the Way you Are " by Bruno Mars animates an audio tape inspired in the art work of Erika Iris Simmons [36].

2.4 A/S for Real-Time or Interactive Media Art

Real-time installations using A/S approaches had been created by Daniel Rozin in his collection of mirrors [13]. He analyzes the real-time captured image of the spectator and recreates it using mechanical and software representations in the synthesis stage. Jim Campbell in his series of "low resolution works" [16] creates a new representation of real images with the combination of quantized digital binary images and diffuse screens.

In two of his pieces on the "Shadow Box" series, Rafael Lozano-Hemmer uses local, small meaningful parts to form the image of the spectator. In "Eye Contact" [46], small videos get into different states, depending on the image of the viewer. In "Third Person," the image of the audience is built out of verbs conjugated in the third person [47].

The versatility of A/S techniques for transcoding, which is, using the signal captured in one media to represent it in another media, was shown by Woody Vasulka and Brian O'Reilly on their scan processor studies [66]. Recordings taken from a Rutt-Etra video synthesizer are used to create a new audio visual experience.

2.4.1 The Creative Coding Era

The creation of open source programming environments like Processing or OpenFrameworks simplified the problem of interfacing the real and synthetic worlds. An important number of A/S experiments can be found on the Internet and in creative coding literature. Processing demos and simple examples for teaching the language show how to build software mirrors based on A/S approaches. Two outstanding sources that are worth mentioning are the book "Generative Gestaltung" [12] which has excellent examples of image redrawing, and the blog of Robert Hodgin [35] which features various examples and explorations. The launching of the Kinect® device, combined with a multitude of open source drivers for reading it, gave the creative coding community the possibility of having a low cost system for the tracking of anatomic parts. The growing interest in the design of new computer interfaces with those devices has created newer A/S applications, including video controlled avatars and face substitution systems [72].

2.5 Chapter Summary

In this chapter a general overview of A/S techniques on different media was presented. A/S techniques are today a standard tool on music processing applications, and sounds altered with A/S strategies are part of the palette of elements for contemporary musicians. But the path fallowed on visual arts has been different. Experimentation was strong during the first years of computer art. The fact that it was difficult to have a faithful visual representation of an image using the hardware of the time pushed the pioneers to explore new modes of figurative computer art. The capabilities of output devices were significantly limited and the stages of analysis and synthesis were clearly separated, even physically!. For instance, to create the 1966 work of Kenneth Knowlton "Studies on Perception I," the original picture was first scanned with a special camera that converted gray levels to analog voltages and then those voltages were digitalized and stored on magnetic tape. A program then did the symbol mapping, and in the last step the data was printed on a microfilm plotter [65]. The next big wave of A/S on images came with the creation of algorithms for NPR. An important number of image processing "filters" based on A/S approaches has been incorporated as a standard part of image editing software environments. Many of those image effects can also be applied on each of the frames of a video sequence, but if the relation between consecutive frames is not managed the result will have disturbing jitter artifacts.

Although the aesthetic appeal of A/S techniques has made them ubiquitous in all media, the application of those techniques is still limited on moving images. The use of them as a narrative tool has been restricted in part because of the problems associated with the lack of temporal coherence. Little work has been done on the creation of algorithms that represent the equivalent of cross-synthesis in the video domain. Those topics will be elaborated upon in subsequent chapters.

Chapter 3

Redrawing Images

This work explores the creative possibilities of redrawing image sequences from live action video. In this particular chapter, the emphasis will be on the spatial part. Temporal aspects will be covered in detail in the next chapter. The general idea of A/S processes is to recreate the original signal with a set of elements or primitives. The way in which those primitives relate can be extremely diverse, in part since the human visual system is undoubtedly skillful in finding relations and patterns. In the next subsection, different categories of image redrawing that were explored are presented and illustrated with examples.

3.1 Redrawing Regions

A region can be defined as a group of connected pixels that share a similar color, luminance or texture. In the same way that quantization of the fundamental frequency produces the new sonic textures of the autotune effect, spatial quantization of the regions gives a new texture to the image. The quantization also introduces high frequency noise that is not presented in the original image. This noise can easily be eliminated by low-pass filtering (e.g., viewing the image from a larger distance) [32] which creates an interesting conflict between the local and the global. Figure 3.1 shows the image of a portrait redrawn after fitting ellipses to connected regions of similar luminance.

Figure 3.1: Replacing connected regions with elliptic shapes.

A different alternative is to manipulate the frequency content of the contour of each region using Fourier descriptors:

- Every point (x, y) in the contour is represented as a complex number $x + iy$.
- The FFT of the complex signal is calculated and the FFT points representing high frequencies are zeroed.

- The inverse FFT is evaluated and the contour is redrawn using the real part of each point as the x coordinate and the imaginary part as the y coordinate.

Figure 3.2 shows the reconstruction of the input image limiting the number of Fourier descriptors in the reconstruction to three and five.

Figure 3.2: Recreating the cumulative connected regions of the input with only three Fourier descriptors (left) and 5 Fourier descriptors (right).

Yet another approach is to build the regions by filling them with known objects or tiles. The alignment, overlap, and gaps in between the elements radically affects the final look of the resulting image. This problem has been studied by the NPR community in order to recreate the look of hand made mosaics (see [62] and [20]). In figure 3.3, the dark regions of a face are filled with full body silhouettes.

Figure 3.3: Filling the black areas of an image with black silhouettes.

3.2 Edges and Lines

The identification of contours is fundamental to object perception [70]. A complete 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. A very well known technique to detect parametric curves is the Hough transform [24]. Figure 3.4 shows examples of the output of the system using parabolas. Figure 3.5 shows the result with spline curves.

Figure 3.4: Synthesis using parabolas of different parameters.

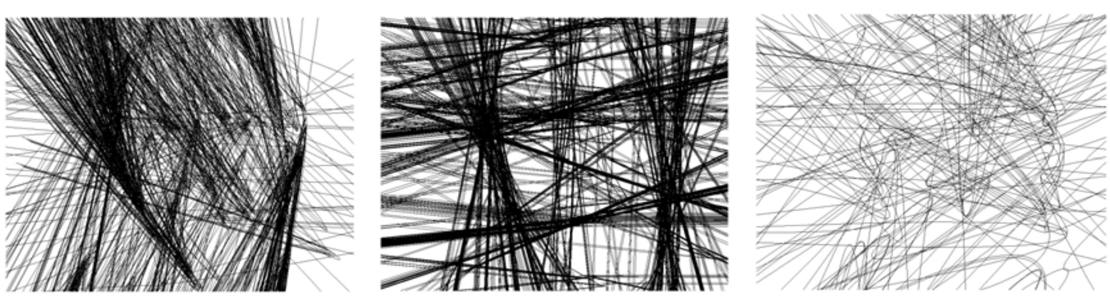
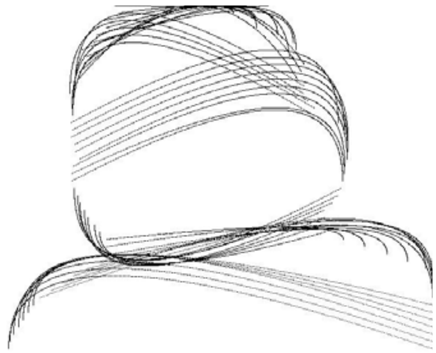


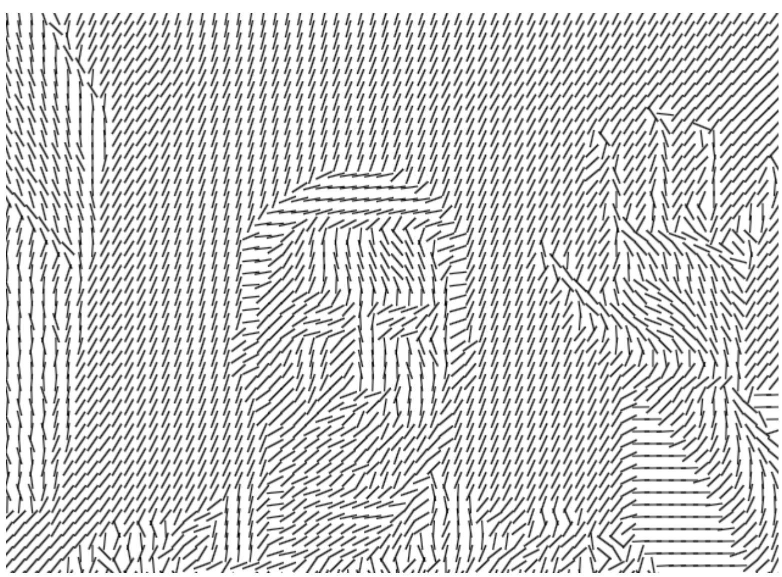
Figure 3.5: Synthesis using Bezier curves.



3.3 Orientation

The human visual system can also separate regions based on orientation. A direct mapping from gray value to line orientation produces the results shown in figure 3.6.

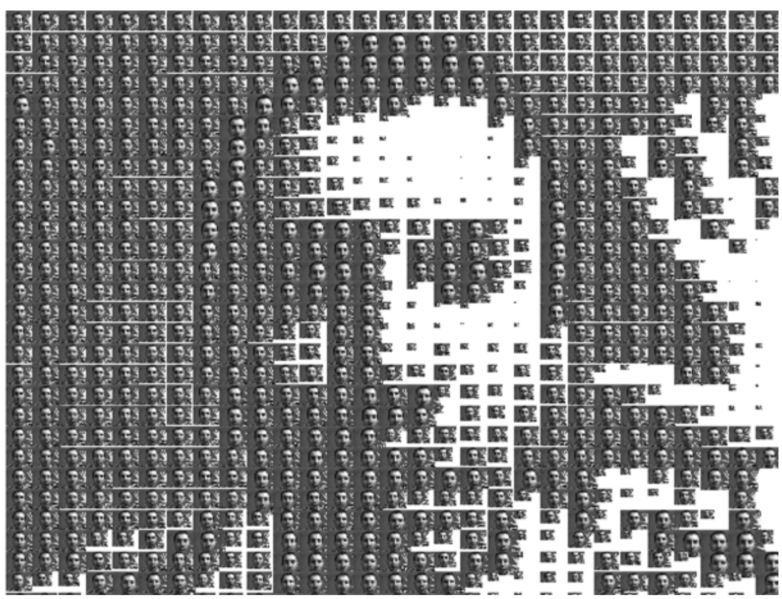
Figure 3.6: A direct mapping from gray value to line orientation.



3.4 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 3.7 shows a visual experiment using size as a mapping parameter. A gray-scaled image is used as input but also as synthesis object.

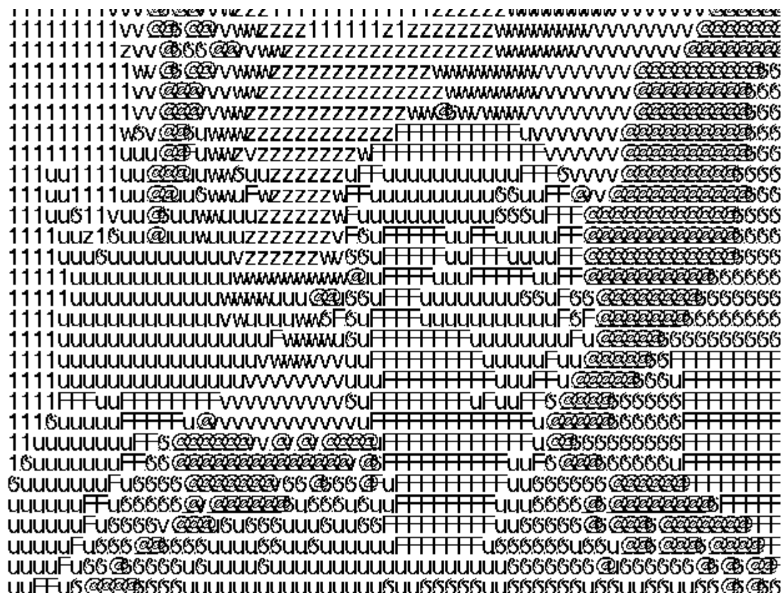
Figure 3.7: Replacing each region with an image of size proportional to the original gray-scaled value.



3.5 Shape

Similar shapes can be perceived as belonging to the same group. Direct substitution of gray value to a different character is shown in figure 3.8. The assignment between character and gray value is done at random, without any consideration for the average darkness of each symbol.

Figure 3.8: Mapping arbitrary characters to different gray-scaled values.



3.6 Random Approaches

Visually interesting results can be obtained using the luminosity information of the input frame to control the amount of random distortion applied to different parameters in the synthesis stage. Figure 3.9 shows an array of squares where the position and orientation of each square is randomly altered depending on the gray value of the input frame. Figure 3.10 shows a mesh where every node is displaced at a random direction. The magnitude of the displacement depends on the luminance values of the input image.

Figure 3.9: Array of squares. The rotation and displacement of each square is driven by the input frame.

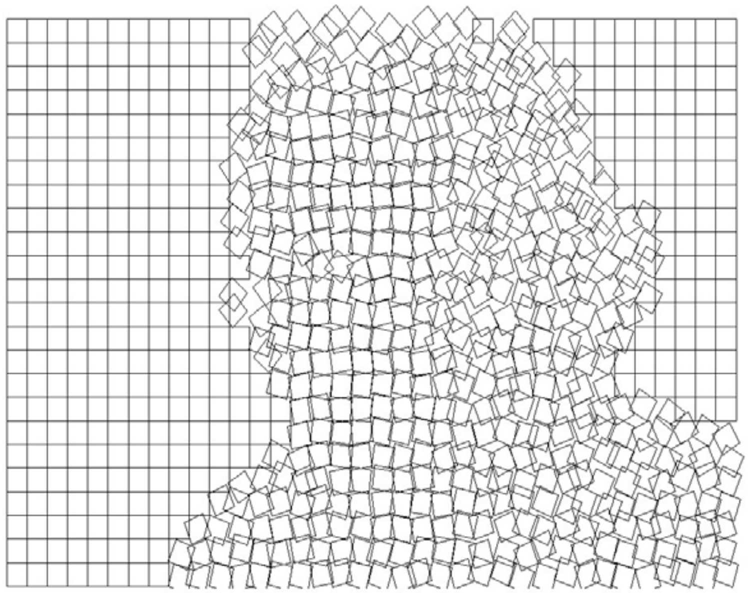
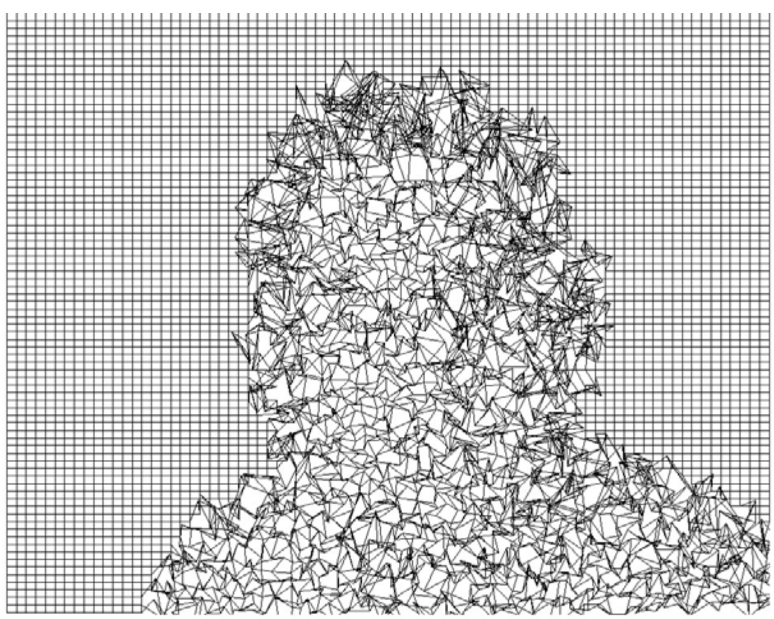


Figure 3.10: The magnitude of the node displacement on the net is proportional to the luminosity of the input frame.



3.7 Density Based Approaches

We perceptually separate regions that have similar element density [70]. 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 [64]. In figure 3.11, the dot density is changed to recreate the gray levels of the input.

Figure 3.11: Using the density of dots as a way to group different regions and to create darker and brighter areas.

3.8 Connectedness

As stated by Kaufmann [70], connectedness is a strong grouping principle, small differences in the way that points are connected can drastically change the look of the result. Figure 3.12 shows different ways of using the output of a dithering algorithm to join the resultant black points with lines following different paths. In Figure 3.13, the path is smoothed using Fourier descriptors.

Figure 3.12: Different ways to connect the same set of points.

3.9 Chapter Summary

In this chapter, different alternatives have been used to redraw the images that were described. These different options can be based on regions or edges. It is possible to use the synthesis objects in different forms to recreate the original regions, manipulating their shape, size, orientation or density. Also, parametric curves can be used to represent the original edges of the image. In this chapter, the Hough transform was suggested as an algorithm to find a set of parametric curves. Although conceptually the Hough transform can easily be extended to all kinds of parametric forms, in practice the dimensions of the search space increases with every new parameter, making it infeasible for elaborated forms. When any of the techniques shown in this chapter are used independently on the frames of a video sequence, the resulting video suffers the problem of lack of temporal coherence. The characteristics of the synthesis objects can change suddenly between

Figure 3.13: Smoothing the path of the connected points.

consecutive frames. Quick changes of spatial position are maybe the ones that are more distracting. In the next chapter, strategies to overcome the problem of lack of coherence will be presented. Those strategies can be combined with the different mappings that have been described in this chapter.

Chapter 4

Temporal Coherence

Most of the different visual mappings that were shown in the previous chapter have the same problem when used on videos, lack of temporal coherence. Temporal coherence is lost when frames are analyzed independently because the parameters of objects in adjacent frames can change abruptly. Even more so, the number of synthesis objects can be totally different in consecutive frames. This rapid appearing and disappearing of objects destroys the effect of motion at the local level and generates a disturbing artifact. It is possible to find different approaches to solve the problem of temporal coherence in the literature of non-photorealistic rendering (NPR). However, there is no perfect solution. There are conflicting goals and every approach requires some trade-offs. Bénard et al. on their state of the art review [9] defined three different goals for a time coherent animation:

- Flatness: It is desired that the resulting image looks flat and the characteristics of the synthesis element do not change when the perspective of the elements on the scene changes.

- Temporal continuity: The attributes of the primitives (particularly their position) should change smoothly.

- Motion coherence: There must be a correlation between the motion of the elements on the scene and the motion of the primitives

From these three goals, the one that is the most concerning to this dissertation is temporal continuity. Flatness is not an issue since in this project the input always comes from live action video, rather than from a 3D model. Motion coherence, as will be shown, can be used as a narrative element and it will never be too low to produce sliding artifacts.

A new trade-off is introduced by the nature of the techniques presented in this chapter: the quality of representation. One side effect of the alternatives used to guarantee time continuity can be the loss of precision in which the primitives represent the input image.

Three different algorithms were designed to generate time coherent A/S mappings from video data. The next section describes the first strategy based on primitive matching in consecutive frames. The second one creates a surface that is used to evolve the parameters of the primitives. The final one uses the elements detected on each frame as attractors for the objects on the previous one.

4.1 Matching and Interpolation on the Parameters Domain

In order to create a coherent animation, the objects on consecutive frames have to be paired. Different criteria can be used to pair objects on consecutive frames.

4.1.1 Minimizing the Sum of Distances

When one element in a frame is matched with an element on the next frame there is some associated cost (e.g.,the euclidean distance between the two elements on the parameters domain). If the number of objects on every frame is the same, the assignment of elements that produces the global minimum sum can be found with a well know combinatorial optimization algorithm known as the general linear assignment problem [15]. The constraint on having the same number of elements can be weakened by allowing objects of null size (invisible objects) to exist on the less populated frames. This strategy is similar to the one followed in the seminal work of McAulay-Quatieri [49] to handle the death and birth of sinusoidal partials on adjacent audio frames. In their algorithm, they allowed sinusoidal partials of zero amplitude to be matched in previous and subsequent frames. With this taken into consideration, the object matching problem can be stated as a linear assignment problem as follows:

Given a cost matrix $\mathbf{C}[n, n+1]$ of size $M \times M$ with elements C_{ij} that represents the euclidean distance between the parameters of every object i on frame n to every object j on frame $n+1$. That is:

$$\mathbf{C}_{i,j} = d\left(\mathbf{z}_i[n], \mathbf{z}_j[n+1]\right) \ . \tag{4.1}$$

We want to find the assignment matrix $\mathbf{X}[n]$ with elements X_{ij} defined as:

$$\mathbf{X}_{ij} = \begin{cases} 1, & \text{if object } i \text{ on frame } n \text{ is matched} \\ & \quad \text{to object } j \text{ on frame } n+1; \\ 0, & \text{Otherwise} \end{cases}$$

Such that the sum:

$$\sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} X_{ij} \ . \tag{4.2}$$

Is minimized.

Subject to:

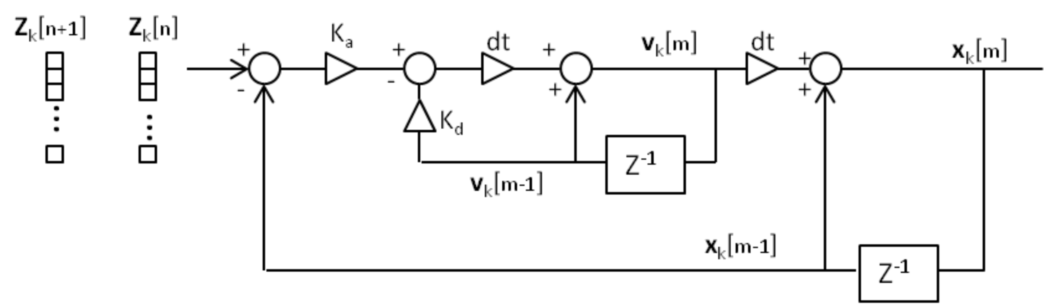
$$\sum_{j=1}^{M} X_{ij} = 1 \qquad\qquad (i = 1.2...M) \ .$$

$$\sum_{i=1}^{M} X_{ij} = 1 \qquad\qquad (j = 1.2...M) \ .$$

$$X_{ij} \in \{0, 1\} \qquad\qquad (i, j = 1.2...M) \ .$$

This is a standard optimization problem known as the linear assignment problem, many alternatives for solving this problem efficiently can be found in the literature [15]. After solving this assignment problem, a time coherent animation can be produced by generating intermediate images between every pair of frames with the interpolated parameters of matched objects. This interpolation is combined with a Euler integration scheme to allow physically inspired dynamics on the parameter evolution. Figure 4.1 shows how intermediate parameter vectors are obtained.

Figure 4.1: The parameter interpolation process.



The k^{th} element on input frame n is represented by the vector of parameters $\mathbf{z}_k[n]$. In the parameters space every object is treated as a particle. Every particle is attracted to a target position by an attraction force defined by the vector that joints the current and target positions. This force is then integrated twice in an Euler integration scheme to obtain the velocity $\mathbf{v}_k[m]$ and new position $\mathbf{x}_k[m]$. The variable dt is the constant that determines the value of the integration step. A different index variable is used since the number of frames at the output is larger than at the input. Note that if the interpolation factor is called I_f, $n = m/I_f \Leftrightarrow m/I_f \in Z$, then the dynamics of the motion of the particles can be controlled by changing the acceleration and damping constants K_a and K_d. After the desired number of intermediate images is generated, the input $\mathbf{z}_k[n]$ is replaced by $\mathbf{z}_k[n+1]$. The equations that describe these dynamics are:

$$F = (\mathbf{z}_k[n] - \mathbf{x}_k[m-1])K_a - \mathbf{v}_k[m-1]K_d; . \tag{4.3}$$

$$\mathbf{v}_k[m] = \mathbf{v}_k[m-1] + F dt \ . \tag{4.4}$$

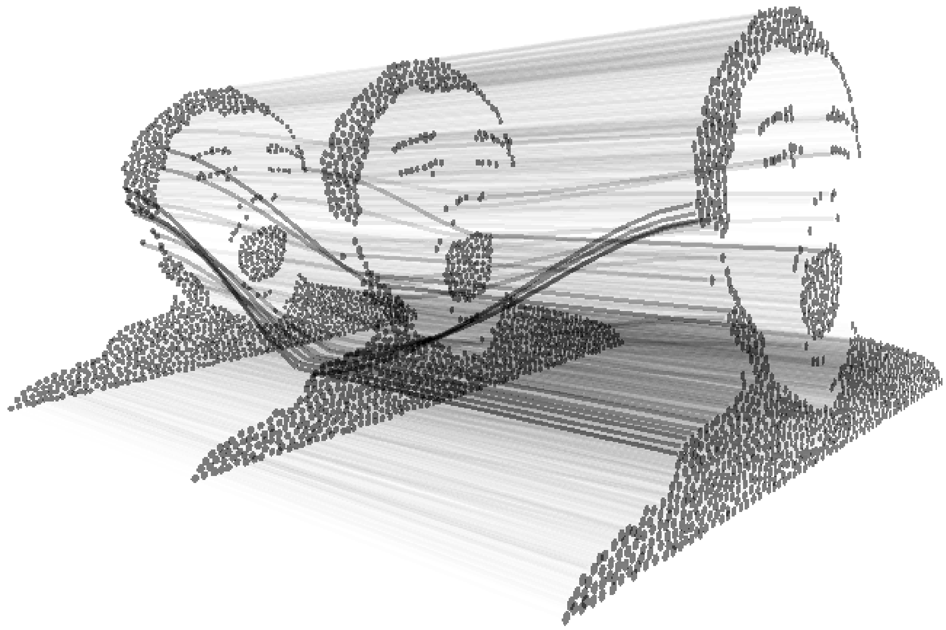
$$\mathbf{x}_k[m] = \mathbf{x}_k[m-1] + \mathbf{v}_k[m]dt; . \tag{4.5}$$

Figure 4.2 shows three frames of an animation where for simplicity of illustration, the synthesis objects are circles with fixed color and size. Every object can be described by a vector of only two parameters (x and y position) and the matching between the objects is done using the minimum sum criterion. Larger distances are plotted with a darker tone. Note how most of the matchings are done between near objects. The curvature of the lines is produced by the interaction of the acceleration and damping constants.

4.1.2 Minimizing the Maximum Error

If the matching criterion is not to minimize the sum of distances but to make the maximum distance as small as possible, the assignment problem is transformed to the bottleneck assignment problem, a very well known variation of the original assignment problem for which efficient algorithms also exist[15]. The problem now

Figure 4.2: Three consecutive frames and the lines joining the matched objects. Darkness of the lines is proportional to the distance.

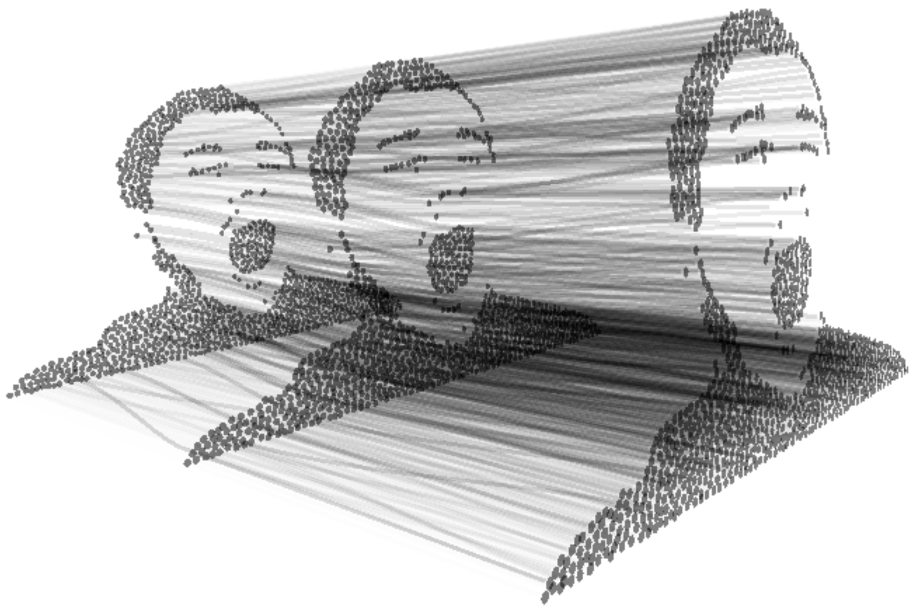


is to minimize the maximum of $C_{ij}X_{ij}$ $\forall i, j = 1 \cdots M$ with the same constrains and definitions presented in the previous section. Figure 4.3 shows the result of this bottleneck assignment with the same three frames. Note that now many matches have significant distances (darker lines), but there are no big jumps.

4.1.3 Direction Filtering

The animations created with the two matching approaches presented above have different dynamics, and each one can be used in different creative contexts. In addition, small changes on the cost matrix $\mathbf{C}[n, n+1]$ can change the animation behavior even more. For example, in two-dimensional vectors certain matchings

Figure 4.3: Same three frames as the previous example but the matchings are done using the minimum maximum criterion. There are many dark lines. However, there are no big jumps.

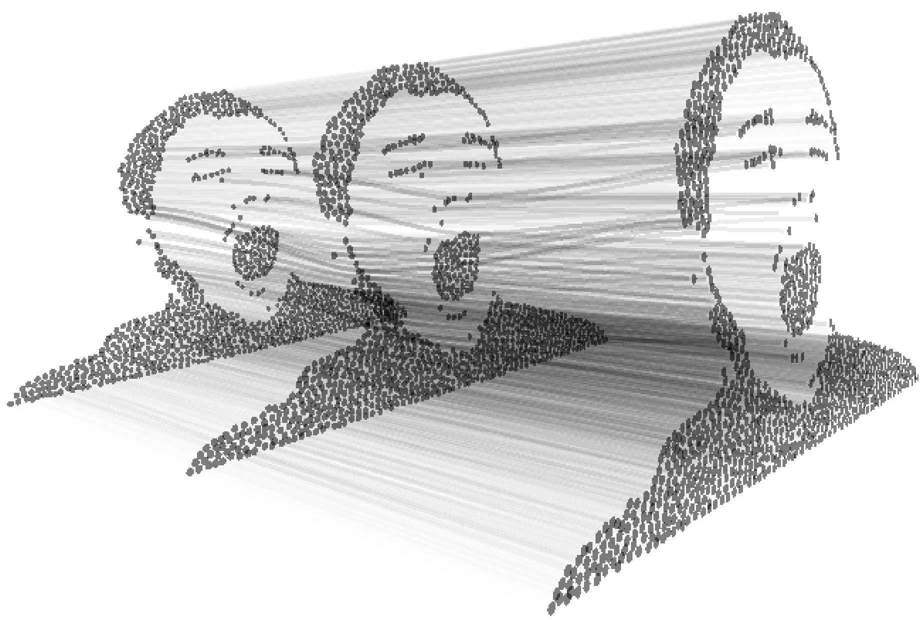


can be penalized if they produce movements that are not temporally or spatially smooth. Redefining the cost matrix in equation 4.1:

$$C_{i,j} = d\left(\mathbf{z}_i[n], \mathbf{z}_j[n+1]\right) + \lambda \arccos\left(\langle V_1, V_2 \rangle\right) \tag{4.6}$$

The second term $\lambda \arccos\left(\langle V_1, V_2 \rangle\right)$ is larger when the direction of the unitary vector V_1 differs more from the direction of V_2 (λ is just a weighing factor). These two vectors can represent the previous and current direction of movement. If a matching would produce a drastic change of direction, this matching is penalized and at the end the temporal behavior is smoothed out. On the other hand, one of the vectors can represent the average direction of the closest objects, and the other one, the direction that a matching between i and j would create. In this case, the effect of the term is to produce a spatial smoothing (closer objects tend to move in similar directions) generating a more cohesive animation. Figure 4.4 shows the three frames again with the resulting assignments after applying these two smoothing strategies.

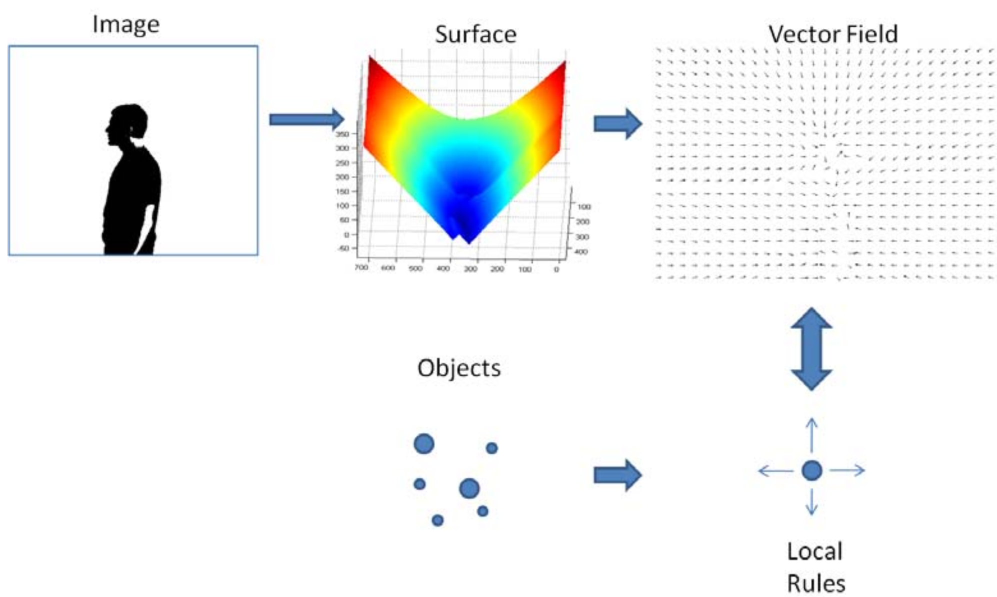
Figure 4.4: Penalizing the matchings that would produce strong (temporal or spatial) changes of direction.



4.2 Gradient Based Approach

Instead of forcing the matching of independently detected objects, these objects can simply be updated according to the contents of each video frame. The local behavior of the synthesis objects is defined by an internal set of rules. Global constraints on the objects motion are determined by the input image from video data. The input image is used to create a surface. The gradient of this surface is used to generate a vector field of forces that will affect the motion of the synthesis elements. Figure 4.5 shows a general diagram of this approach.

Figure 4.5: The generative approach to time coherent animations.



This approach resembles in some way the practice of generative art [28], where a set of rules is defined over a collection of elements that then behave with some degree of autonomy.

4.2.1 Example 1: A Dynamic Mesh Grid

Figure 4.6 shows a vector field, in this case generated directly form the gradient of a gray-scaled image. This vector field is used to attract the nodes of a grid to the dark areas of the input image and repel them from the bright ones. The resulting image is a grid that reassembles the video input, see Figure 4.7. Internal forces of the mesh, such as tension or drag can be modified to produce different visual results.

Figure 4.6: The vector field obtained with the gradient of a gray-scaled image.

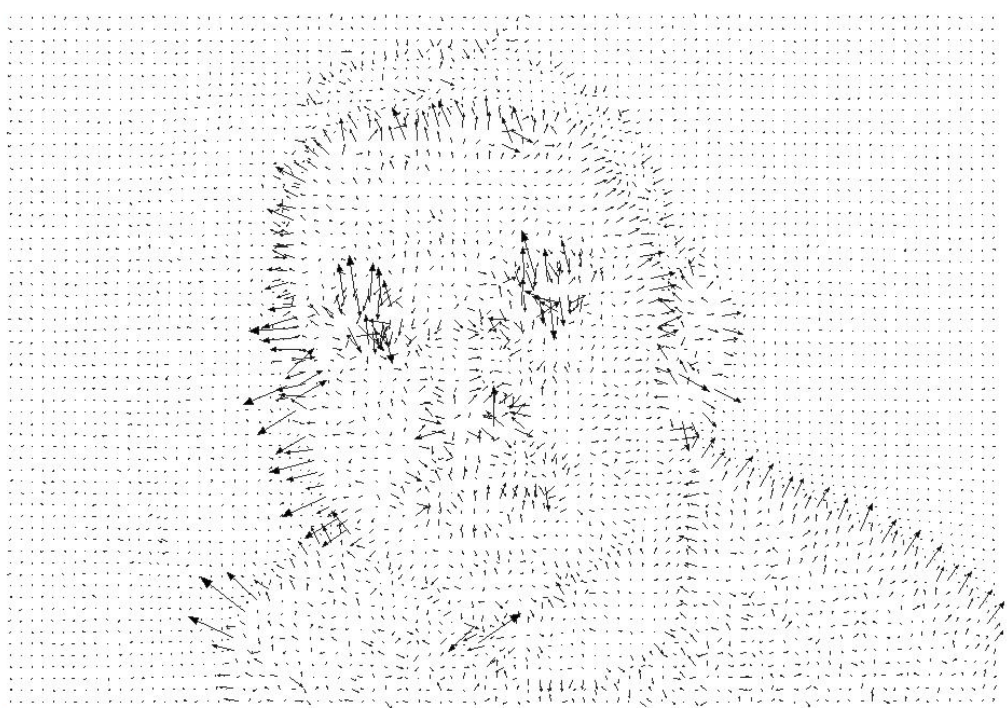
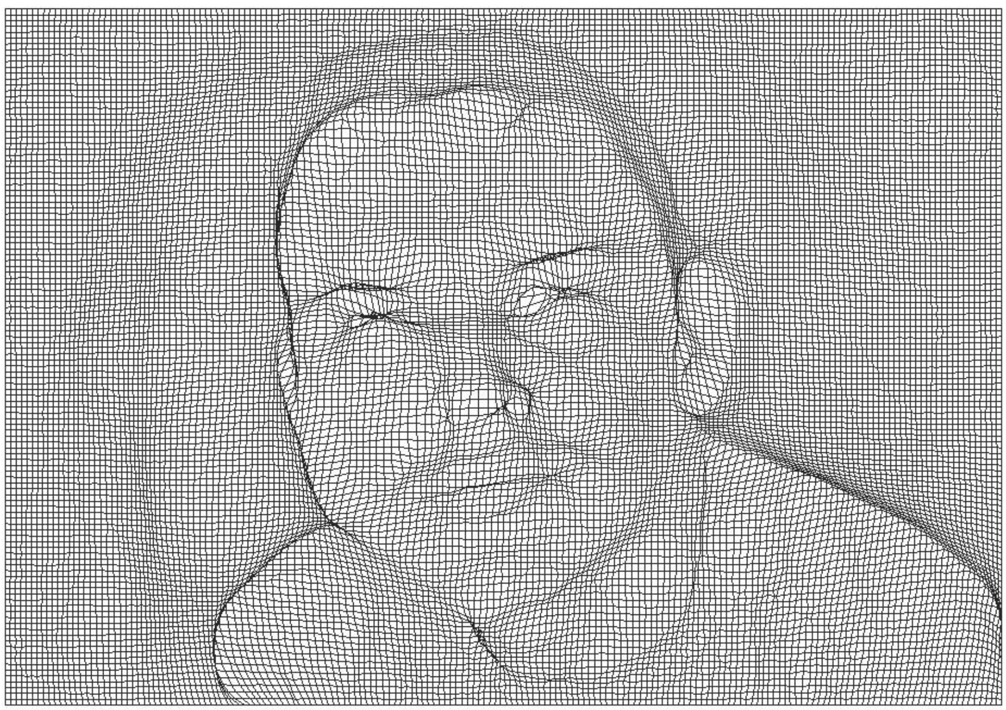


Figure 4.7: Mesh grid generated with the vector field.



4.2.2 Example 2: Circle Packing

Figure 4.8 shows a sequence of frames, where a set of circles of two different sizes are being attracted by the black areas of a binary video. In order to produce this attraction, the black and white image is used to generate a surface by means of the distance transform. If we define the black pixels of the binary image I, as foreground F and the white ones as background B, then the value of a pixel p on the distance transform image is defined by:

$$Dt\left(p\right) = \underset{q \in F}{min}\left(d\left(p,q\right)\right) \ . \tag{4.7}$$

A surface is then calculated as:

$$S\left(p\right) = Dt\left(p\right) - \bar{D}t\left(p\right) \ . \tag{4.8}$$

Where $\bar{D}t\left(p\right)$ is the distance transform of the logic complement of image I evaluated at pixel p. The gradient of this surface combined with a circle packing algorithm will determine the motion of every circle. Figure 4.9 shows the black and white input image; the image $S\left(p\right)$ as gray-scaled; and the image $S\left(p\right)$ as a surface in 3D space.

Figure 4.8: Some frames of a gradient based animation. The circles started at random position and then they moved according to a set of rules and constrained by a surface created with the input image.

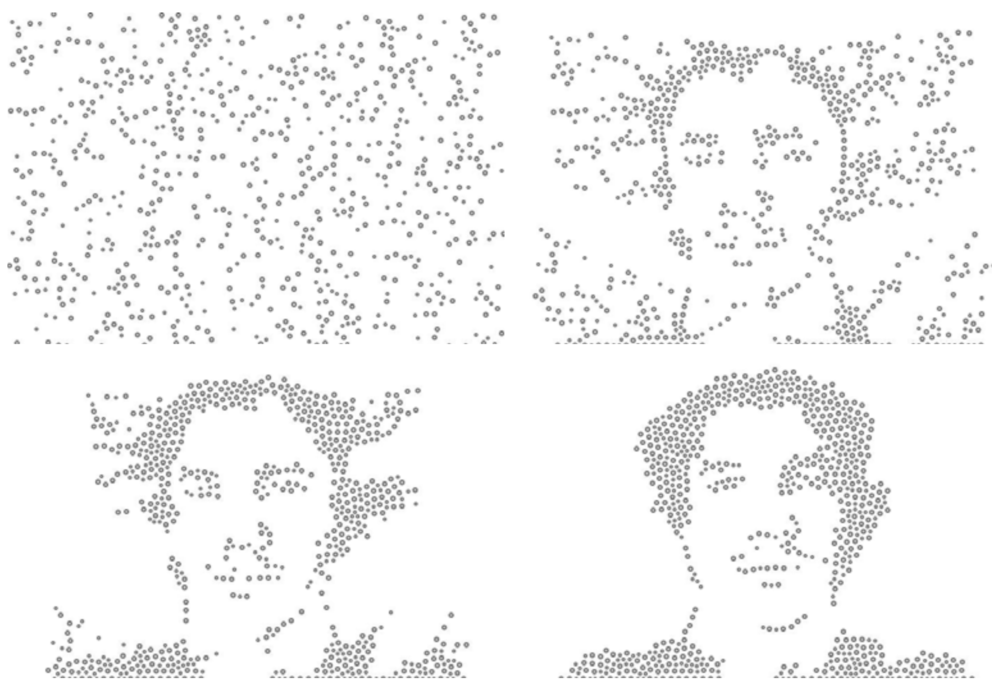
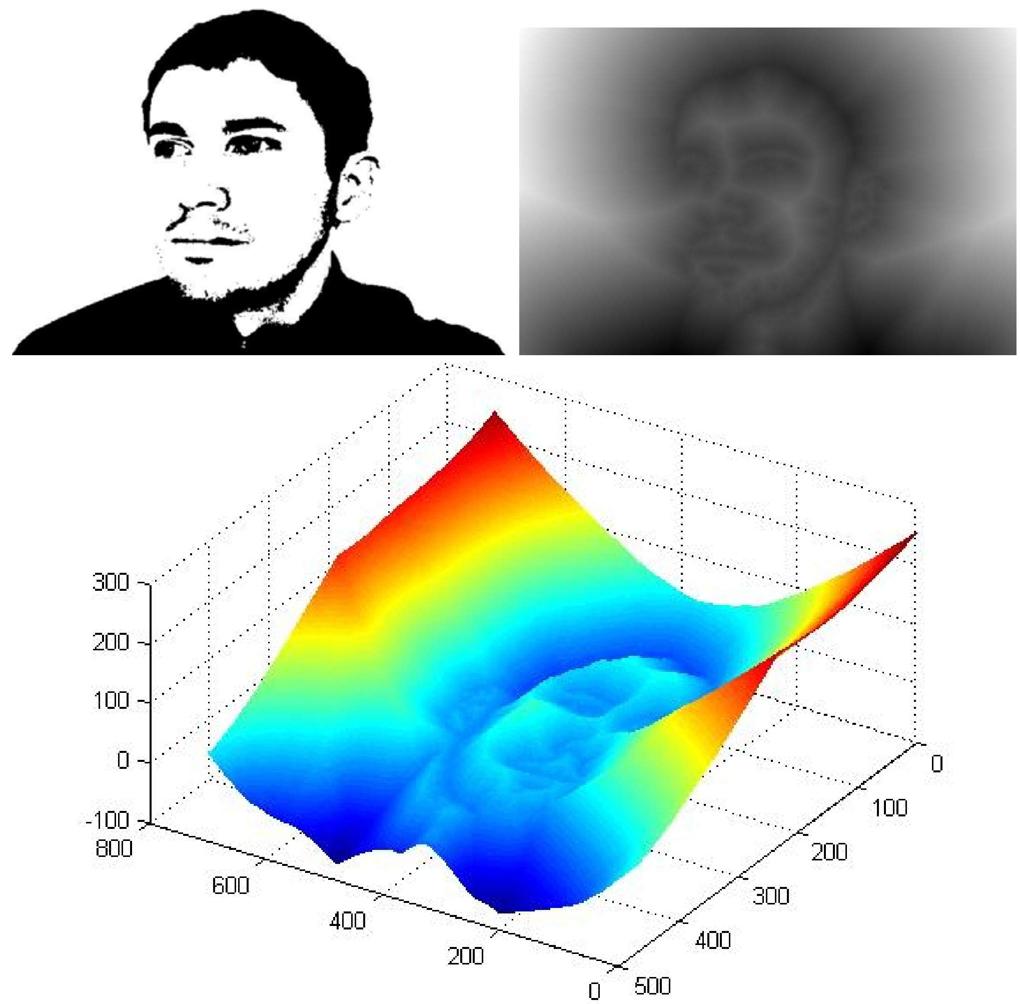


Figure 4.9: A black and white image in the upper left. The upper right is the surface created after applying the distance transform to it. The bottom image is the same surface in 3D space.



4.2.3 Example 3: Coherent Straight Lines

Almost the same set of steps used in the previous example can be applied on a different domain. For example to recreate the input image with straight lines moving with temporal coherence the following strategy can be used:

- Obtain the edge map of the image.

- Calculate the Hough transform using the parametrization from [24].

$$r(\theta) = x\cos(\theta) + y\sin(\theta)\ . \tag{4.9}$$

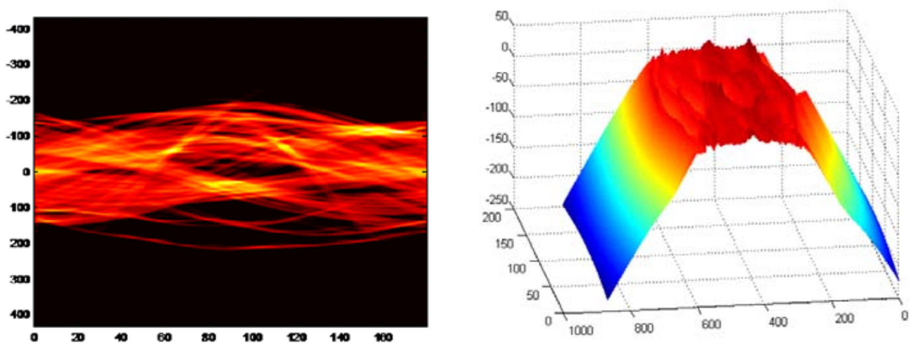
  (x, y) are the coordinates of pixels that belong to an edge.

- Low pass filter the Hough plane image, see left of Figure 4.10.

- Get a black and white image by thresholding the resulting image from the previous step.

- Generate a surface using the distance transform as shown in equations 4.7,4.8. See right of Figure 4.10.

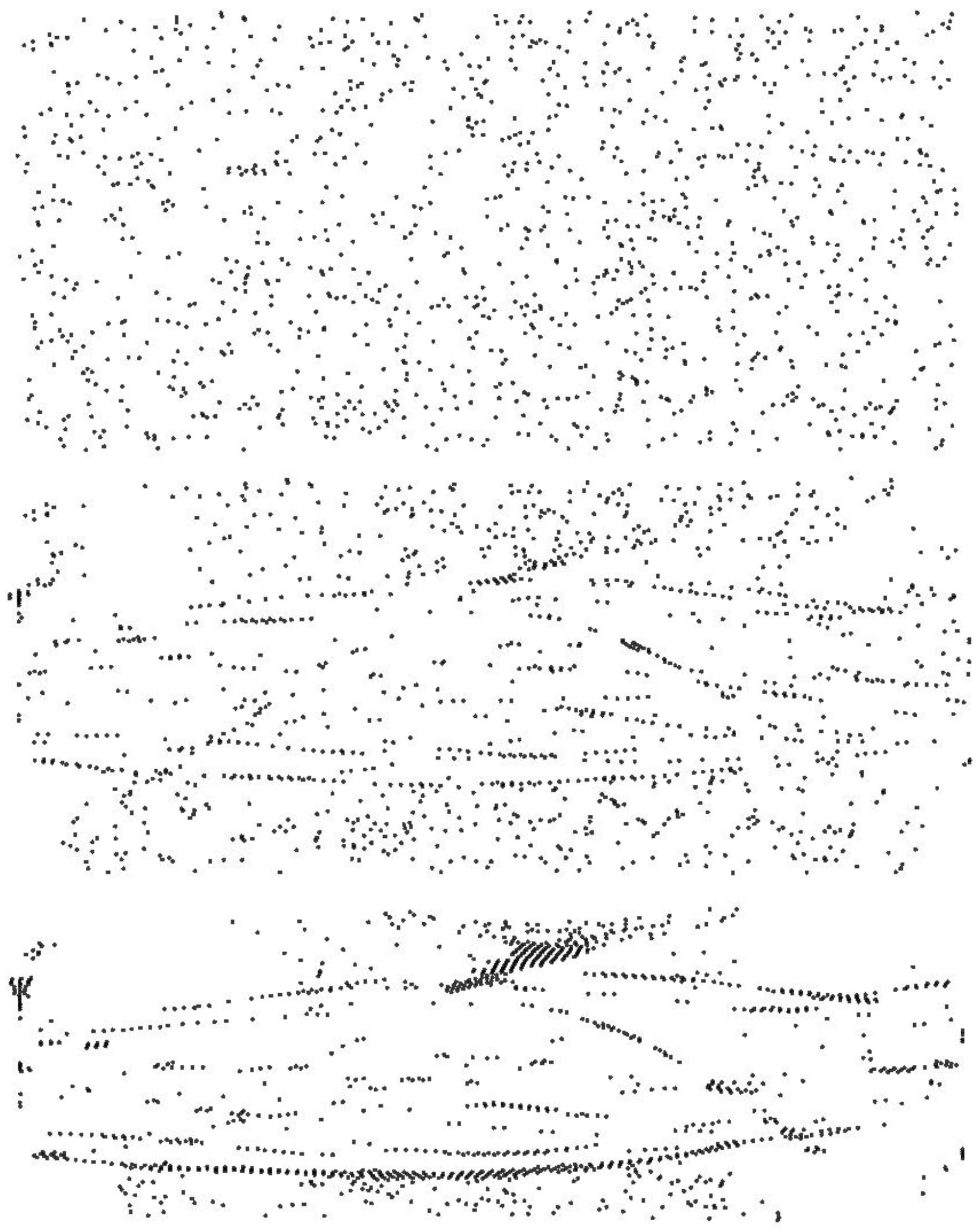
Figure 4.10: On the left, the low-pass filtered version of the Hough transform of an edge image. On the right, the surface created applying the distance transform to the Hough plane.



- Generate set of random points on Hough space uniformly distributed. Figure 4.11 (Top).

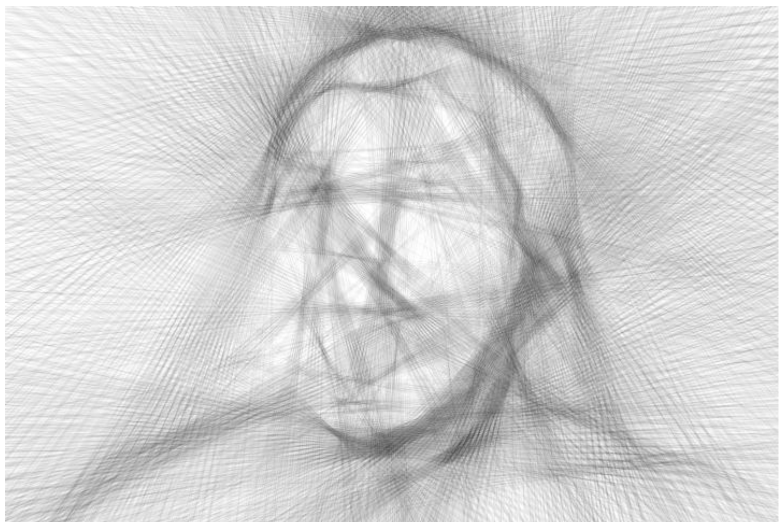
- Update the position of every point using the gradient of the distance transform and the circle packing algorithm to avoid point collisions. Figure 4.11 (Middle - Bottom).

Figure 4.11: On the top are random points uniformly distributed on the Hough plane. On the middle and bottom are dots in the Hough space moving in the direction of the peaks.



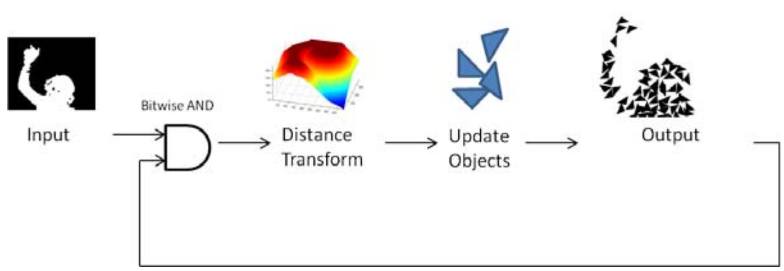
- Convert points on Hough space back to lines on the original domain. Figure 4.12

Figure 4.12: The image transformed back to the original domain. Every point on Figure 4.11 is now a line.



A modification that works well with silhouettes is to replace the circle packing algorithm by a feedback loop and calculate the distance transform of the error signal. See Figure 4.13.

Figure 4.13: The circle packing part of the algorithm can be replaced by a feedback loop.



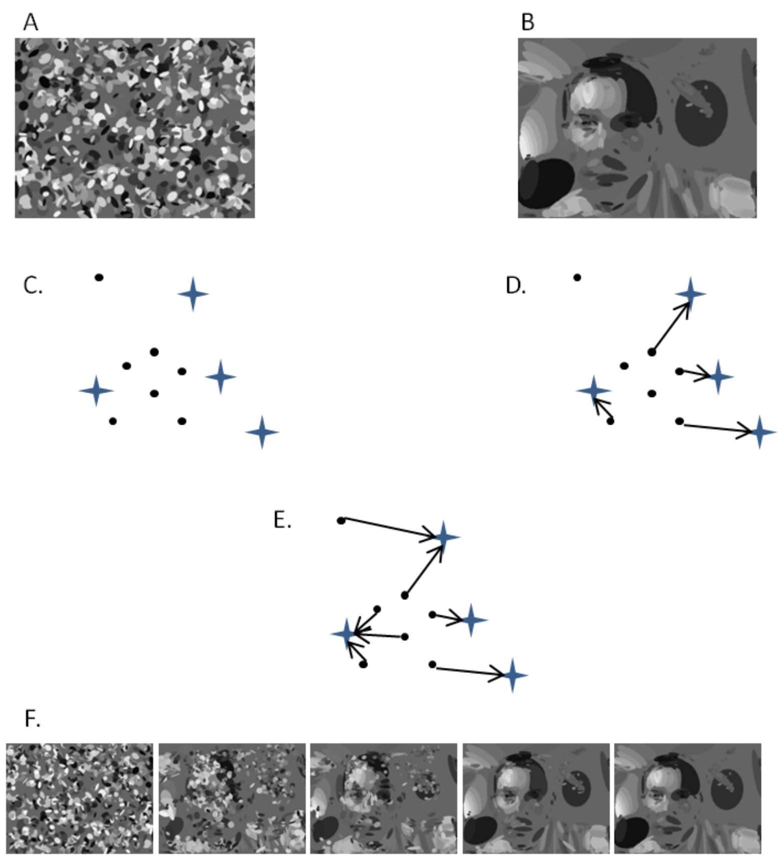
4.3 Attractors Based Algorithm

The algorithm presented in the previous section needs the creation of a surface on the parameters domain. The extrema of that surface are the target points, the points on the parameters space that represent the detected objects. When the extrema points are sparse, instead of creating a surface it is more convenient to use the limit points as attractors. The algorithm is described next using as an example a collection of ellipses as synthesis elements.

Each ellipse can be seen as a point in a six dimensional space (x-y position, width, height, angle and color). In figure 4.14A, a set of ellipses is placed at random and they are represented by the circles in Figure 4.14C. Lets call this collection of objects O_j, $j = 1 \ldots M$. Figure 4.14B shows the detected primitives. These ellipses were detected in the analysis stage. The connected region on different gray levels were fitted with ellipses. We call the detected primitives T_i, $i = 1 \ldots N$, with $M \geq N$. Those primitives are represented as stars in figure 4.14C. The algorithm proceeds as follows:

- Each of the target objects T_i is paired with the closest non assigned synthesis object O_j. The object is then marked as occupied. Since the number of synthesis objects is greater or equal than the number of targets, it is always possible to find a free object (Figure 4.14D).

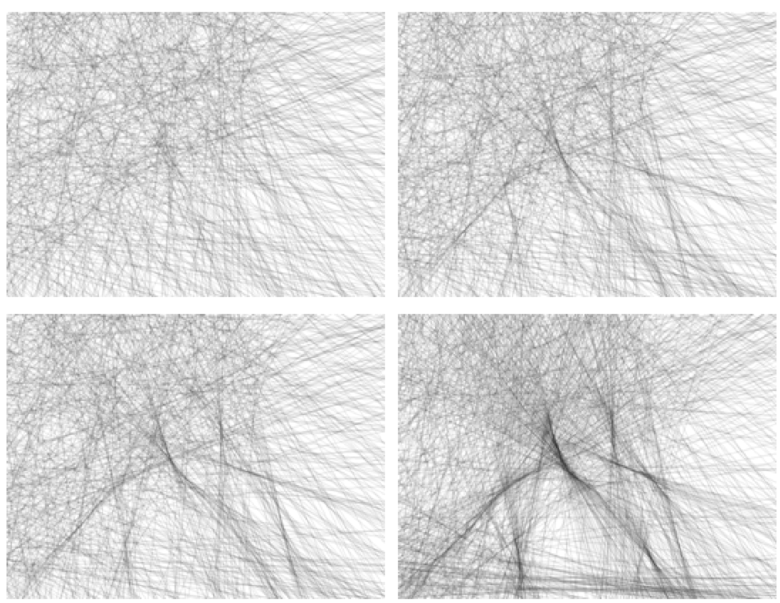
Figure 4.14: Attractor based algorithm. The objects that are the result of the analysis stage (Ellipses in B) are used as attractors for the current synthesis objects (Ellipses in A). The number of objects (circles in C, D and E) is always equal or greater than the number of attractors(stars in C, D and E). The algorithm guarantees that every attractor has at least one object assigned. F shows the elliptic objects moving towards their targets.



- After all the targets have been assigned, each of the remanent free objects O_j is assigned to the closest target, even if it is occupied(Figure 4.14E).
- An attraction force is applied on every object from its current position to its assigned target.
- New velocities and positions are calculated with an Euler integration scheme.

Figure 4.14F shows the evolution of the ellipses from random configurations to their targets. Figure 4.15 shows another example, again using the parametric representation of lines as a vector of parameters and the peaks of the Hough transform as the targets.

Figure 4.15: The attractors algorithm using the peaks of the Hough transform as targets.



4.4 Comparison

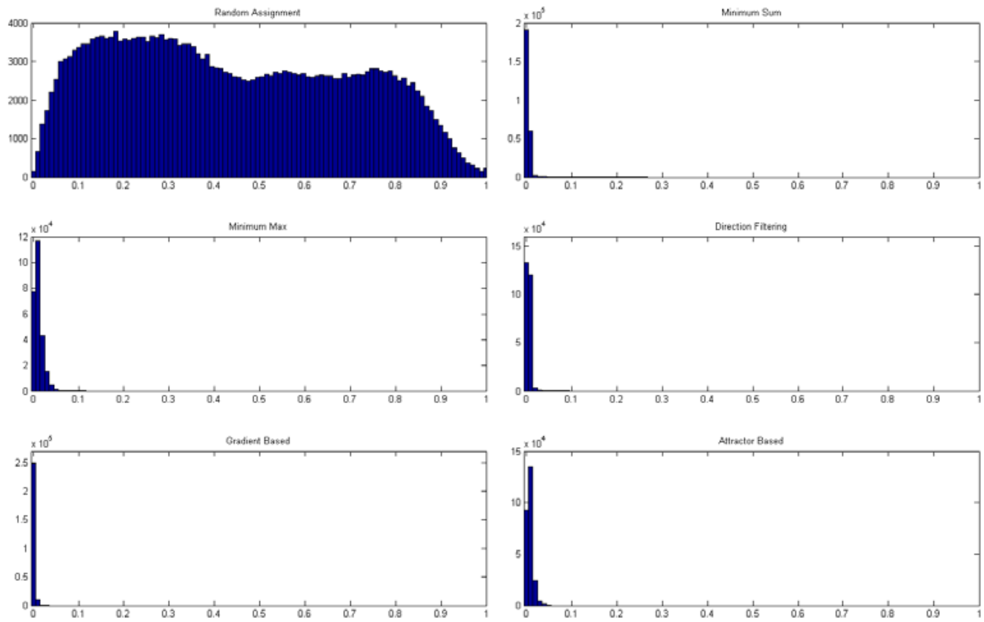
Three different strategies to solve the problem of temporal coherence when using A/S techniques were presented in this chapter. The first one uses combinatorial optimization techniques to match primitives in consecutive frames. The second one uses a gradient surface on the parameters space, and the third one uses the results of the analysis stage as attractors on parameters space. In order to evaluate the performance of the different algorithm and their variations on different aspects, all the techniques were applied to the same sequence, a slow-moving face. The synthesis elements are 2D objects, circles with the same size and color. Some quantitative measures can be obtained with comparison purposes:

4.4.1 Displacement Distribution

Small changes on object parameters in between frames are preferable over large changes. The different algorithms use different criteria to build the trajectories of the synthesis objects, thus the distribution of the amount of change is different in each case. Figure 4.16 shows the histogram of the magnitude of the displacement of every object on each algorithm.

With random assignments, all the jump sizes are equally probable. All the purposed algorithms tend to prefer the small jumps, being the gradient based algorithm the one that has smallest updates on the objects position. Minimum sum, direction filtering and attractor based algorithms behave in a very similar

Figure 4.16: The histogram of the distance that separates the consecutive position of each object. Each histogram represents a different algorithm.



way, although the minimum sum algorithm is more likely to find a small number of very large jumps. The minimum maximum algorithm on the other hand has a more sparse distribution of small jumps and it is less likely to have a big one.

4.4.2 Smoothness of the Matching

Temporal coherence is accomplished if there is continuity in the motion of objects, if they do not appear or disappear suddenly. But even if there is continuity, abrupt changes in direction or speed can look unnatural. Smoothness is definitely a desired characteristic in an animation sequence. One way that has been used to define smoothness in trajectories is to use the minimum jerk rule [68]. Jerk is the third derivative of position, natural movements (e.g., hand movements) tend to be planned to minimize the jerk. A cost on the jerk of a trajectory can be defined as:

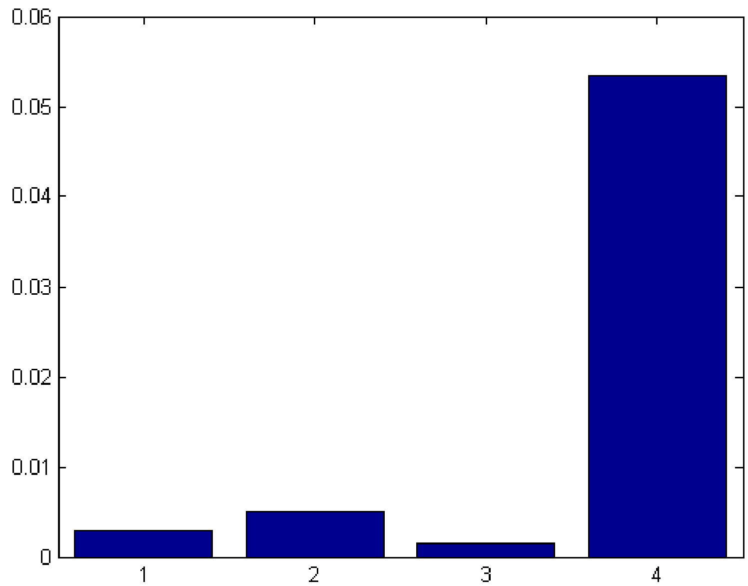
$$CJ = \sum_{n=1}^{N} \left(\dddot{x}\,[n]^2 + \dddot{y}\,[n]^2 \right) \tag{4.10}$$

Where: $\dot{x}\,[n] = x\,[n] - x\,[n-1]$; $\ddot{x}\,[n] = \dot{(\dot{x}\,[n])}$ and $\dddot{x}\,[n] = \dot{(\ddot{x}\,[n])}$

This cost can be averaged over all the trajectories in order to have a total cost for the whole sequence. The different matching strategies produce trajectories with different smoothness. The following graph shows the average jerk of four of the techniques, the gradient based algorithm is not considered here since the graph is comparing the jerk due only to matching and there is no explicit matching in the gradient based algorithm. The effects of the dynamics (attraction and damping forces) on the smoothness will be considered later.

All the algorithms have an important improvement over random assignments (not shown in the graph). The bar graph shows that the minimum sum matching produces a smoother animation than the minimum max assignment strategy. It also shows that this value can be reduced more with the temporal spatial filtering as suggested in section 4.1.3. The attractor based strategy is the less smooth of all the ones shown, that result can be expected since it is not optimal in any sense and it is sensitive to the ordering condition of the targets.

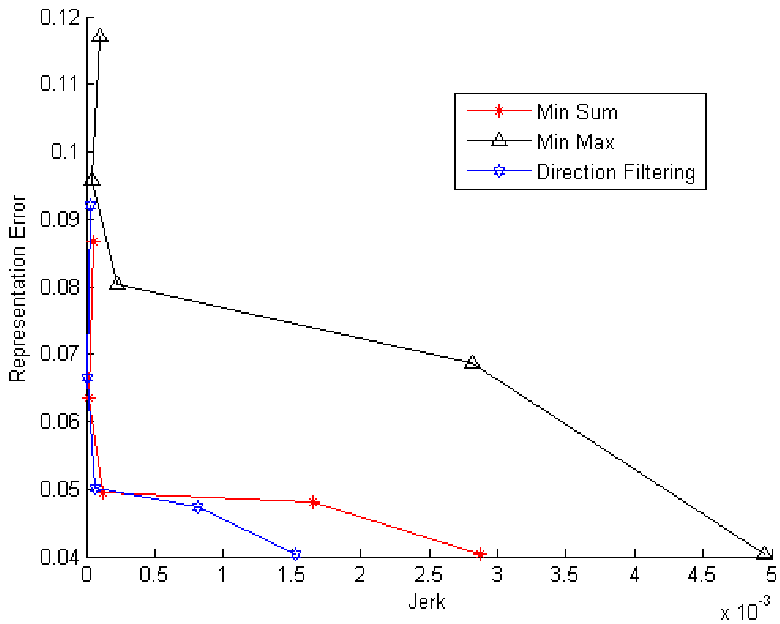
Figure 4.17: The smoothness of the algorithms with different matching strategies: 1) Minimum sum assignment, 2) Minimum maximum assignment, 3) Direction smoothing of(1) 4) Attractor based.



4.4.3 Smoothness Versus Representation

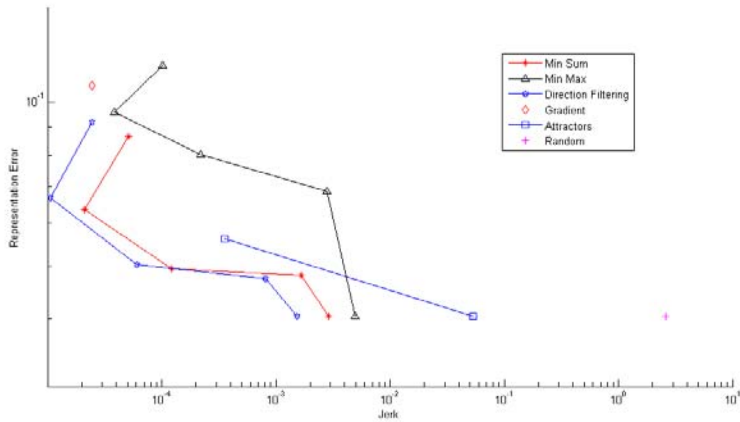
The motion of the objects can be altered by changing the dynamics of the update equations. The evolution on the primitives parameters can be made smoother, but by doing this, the quality of the represented image is affected. Figure 4.18 shows the jerk as a measure of smoothness versus the representation error. Both values were calculated over the same slow-moving head sequence. Jerk was calculated using equation 4.10 and the error is the number of pixels with the wrong value over the total number of pixels on the image.

Figure 4.18: Smoothness versus quality of representation. Different configurations for three of the algorithms, the graphs show the trade-off between smoothness and representation error.



The points with less representation error on each plot are the trajectories using only the matching information (same as in figure 4.17). On the other extreme, the points with low jerk and bigger error represent a configuration with low value of damping (see equations 4.3 to 4.5). The third point on each line, represents a good trade-off between the two aspects. It is evident that the worst behaved algorithm when evaluated with those two criteria is the minimum maximum algorithm. The direction filtering algorithm tends to work better than the minimum sum one, but they became similar on the ideal operation area. Figure 4.19 also compares the other algorithms now using a logarithmic scale.

Figure 4.19: Smoothness versus quality of representation. All the algorithms in different operation points.



The new graph shows that the gradient algorithm, although it tends to favor smoothness it compromises the quality of the representation. This result was expected given the problem of the algorithm with local minimum and sensitivity to starting conditions. The attractor based algorithm significantly improves smoothness when dynamics are incorporated. Also shown in the far right of the graph is the high jerk value that arbitrary matchings would generate.

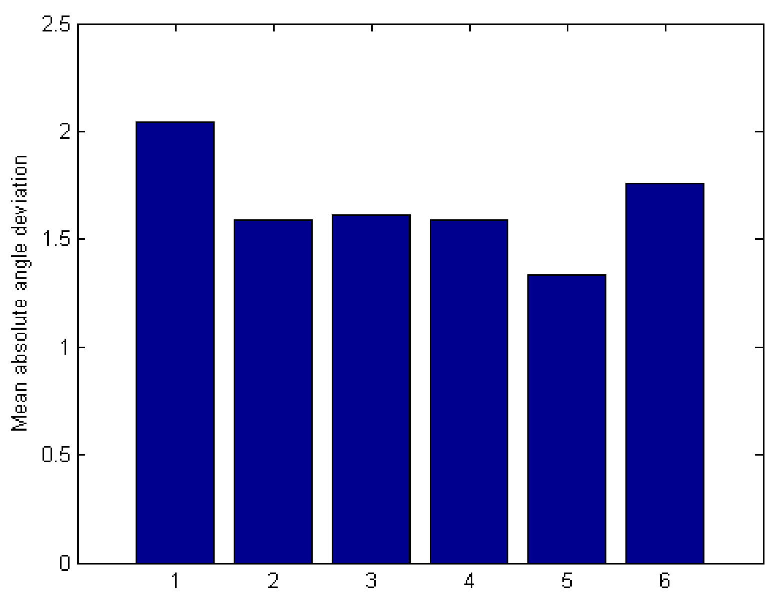
4.4.4 Cohesion

Maybe the most distracting characteristic of the proposed algorithms is the lack of cohesion of the synthesis elements in the results. The primitives tend to move independently from the ones that are spatially close. One way to measure the degree of dispersion is to calculate how far the direction of motion of an object is to the average direction of its neighbors. Equation 4.11 describes how this value was calculated.

$$DD = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \arccos\left(\left\langle V_{m,n}, \overline{V_{m,n}(d)} \right\rangle\right) \tag{4.11}$$

Where $\overline{V_{m,n}(d)}$ is the average direction of the d closest objects to element m of frame n. Lets call this measurement the direction dispersion (DD). Figure 4.20 shows the result of the DD for the five algorithms and random assignment. The original sequence is the same slow-moving head that was used in the previous measurements.

Figure 4.20: The direction dispersion for: 1)Random assignment. 2) Minimum Sum. 3) Minimum Maximum 4) Direction Filtering 5) Gradient 6) Attractors



The graph shows that the gradient algorithm is the one that has smallest variation. By design, in the gradient algorithm, spatially closer objects will be placed on the same area of the surface and they will tend to move in the same direction. On the other hand, the two first combinatorial optimization strategies (minimum sum and minimum maximum) behave similarly and perform much worse than the gradient. The graph also reveals that the direction filtering algorithm does not significatively improve the results of the combinatorial optimization techniques and shows that the attractors technique is the one in which the dispersion is worse.

4.5 Chapter Summary

Three novel approaches to generate time coherent animation from video sequences using analysis and synthesis techniques were presented. The first one uses the matching information of objects in adjacent frames to create smooth transitions or to generate interpolated objects in new in-between frames. The matching is done using variations of combinatorial optimization algorithms. When the matching is performed using the minimum sum criterion, the results in terms of trade-off between smoothness and quality of representation and the cohesion of the synthesis elements are numerically better than the ones obtained with the minimum maximum approach. As a downside, the minimum sum approach is likely to have few but very large and thus distracting jumps in the position of the objects. A direction filtering scheme was designed in order to improve the smoothness and cohesion of the assignment algorithms. The filtering operates by penalizing in the cost matrix the matchings that will produce more dramatic changes on direction or the assignments that would generate a direction that diverges from the average direction of the closest objects. This filtering indeed improves the smoothness but changes in cohesion were not significant. The combinatorial optimization approaches can not be done in real time since the solution of the linear assignment problem and the bottleneck assignment problem are computationally expensive. The resulting animation will always be at a slower rate than the original video since intermediate frames need to be generated. The algorithms were illustrated with synthesis objects that can be described on a two-dimensional parametric space, but the linear assignment and bottleneck assignment strategies can easily be extended to higher dimensional spaces.

The second technique uses a gradient based approach consisting of the creation of a surface in which the synthesis objects can move smoothly. The objects have to be aware of themselves so they do not sink together into a local minimum. To avoid this colliding behavior, the gradient strategy is always combined together with a circle packing algorithm or a feedback loop. Since this approach is based on incremental updates of the objects parameters it is adequate for real-time implementation. In addition, it was demonstrated that this technique can be extended to parametric curves like straight lines, after a preliminary domain transformation, such as the Hough transform. This gradient strategy performs well in terms of smoothness and cohesion, but at the expense of compromising the quality of representation. The starting distribution of the objects can prevent that certain areas get filled correctly and the circle packing algorithm produces some visible artifact in the distribution of objects.

If the detection algorithm produces sparse points it is better to use a strategy based on attractors, that is the third algorithm presented here. The attractors based algorithm can be considered a sub-optimal alternative to the combinatorial optimization algorithms. It performs badly on smoothness and cohesion but it is adequate for real-time implementation and is easily configurable to handle different numbers of objects.

A different set of measurements were also presented at the end of the chapter. These types of measurements are not common in the non-photorealistc animation literature. This set of quantitative measurements constitutes another contribution to the present work. In spite of the numerical outcomes and from a creative point of view, the results obtained with the different strategies are visually dissimilar and potentially useful in diverse contexts.

Future work can concentrate in producing algorithms that generate animations that are more cohesive, where the synthesis elements behave more like a group. This is maybe the more distracting characteristic of the combinatorial optimization approaches presented here. Some effort can also be put in tuning the gradient algorithm to make it more accurate in the image representation by forcing uncovered areas to attract objects, even if these objects are already in stable positions.

Besides uses for narrative animation and interactive installations, other scenarios can be suggested. The possibilities of manipulation, creation, and gradual destruction of images are useful not only for narrative purposes but also for face perception studies or gesture amplification. The combinatorial optimization algorithms used in this work are well known, and are used in many different applications. The animations generated by this work are an unexpected contribution to the visualization of some variations of the linear assignment problem.

Chapter 5

Cross Synthesis and Ambiguous Images

As it was pointed in Chapter 2 one of the uses for A/S techniques is to mix two signals into a third one. Even in the audio domain, there are many interpretations on how to do that [57]. In images and videos, this concept is even less defined. One possible interpretation for cross-synthesis in images and videos emerges when the synthesis elements have an identity of their own. Then, it is possible to interpret an image in two different forms, depending on if the focus is on the local or in the global. There is a long tradition in the arts of exploring this approach. Figure 5.1 shows the work of the XVI century painter Giuseppe Arcimboldo, who mixed figures of fruits and vegetables to create portraits.

Figure 5.2 shows two classical examples from the early 1900s, “All is vanity” from the artist Charles E.Gilbert and “Society: A portrait” from George A. Whitherspoon. Dalí, also explored the possibilities of ambiguous images in paintings such as “The Mysterious Lip that Appeared on the Back of My Nurse”

Figure 5.1: Portrait of Rudolph II, Arcimboldo, 1591.

and "The Great Paranoiac" [11]. Other artists like Sandro del Petre and Octavio Ocampo are specialists in the creation of double images.

Automatic systems to generate these embedded images are not known by the author, neither examples of them on video. However, the controlled synthesis of natural scenes has been formally investigated. Prusinkiewicz and Mech [52] explored the generation of botanic structures with external environmental constraints. Anderson et al [4] and Xu et al [73] created flocks of birds and butterflies with the capability to align in predefined shapes.

In this chapter, the algorithms for temporal coherent synthesis that were presented in the previous chapter are extended to explore the concept of cross-synthesis on moving images. Three different alternatives are presented: the sub-

Figure 5.2: Ambiguous images "All is vanity" by Charles Gilgert (Left); "Society: A portrait" by George Witherspoon (Right).

stitution of synthesis elements by recognizable entities, the direct detection of those entities and the use of synthesis elements as generative seeds.

5.1 Substitution

A direct alternative to creating scenes that are meaningful in the local and in the global is to replace abstract synthesis objects with well known elements, such as animals that move in groups (school of fishes, flock of birds etc). The rules for the synthesis can be combined with rules that simulate swarming behavior[55]. Figure 5.3 shows a frame from the real-time installation "Ant Theater". The points of the output of a Shi-Tomasi corner detection [60] are used as attractors on algorithm 4.3. The points are then replaced by "ants" textures. The orientation of the velocity vector, from the time-coherent algorithm is used to guarantee that the “ants” are aligned with their direction of motion. The walking cycle of ants can be simplified to six different leg positions that alternate while the virtual creature is moving.

Figure 5.3: Ants taking the places of the outpout points of a Shi-Tomasi corner detector.

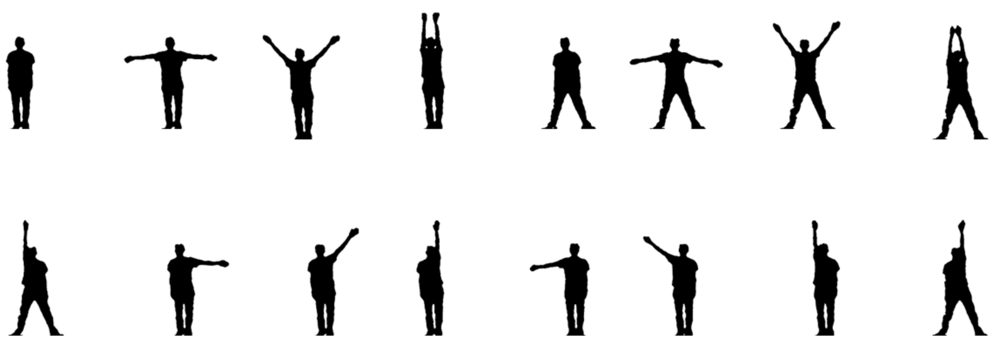
5.2 Direct Detection

More complicated synthesis elements can not simply substitute basic primitive shapes like points, circles or ellipses. Algorithms for the detection of the target virtual creatures in specific positions must be used. Before the detection, it has to be defined how much variation the synthesis element is allowed to have. Can the elements change position, orientation, scale and/or state?. All of these constrains define a model for the synthesis element, and dictate what to look for during the analysis stage. Depending on the specific circumstances, standard techniques for template matching can be used (Moments, Fourier descriptors, Correlation, etc). [14].

5.2.1 Template Detection

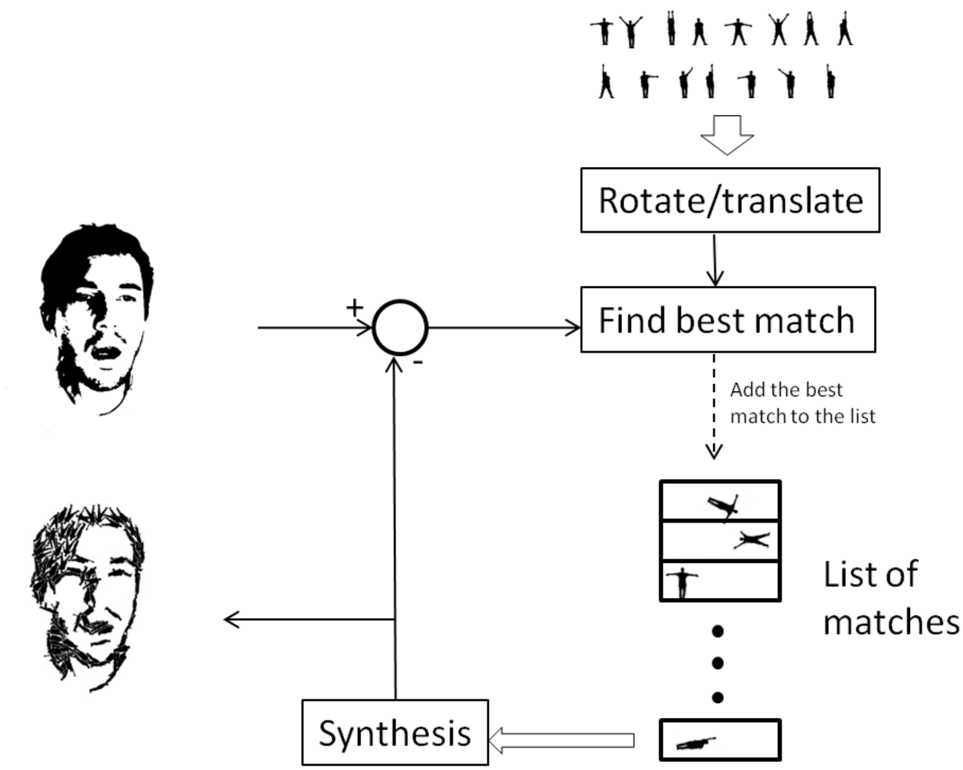
As an example, figure 5.4 shows different states of a synthesis element, a human silhouette.

Figure 5.4: Different positions of a human silhouette as synthesis elements.



Different translations and rotations of each silhouette are matched against a target image as it is shown in the diagram presented in figure 5.5. The best match is stored in a list and it is subtracted from the original image. The process is repeated until the desired number of element is placed.

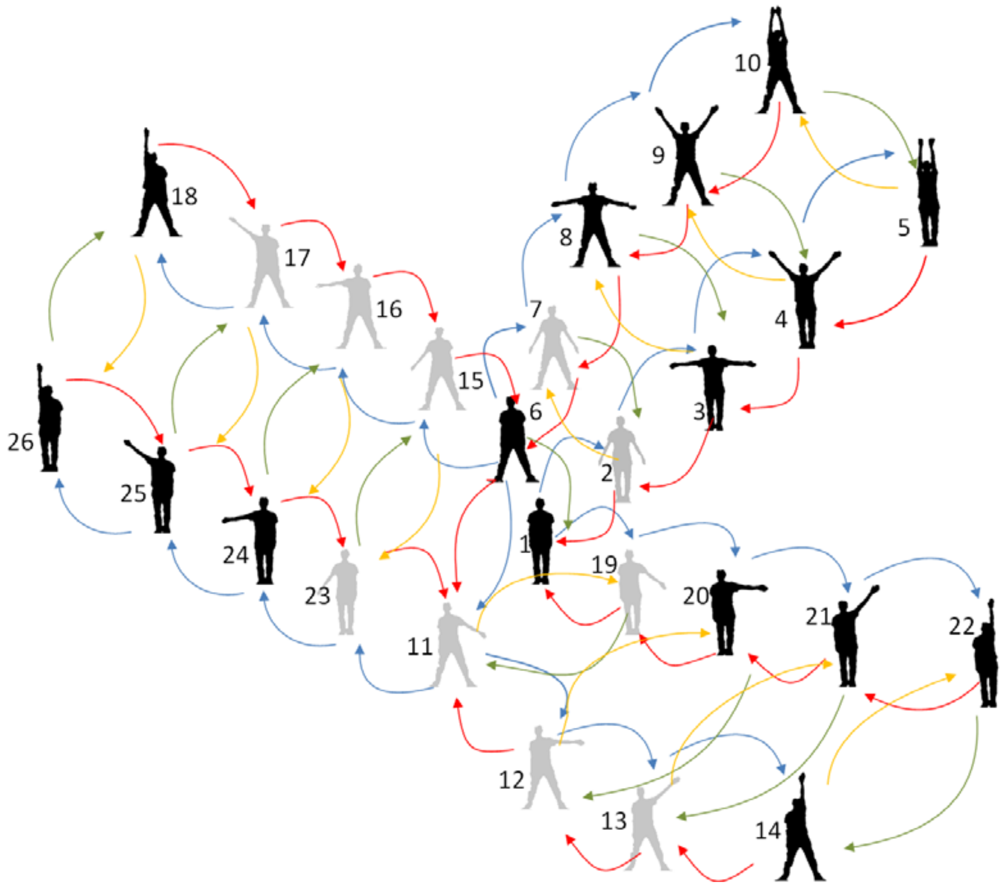
Figure 5.5: The extensive template matching algorithm.



5.2.2 Matching

Once the synthesis elements have been detected on every frame, one of the matching strategies discussed in section 4.1 can be used. In the case of the human silhouettes every element can be represented with four parameters: $x-y$ position, orientation and state. The fourth parameter is discrete and a euclidean distance between state indexes is not a representation of closeness. To create a measure of distance between two different states, a transition diagram is built. It is shown in figure 5.6.

Figure 5.6: The Markov chain model of the synthesis elements.

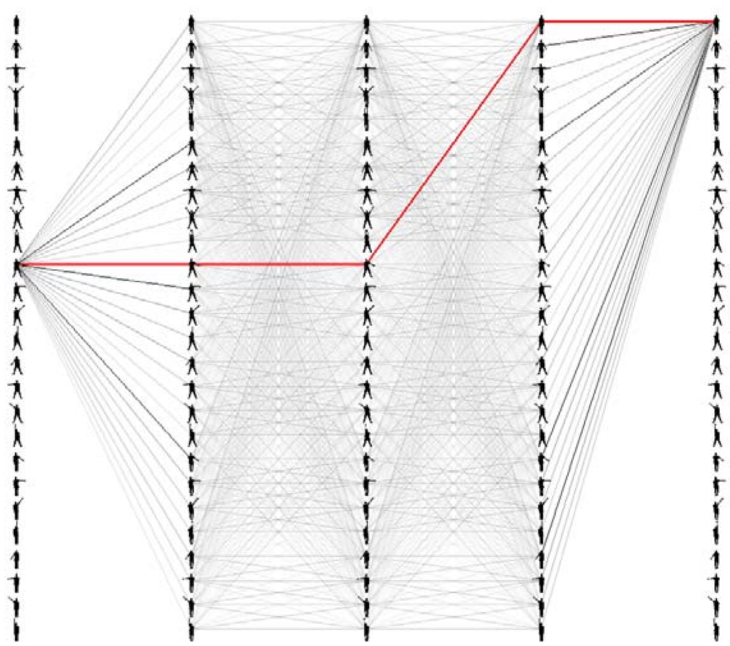


The distance between two states can be defined as the minimum number of transition that separates them. In figure 5.6 the distance between states 6 and 20 is 3 (with the sequences 6-1-19-20 or 6-11-12-20). Having a distance measure allows us to calculate the matching between frames using any of the combinatorial optimization approaches illustrated in section 4.1. However, what about the interpolation? How can we generate intermediate images between two states?

5.2.3 Interpolation

The continuous parameters can be interpolated with the Euler integration scheme shown in figure 4.1, but another strategy should be used for the discrete states. It is desirable that the figures change states smoothly when going from one state to another. One way to find a path, given two states and the wanted number of intermediate states is to use the Viterbi algorithm [67]. The Viterbi algorithm is a dynamic programing strategy that finds the more likely trajectory to get to every state on every time step. Only the less expensive path for every step is kept [45]. Figure 5.7 shows the trellis diagram of the model and the best trajectory between states 11 and 1.

Figure 5.7: Trellis diagram showing the best path beteween states 11 and 1.



Mixing the combinatorial matching strategies with the Viterbi interpolation process, it is possible to generate animations where in the local level the motion of the elements is consistent with a model, and in the global level, an external image is recreated.

5.3 Generative Approach

A different alternative for the generation of cross-synthetic images is to use the detected elements as a "seed " for a generative algorithm. This generative algorithm should create a known structure while keeping the global image identifiable. In "Herbaceous", the points with more positive error value from a dithering algorithm are used as the terminal nodes of a tree. The tree is then built from the nodes to the truck using the particle system of plant modeling described in [58]. (Figure 5.8).

5.4 Chapter Summary

Three methods for the creation of cross-synthesis scenes were described in this chapter. The first one is straightforward to implement, given the time coherent algorithms from the previous chapter. Since the abstract elements are usually simple shapes this approach is limited to elements of very simple behavior. The natural motion of living creatures also imposes some new conditions on the algo-

Figure 5.8: Using the results of the dithering algorithm as a generator.

rithms. Most creatures move with their body oriented towards the direction of displacement and they assume a totally different position while in rest. Depending on each case, the algorithms have to be modified to implement these conditions and new artifacts can appear. Creatures with more elaborated behavior can be animated with the second method. This method inherits the same cohesion problems of the combinatorial optimization algorithms for matching and it is limited to objects that can be described with a discrete states Markov model. The last method is useful to create more structured objects, but these new objects do not necessarily maintain time coherence.

Chapter 6

Examples

6.1 The Fitting Dance

The fitting dance is a short animation that illustrates some of the possibilities of A/S approaches and how to use them to support a narrative. It has been presented at different events, including the 13^{th} Japan Media Arts Festival (2009). The piece is divided into different phases:

- Creation: At the beginning there is only one circle on the screen moving according to the background music. Other circles start to appear quickly and together they form a face. The circles turn into ellipses aligned with the axis but then they are allowed to have an arbitrary angle. (Figure 6.1).
- Chorus: The contours of the ellipses are modified by the RMS value of the accompanying voice signal (Figure 6.2).
- Shape Quantization: The contours are transformed from continuous elliptic curves to regular polygons. The number of sides is reduced until only

Figure 6.1: First Phase: A face is created starting from a single circle

Figure 6.2: Second Phase: The contours are modified by the voice signal.

rectangles are used. The chorus is repeated using a square wave as a superimposed contour. The voice is altered to mimic the “artificiality” of the squared shapes on the audio signal. (Figure 6.3).

Figure 6.3: Third Phase: The contours are turned into straight edges. Ellipses become polygons.

- Spectral analogy and destruction: The complexity of the contours is modified according to the spectral richness of the music playing. Image and music are taken to a climax of spectral content until the fragments of the image explode, going back to the silence and empty space that started everything (Figure 6.4).

Figure 6.4: Final Phase: The spectral content of the contours mimics the audio. The animation ends with an explosion of the figure.

6.2 MeshFlow

"Meshflow" is a mirror-type video installation consisting of a camera pointing to the viewer and a software application that turns the camera image into a new visual and sonic representation. This installation was presented as a creative showcase in the 2010 edition of the International Conference on Advances on Computer Entertainment (ACE 2010). In "Meshflow," a set of points connected on a grid with almost no other property than its position on the screen, evolves continuously following physics laws and morphing into the image of the spectator. The motion of the points triggers a particle-based maraca synthesis algorithm that completes the experience.

In "Meshflow," the role of a grid is transformed. The grid is no longer used as a system of reference where points can be found, but it is converted into the active element that creates a mirror. The spectator sees how the mesh moves organically until it takes the form of whatever the camera is looking at. The motion is continuous and every frame evolves from the previous one. This motion represents the fact that what we are now is caused by a series of events, and our present actions will determine what our future will be. Changes in the position of the nodes on the mesh are used to generate sound. The algorithm chosen for the sound generation is a physical model for maraca synthesis [17]. A physical model based upon synthesis was chosen since it was important that the sound resembles in some way the one that can be heard when shaking a real object.

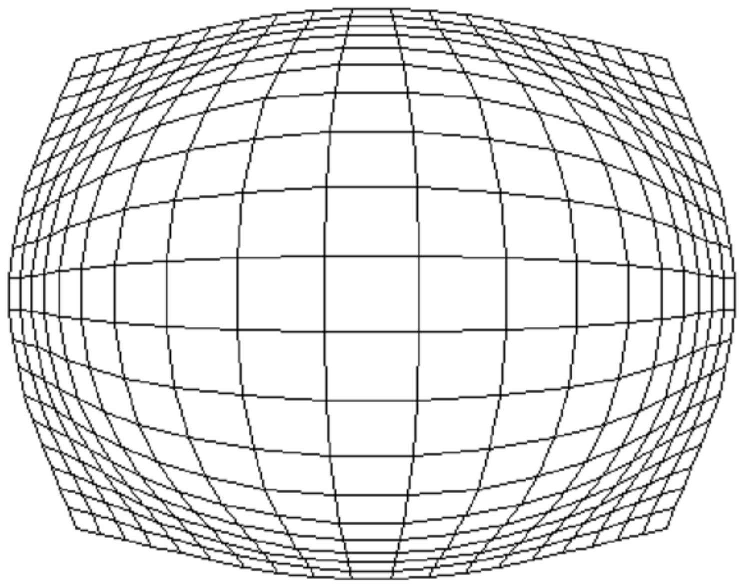
6.2.1 The "Meshflow" Installation

"Meshflow" explodes some common expectations and characteristics of the human visual system. On one hand, the impression of having different levels of gray can be created using only black dots on a white background, by altering the density of dots in an area. This characteristic has been used for years by black and white publications and it is known as halftoning [40] (See also section 3.7).

When we are observing a grid, there is also a presumption of regularity that creates the impression of three dimensionality, as is observed in figure 6.5.

This characteristic has been used for centuries to represent three-dimensional objects in the plane. The artist Marius Watz has created a series of grids that when perturbed in their regularity creates a strong feeling of three dimensionality [71]. In "Meshflow," the nodes on the grid are attracted to the darkest areas of the

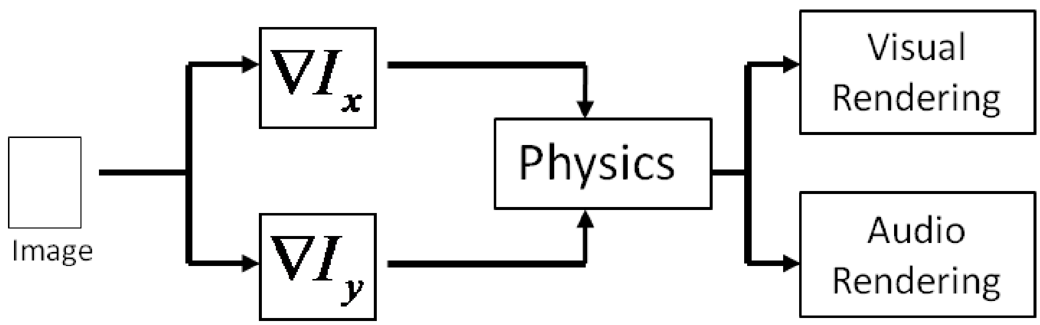
Figure 6.5: A two-dimensional non-uniform grid gives a feeling of three dimensionality.



image, but the grid structure is kept. The motion of the nodes is constrained by physics laws like the attraction between neighbor nodes or the drag force. Figure 6.6 shows the block diagram of the applications.

The horizontal and vertical gradients of the image are approximated using the Sobel operator [37]. Those gradients are used as external forces over the nodes. Internal forces like the spring constant associated with the line segments joining the nodes and drag forces proportional to the nodes velocity are also applied. The set of forces is used to update the position of the nodes and this position is used to render the audio and visuals of the system. The total forces applied to a node

Figure 6.6: Block diagram of the "Meshflow" installation. The horizontal and vertical gradients of the input image are extracted and used as external forces over the nodes. These forces are used by the physics module in conjunction with internal forces as tension and drag to update the position of the nodes on the grid. The position of the nodes is then used by the graphic render system to plot the mesh, and by the audio render unit as a parameter for the sound synthesis.



at position (n_k, n_q) are:

$$F_x = w_t\left(T_x\right) - w_g\left(\bigtriangledown_x\left(C_x\left(n_k, n_q\right), C_y\left(n_k, n_q\right)\right)\right) - w_d\left(V_x\left(n_k, n_q\right)\right) \tag{6.1}$$

Where w_t, w_g, w_d are adjustable weights for the tension, gradient and damping forces; $C_x\left(n_k, n_q\right)$ is the spatial x position of the node with coordinates (n_k, n_q) in the mesh. $\bigtriangledown_x\left(C_x\left(n_k, n_q\right), C_y\left(n_k, n_q\right)\right)$ is the horizontal gradient of the input image evaluated at the position of the node. $V_x\left(n_k, n_q\right)$ is the velocity of the node at (n_k, n_q). T_x is the tension force given by:

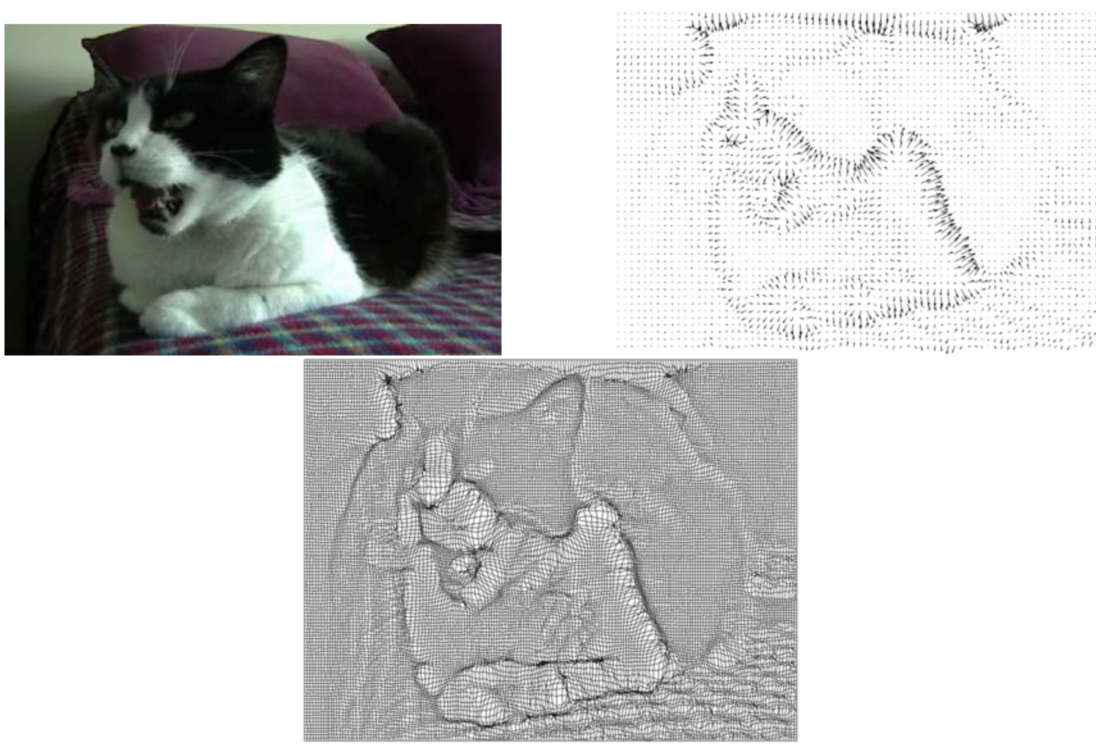
$$\begin{aligned} T_x = &\left(C_x\left(n_k - 1, n_q\right) - C_x\left(n_k, n_q\right)\right) + \left(C_x\left(n_k + 1, n_q\right) - C_x\left(n_k, n_q\right)\right) \\ &+ \left(C_x\left(n_k, n_q - 1\right) - C_x\left(n_k, n_q\right)\right) + \left(C_x\left(n_k, n_q + 1\right) - C_x\left(n_k, n_q\right)\right) \end{aligned}$$

Similarly, the sum of forces for the vertical coordinate is:

$$F_y = w_t (T_y) - w_g (\bigtriangledown_y (C_x (n_k, n_q), C_y (n_k, n_q))) - w_d (V_y (n_k, n_q)) \tag{6.2}$$

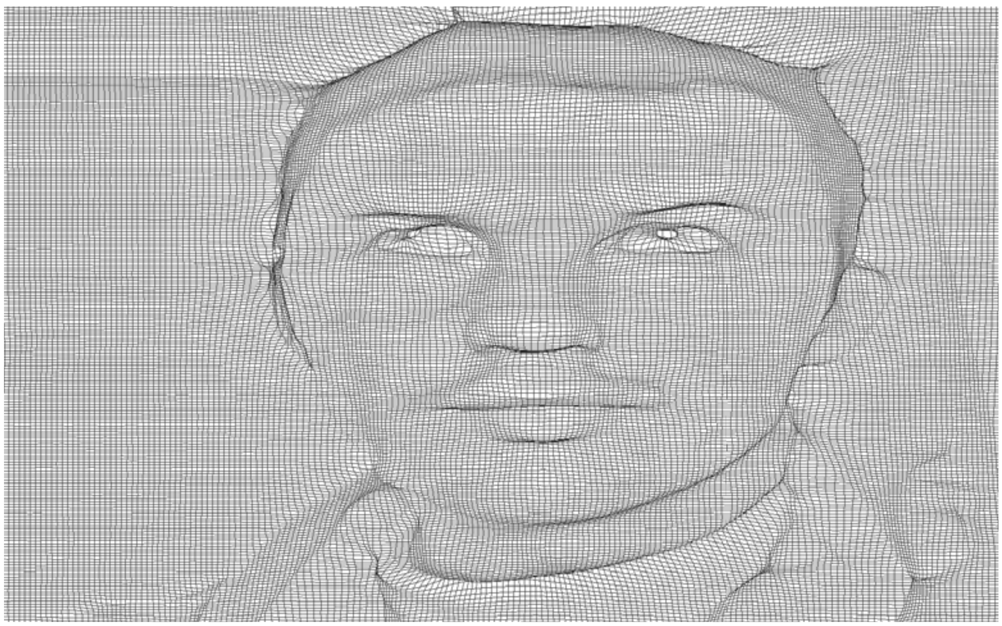
After the forces are evaluated, the position of the nodes is updated following a standard Euler numerical integration [30]. Figure 6.7 shows the original image, the vector field of the gradient and the resulting mesh after applying external (gradient) and internal forces to a regular grid.

Figure 6.7: A sequence showing the original image; the (negative) direction of the gradient (arrows point in the direction of stronger changes from white to black) and what was a regular mesh after applying the gradient and the internal forces.



By using different weighting on the forces that determine the nodes behavior, different results can be obtained. Figure 6.8 and figure 6.9 show the grid output using different configurations on the three weights w_t, w_g, w_d.

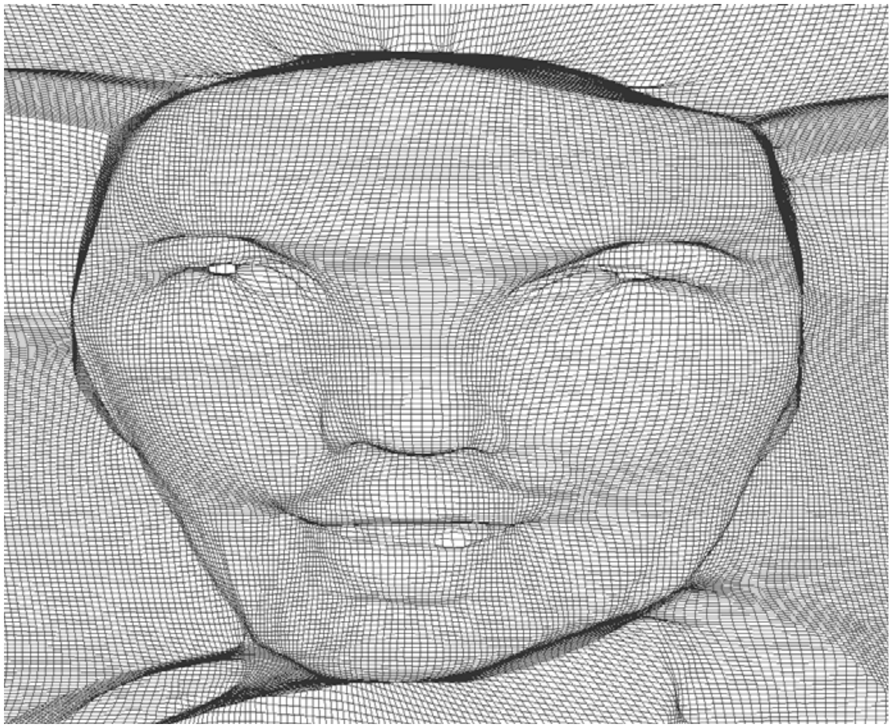
Figure 6.8: The output of the system with a human face as input.



6.2.2 Sound

The audience experience is enhanced by the use of sound. The motion of the mesh is sonificated to increase realism. The sound generation algorithm produces louder sounds when larger changes are detected in the synthetic image. To create the audio, the motion of the nodes on the mesh is used to feed a maraca synthesis algorithm [17]. The visual and the aural are synchronized by coupling the dis-

Figure 6.9: The same person, but with a different parameter configuration than the previous figure.

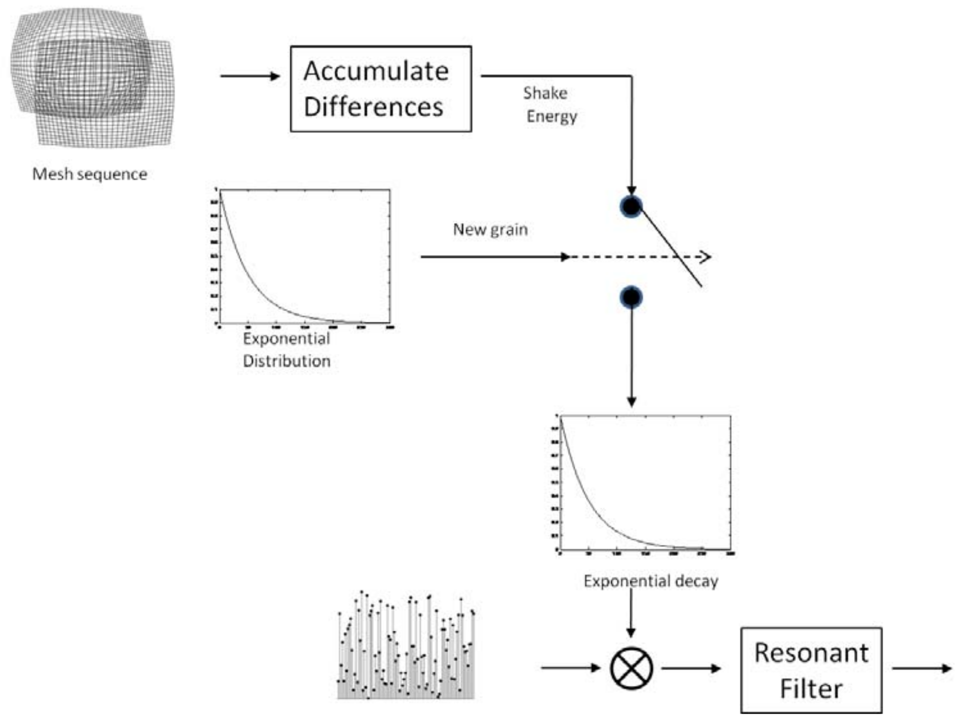


tances traveled by each node in the mesh with the shaking energy of the maraca. Figure 6.10 shows a schematic diagram of the sound rendering block.

6.2.3 Results

The installation runs in real time, it was built using C++. OpenCV was used for the image processing and capture. The graphics were rendered using OpenGL and the sound was made using RTaudio. The installation was presented as a creative showcase during the 2010 edition of the ACE conference. Figure 6.11 shows the hardware setup as it was implemented at the conference.

Figure 6.10: Block diagram of the audio rendering module. The amount of change in the position of the nodes on the grid is used to control the energy of the maraca shake.



6.3 Slave of Your Words

"Slave of Your Words" is a real-time installation that was exhibited at the End of The Year show "Questionable Utility" at the MAT department at UCSB in 2011.

Figure 6.11: The setup at the exhibition place. Minimum hardware requirements: A screen, a P.C and a webcam.

6.3.1 Description

We live in times where decisions that strongly affect our lives are taken by leaders and corporations looking out mostly for their own well-being. Unemployment rates are high, inequality indices are increasing all over the world while the earnings in the banking system are constantly getting better. People from all over the world are starting to raise their voices, and when they speak they become visible. Their problems are seen by others producing a mirror effect that recalls other historical social fights from the past. “Slave of Your Words” is a sequence of software mirrors activated by sound. The image captured with a webcam is build again using the audio information from a microphone. This installation explores the possibilities of transcoding audio and video signals in real time.

Aural and visual information are radically different. While images persist and have a place in space, sounds are volatile and they just fade out into the air. Those characteristics are transmuted in “Slave of Your Words.” Images of the world are only presented when there is sound getting into the microphone. They became volatile. Sound now has a spatial representation and the identity of different sounds is part of this spatial information.

The name “Slave of Your Words,” also reminds us that we are defined by what we say, not only as individuals but also as a society. Our words have an effect on ourselves and on others. The installation represents all of these concepts using three different modes with three different sound activated mirrors.

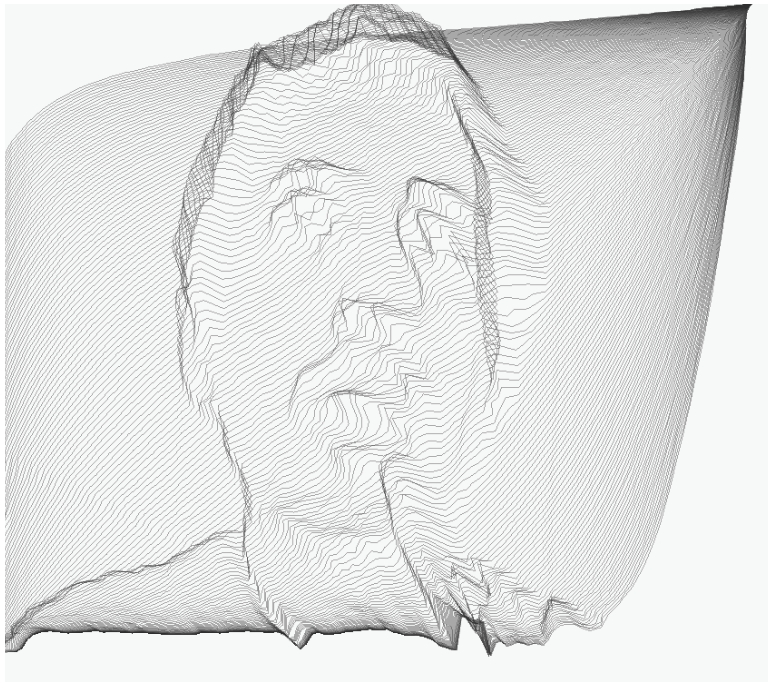
6.3.2 Mode 1. A Family of Curves

In this mode the webcam image is used to manipulate a family of curves with the audio information. In the output image it can be seen how different sounds have a different visual identity.

6.3.3 Mode 2. A Lissajous Array

An array of Lissajous figures created with the audio information are modulated by the brightness of the input image. Periodical signals produce more regular

Figure 6.12: Mode 1 of the installation.

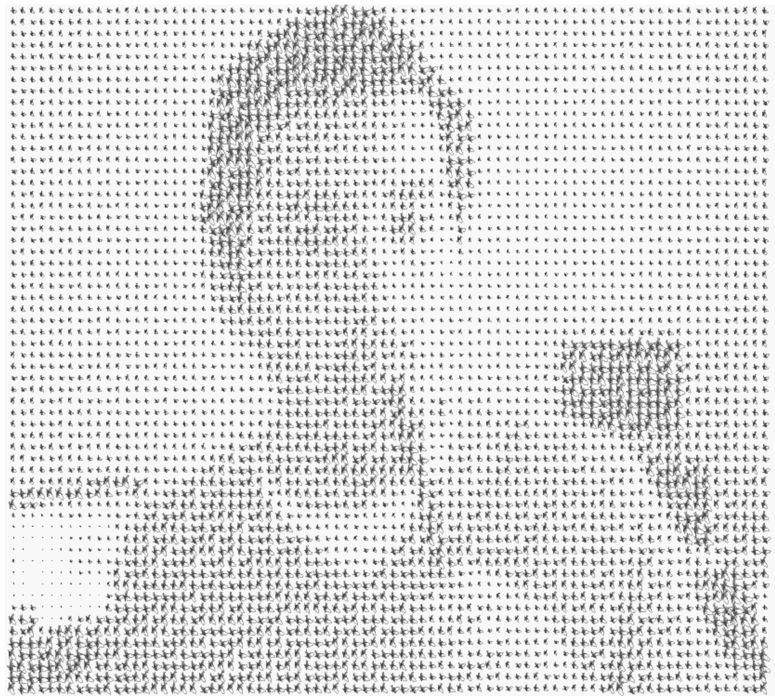


shapes than noisy ones. Characteristics that belong to the audio signal can now be perceived visually.

6.3.4 Mode 3. Circles

The audio information is superimposed on the contours of an array of circles. The amplitude of the contour is altered proportionally with the values of the input image in that region.

Figure 6.13: The installation in mode 2.



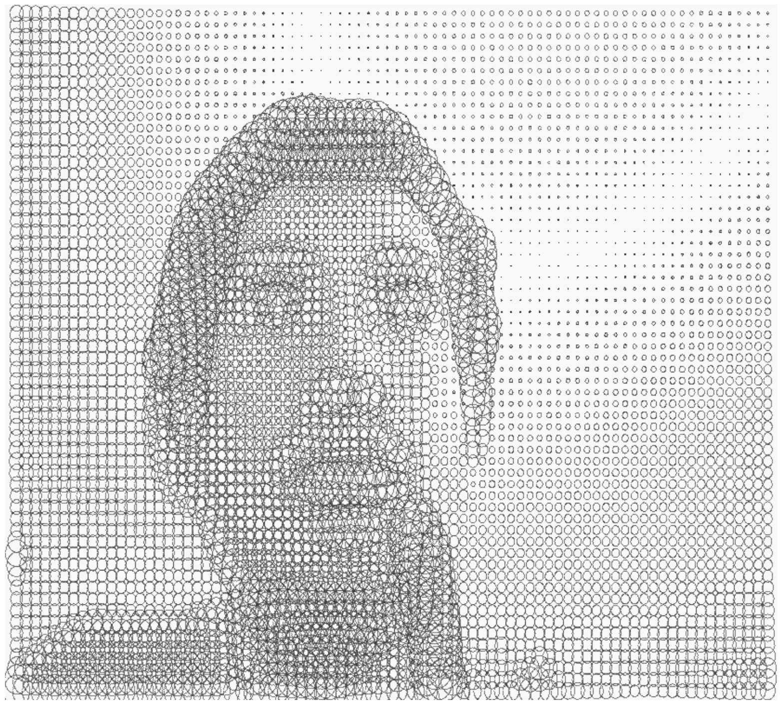
6.3.5 The Installation

The installation runs in real time using a P.C, a webcam and a flat screen. It was created using multiplataform open source libraries like OpenCV, OpenGL and RTAudio. Figure 6.15 shows the audience exploring the installation during the end of the year show "Questionable Utility" in Spring 2011.

6.4 On the Selection of the Subject Matter

Most of the examples presented in this dissertation show frontal human faces as subject matter. This decision does not imply a restriction on any of the algorithms

Figure 6.14: Mode 3 of the installation.



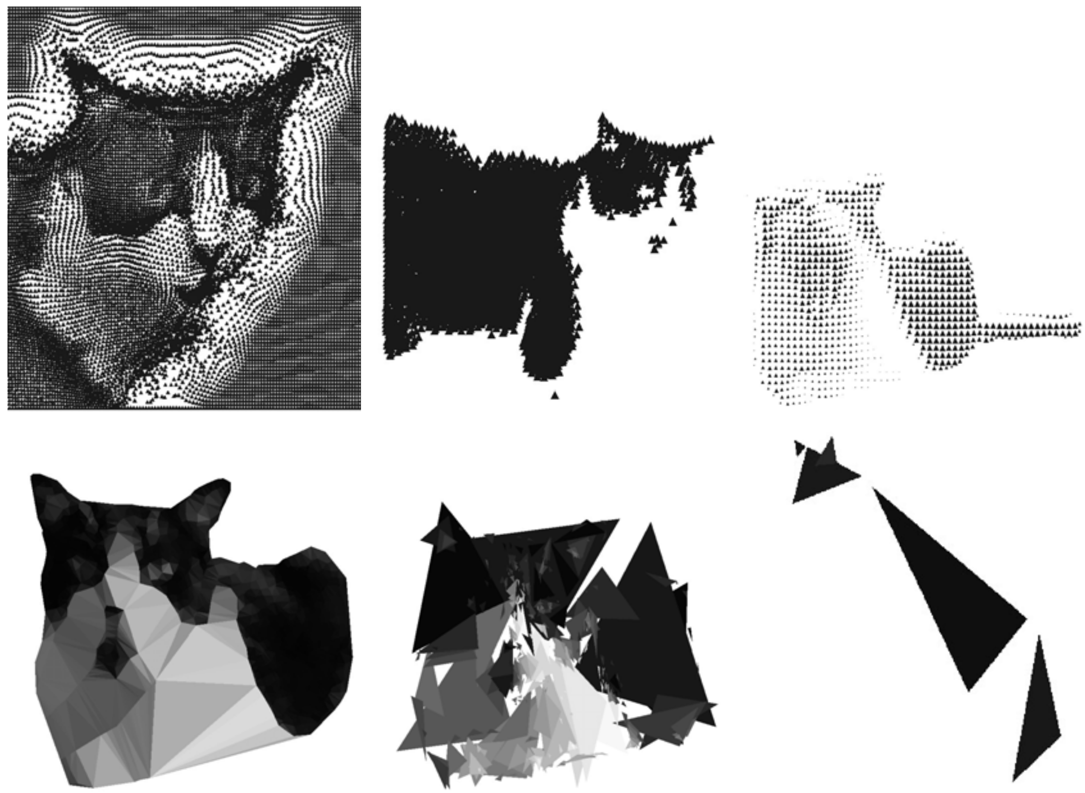
presented but are intended to be convenient for both testing and perception. One clear exception is the short animation presented in the MAT End of the Year Show of 2010: "Triangles and Cats." In this animation, the subject matter is always a cat that is recreated using different configurations of triangles. The animation was created to show that even restricting the synthesis objects to a specific shape like triangles, there are multiple alternatives to recreate the input image (see figure 6.16). Cats were selected as the subject matter because of their ability to communicate with humans through body language and because their shape can be recreated with different levels of complexity and still be identifiable (see figure 6.16).

Figure 6.15: The installation.

6.4.1 Why Faces?

Although in most of the cases it was made for pragmatical reasons, the selection of a face as a subject matter has important consequences. Human observers are particularly good at detecting faces. We need less time to identify an array of elements as a face [33]. This perceptual advantage allows the human visual system to be used as the last part of the A/S process, thus offering the audience the possibility of having some degree of anticipation when the image is being created and to subconsciously participate in the construction of it. A similar perceptual strategy was used in the animation "Please Stand By" presented in the MAT End of the Year Show of 2011. In this animation, very few synthesis elements (18 color bars from the SMPTE NTSC signal) were used to recreate a dancing silhouette. Since humans can identify patterns of human motion from very limited

Figure 6.16: The image of a cat recreated from live video using only triangles as the synthesis element.

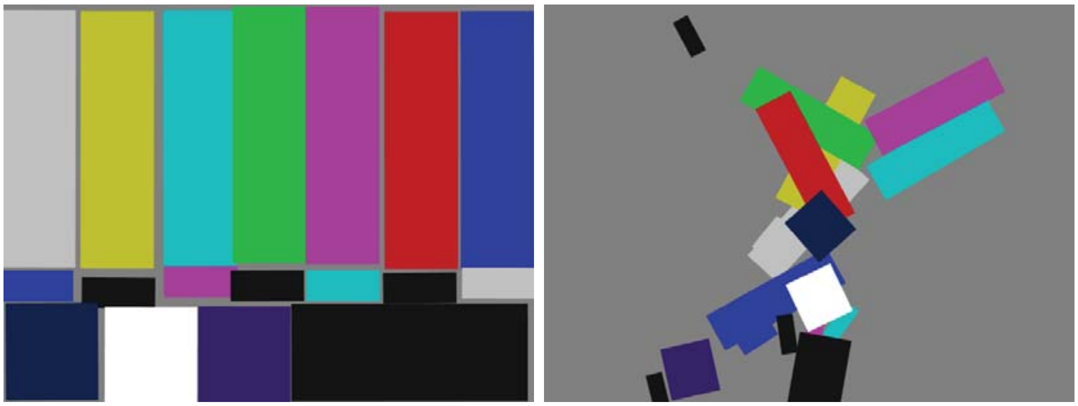


information (see Johannson figures [63]), the motion of the bars is easily identified as a dancing person.

6.4.2 The Real-Time installations

All the real-time installations presented in this dissertation are designed as mirrors. The camera is always pointing to whatever is in front of the output screen. This configuration invites the audience to a mediated self-recognition and to play with its artificial reflection. The installation also challenges the audience

Figure 6.17: Two frames from the "Please Stand By" animation.



to discover the algorithm that is behind. This interaction is more difficult if the camera is pointing to a direction that is out of the control of the viewer.

6.4.3 A Local and a Global Subject Matter.

As it has been shown previously, when using A/S approaches, the synthesis elements are not solely the construction material. Even if they are just an abstract shape they can evoke a particular meaning (e.g., the color bars in "Please Stand By"). When using the cross-synthesis strategies described in chapter 5, it makes sense to talk about the subject matter at two different levels, global and local. One of the alternatives for cross-synthesis presented in chapter 5 is the direct substitution of abstract elements by concrete entities. Left of figure 6.18 shows a frontal face recreated with baguettes while right shows the same person recreated with ants. Note that, the ants image is more believable. This is conceivably because ants tend to work together and arrange themselves in arbitrary ways.

When doing cross-synthesis in images, there is a global subject matter (in these examples it is always a frontal face), but there is also a local subject matter (baguettes or ants) and the different connotations of these elements can influence their perception as adequate synthesis objects.

Figure 6.18: Face recreated with baguettes(left) and ants(right).

6.5 Chapter Summary

This chapter describes some of the pieces that were created using A/S approaches. All of them have been exhibited in front of diverse audiences. The purpose of this chapter was to show that A/S processes are indeed a powerful tool to assist in the creation of media art content. The first example was "The Fitting Dance," a short animation that uses region based mappings as described in section 3.1. In addition, it also shows some of the possibilities of transcoding in audio visual pieces. The audio information is used to significantly alter the shape of the synthesis objects. The same concept is taken to extremes in the third example “Slave of Your Words” where the drawing primitives are the audio waveforms themselves. Three different modes to draw the input image with the audio waveform were explored and presented alternatively during the installation. The other example presented here was also a real-time installation created using the gradient based approach for coherent synthesis as described in section 4.2. Different configurations can be used to produce variate effects on the audience depending on the specific setup. For instance, if the system is placed in a hallway, the mesh can be configured to react quickly, so people that are walking fast can notice that something happened. A totally different setup will work best if “Meshflow” is placed in a gallery where curious people can approach to the system to play with it. There are possibilities for future extensions like interaction with sound or other external signals. To incorporate these new signals into the system, the new stimuli only needs to be added to the total force calculation of each node. The chapter ends with some reflexions about the reasons and implications of choosing a particular subject matter are included.

Chapter 7

Discussion

The motivation for this dissertation was to show that A/S techniques can be considered an effective tool for the creative manipulation of video signals. In order to do that, different pieces and examples were created. Solutions to problems inherent to video signals like the temporal coherence of the synthetic objects were developed. In this chapter, the contributions and findings of the dissertation are summarized and analyzed. The animations and real-time installations that were created and exhibited are used to illustrate these contributions. Some comments about possible extensions and future works are also included.

7.1 Contributions

This dissertation explores the possibilities of Analysis/Synthesis interpretations for the manipulation of video sequences. Different areas of contribution can be highlighted.

1. *Creative Use of Mapping Algorithms.*

   Our brain filters information in order to maximize clarity of structure. Variety is structured into unity to simplify a scene. Chapter 3 showed different ways to insinuate the grouping of objects to the human visual system. All these different synthesis strategies can be used to support a particular narrative interest. During this dissertation, different examples of videos and installation were created to show the creative possibilities of such mappings.

   In the creation of the animation "The Fitting Dance" described in detail in section 6.1, the shape and the number of elements in the scene are used as a way to control the narrative. In the final part of this animation, an audio signal is used directly to manipulate the contours of the synthesis objects showing the versatility of A/S approaches when combined with external signals. In the creation of this short animation, frames were processed independently. However, the perception of causality was not affected since the selected original scene was controlled and the amount of motion was limited. The animation was engaging and audience reactions were enthusiastic. The coupling between the changes in music and the modification of the object contours got the audience connected to the piece. Viewers manifested surprise when a face was slowly created out of ellipses. Ellipses were chosen for being the smoothest convex shapes that can carry information about the size, color, and orientation of a region. Additionally, with Fourier descriptors, a region representation can be evolved smoothly from ellipses to the full contour. This strategy was used in the climax scene of "The Fitting Dance" to increase the tension of the visual elements in synchronicity with the music.

The interactive installation "Meshflow" described in section 6.2 also uses mapping strategies creatively. Proximity and connectedness cues are used to recreate the camera input. "Meshflow" is designed as a mirror installation. For the experience, it is important that the camera is pointing to an area that can be affected by the audience. In this way, the participants can play with their own resynthesis and get immersed in a process of discovery. New observers are usually attracted when they see somebody else playing with the system. Audiences are always curious about the sensor type since the mesh distortion creates the illusion of tridimensionality (See figure 6.5). The "Meshflow" installation can be set on different configurations and diverse setups can be used in different scenarios. For instance, if the installation is in a hall where people walk fast, the mesh can be configured to respond quickly. On the other hand, in a gallery the installation can be set into a more intimate mode where the effect of the previous positions of the nodes have a more dramatic influence on the current output. With this setup, the response time of the system will be longer so audiences need to be more patient to see their image recreated.

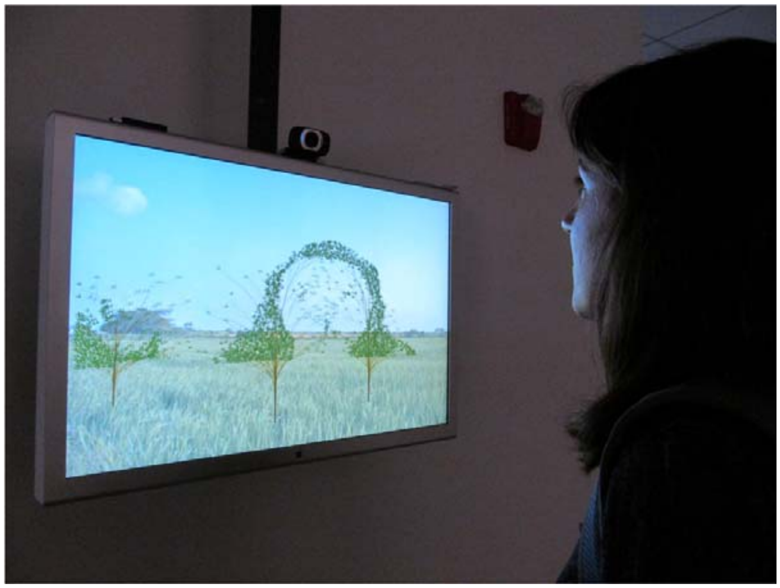
2. *Algorithms for Temporal Coherence.*

   Lack of temporal coherence is indeed the most distracting artifact in video sequences created through A/S processes. In this dissertation, this problem was addressed and three different solutions were provided. The first solution uses a combinatorial optimization approach. This solution is novel in animation and when different matching criteria are used, the animated output has a different look and feel. The optimization criteria becomes a parameter with a strong effect on the aesthetics of the results. This strategy was shown in the first two scenes of the animation "Background Singer," presented in the MAT end of the Year Show of 2012. The criterion of the minimum sum is applied to match the silhouettes in consecutive frames. Then, intermediate positions are generated with the Viterbi interpolation algorithm. Despite that with this technique the results can have a lack of cohesion (see section 4.4.4), the illusion of small bodies recreating a face in motion is clear and it was understood by the audience. The second solution for temporal coherence is based on the creation of a surface where the synthesis objects can move. This was the technique used in the real-time installation "Meshflow." In "Meshflow," the representation problem inherent to the gradient-based technique (see section 4.4.3) is solved with spatial constraints in the synthesis elements (the nodes). The third strategy for temporal coherence was presented in section 4.3. It is based on the use of attractors in the parameters domain. This technique was used in the creation of cross-synthesis applications that will be discussed next.

3. *Automatic Generation of Ambiguous Images/Videos.*

   An additional contribution of this research is the development of algorithms for the automatic and assisted generation of ambiguous images and videos. An example that illustrates the creation of ambiguous images is the real-time installation "Herbaceous." In "Herbaceous," the image of a group of trees is derived from a dithered representation of the input image. Both the images of the trees and observer can be recognized. The installation is simultaneously a mirror that returns the image of the viewer, and a window to a landscape with virtual plants. The algorithm successfully creates ambiguous images (see figure 7.1). When the input is a video, the algorithm based on target points that was presented in section 4.3 can be used to generate a time variant version of the experience. The local subject matter is a group of trees. Trees are adequate candidates to represent the input sequence since trees are symbols of life, evolution and hierarchy. The global subject matter is the audience itself, since the algorithm is a mirror installation. The members of the audience are allowed to participate in the experience not only as observers but also as creators and manipulators.

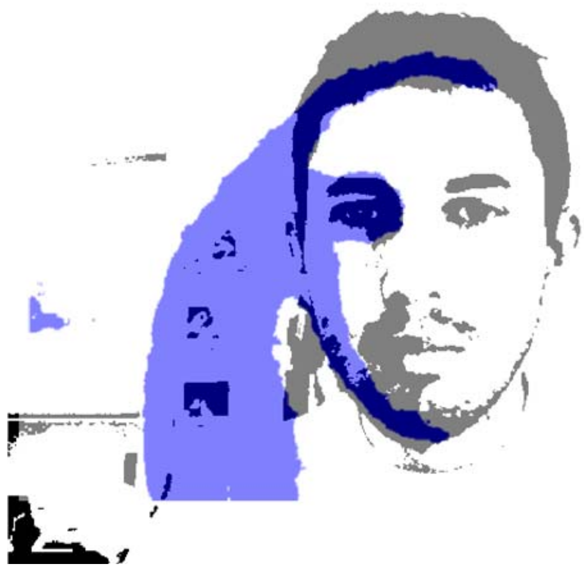
Figure 7.1: The “Herbaceous” installation.



4. *The Use of Cross-Synthesis for Parallel Narratives.*

   This dissertation presents ambiguous videos as an alternative to what cross-synthesis is in the audio domain, but in the video domain possibilities go beyond the mixing of two signals. With strategies like the ones presented in chapter 5 pieces with two coexisting narratives can be created. Real-time installations like “Herbaceous” or “Ant Theater” (section 5.1) show that it is possible to create two simultaneous realities. A mirror that reflects the observer and a window to a different world (the trees or the group of ants). Also, the last two scenes of the short animation “Background Singer” use the juxtaposition of global and local information to present two parallel stories, one at each level. The creation of these scenes was not automatic, it was computer assisted (see figure 7.2). The global subject matter is again a frontal human face so the extraordinary ability of human viewers to perceive faces is exploited. The local subject matter is given more explicitly and consists of human silhouettes adopting different positions in order to recreate the global human face (figure 7.3).

Figure 7.2: Recreating a global image for "Background Singer" with the semiautomatic application.



By presenting these two different layers (local and global), this animation shows the narrative possibilities of cross-synthesis for parallel storytelling.

7.1.1 Technical Contributions

In the development of this dissertation, different algorithms were created to explore the creative possibilities of A/S approaches on video signals. In many cases the algorithms are well known in different fields of engineering and the contribution

Figure 7.3: Two frames from the "Background Singer" animation.

resides in identifying them as solutions to the specific problems of this dissertation or adapting them to show their possibilities for creative composition. For instance, the algorithms used to fit ellipses and other objects are standard, but their use for the controlled construction of images is innovative (for example "The Fitting Dance" in section 6.1). The following is a list describing the origin, modification and repurposing of the most important algorithms used in this work.

1. The "Minimum Sum" and "Minimum Maximum" algorithms find optimum solutions to the matching problem (see section 4.1). Those two algorithms are part of the standard tools of combinatorial optimization theory. Their use as an alternative for interpolation in animation is new, and the modifications on the cost matrix to get directional filtering of the trajectories is exclusive to this work.

2. The two other algorithms for temporal coherence were developed directly as alternatives to the problem of how to create real-time temporal coherent synthesis. The general structure of the surface-based algorithm (see section 4.2) resembles digital art approaches like generative art. The use of the distance transform to produce a surface from the image and the use of this surface combined with feedback loops or packing algorithms in the parameters space gives this algorithm its originality.

3. The attractors-based strategy is a sub-optimal approach to the matching problem. It was built from scratch and inspired by a simplification of the McAulay-Quatieri algorithm used in audio signals to match partials but adapted to the multidimensional case. This is the temporal coherence algorithm used in "Herbaceous" and "Ant Theater."

4. One of the most important technical contributions of chapter 4 is the development of quantitative measurements for the objective comparison of the temporal coherence techniques presented in the same chapter. The use of closed formulas is not common in the evaluation of non-photorealistic animations. The calculation of the jerk as an indication of the smoothness of the trajectories has been used in different disciplines such as movement planing, robotics or ergonomics, but it is novel in animation. The formula to compare cohesion was created for this dissertation.

5. The technique presented in section 5.2 describes how to combine a template matching algorithm with the optimum assignment approaches from chapter 4, but includes the constraint that every object needs to have a discrete state model. Intermediate states of the synthesis objects are calculated using the Viterbi algorithm. The Viterbi algorithm finds the more likely trajectory between two states in a determined number of steps. This algorithm is extensively used in diverse applications such as the detection of convolutional codes and the recognition of symbol sequences [45]. It has also been used to find a smooth sequence of mouth positions in visual speech synthesis [5], but the use of the Viterbi algorithm to generate in-between frames is new in animation and it is an important contribution of this work.

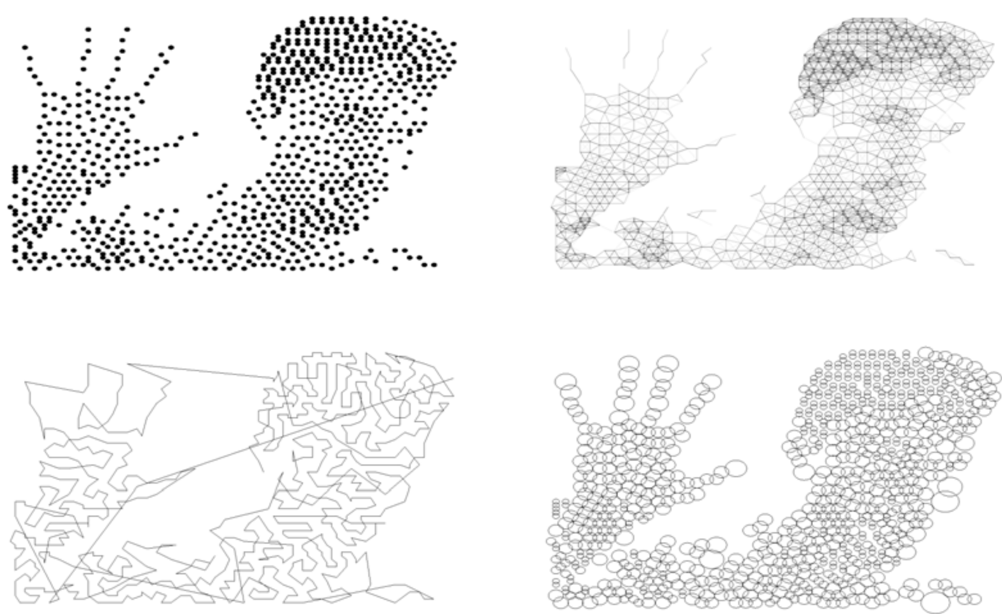
7.2 Conclusions

In this dissertation different alternatives for the manipulations of video signals using A/S techniques have been shown. The most important findings of this exploration process were the following:

1. *The grouping principles of the human visual system can be exploited effectively with A/S strategies.*

   Although this dissertation was inspired by the role that A/S techniques have in the audio and music processing communities, the approaches that were explored to recreate moving images are significantly different from the ones used in the acoustic domain. The human visual system simplifies its work by grouping elements to assign structure and create unity. As it was show in chapter 3, diverse strategies can be used to suggest the spatial grouping of objects when recreating the input image (e.g., regions, density, proximity). With A/S techniques, the analyzed data can be manipulated and drawn in different ways to take advantage of the diverse grouping principles. Figure 7.4 shows the same input image recreated in 4 different ways (also see chapter 3).

Figure 7.4: The same set of points used in different ways to recreate the input image.



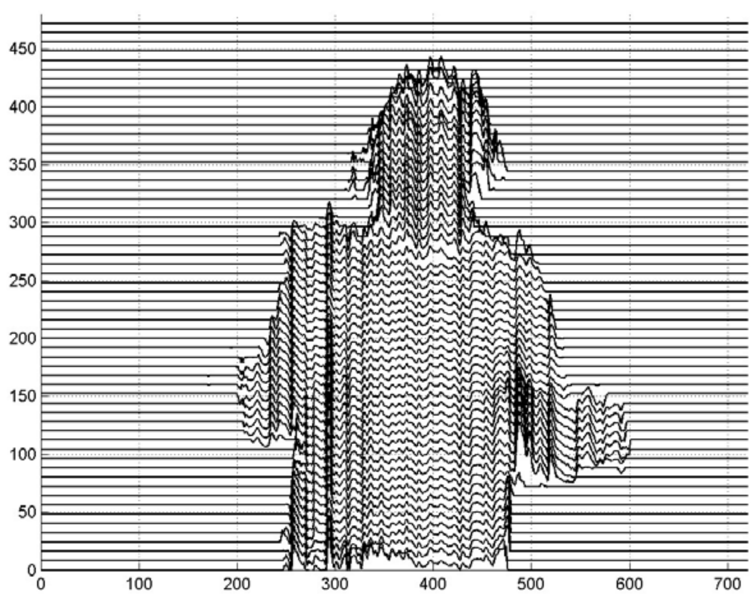
2. *A representation of synthesis objects in the parameters domain is convenient for synthesis and interpolation.*

   If the synthesis objects can be represented as points in the multidimensional parameters space, it is possible to apply to them the temporal coherence strategies explained in chapter 4. A distance between objects can be calculated and this distance can be used to create the cost matrix for the matching algorithms presented in section 4.1. This distance measurement can also be weighted to give more sensitivity to the dimensions that are perceptually more important. Furthermore, the interpretation in the parameters domain strategies traditionally utilized in particle system animation can be applied on arbitrary synthesis objects (e.g.,attracting and repelling forces, emitters, sinks).

3. *A/S techniques can lead to strongly coupled audiovisual pieces.*

   Audio and musical information can be incorporated directly during the synthesis stage. Features extracted from the audio signal can be mapped on the parameters of the synthesis objects. Changes in the acoustic signal will be seen immediately in the recreated image. In this dissertation these possibilities have been explored in the animation “The Fitting Dance” (see section 6.1). In this animation, the RMS value of the audio signal is used to change the amplitude of a sinusoidal signal superimposed on the contours of the synthesis objects (ellipses in this case). In other scene of the same animation, the spectral content of the acoustic channel determines the number of frequency components used in the reconstruction of the connected regions. Spectral richness in audio is mapped into spectral richness in the contours of the synthesis elements. In the animation "Narrative Lines" (2nd place in the animation category of the Swan Lake: Moving Image & Music Awards 2011), the waveform of an existing piano track is used to draw the input image (see figure 7.5). The real-time installation "Slave of Your Words" presented in section 6.3, shows a similar strategy. The waveform of the audio signal captured with a microphone is the synthesis object itself. With these strategies, pieces with a strong coupling between audio and visual can be generated.

Figure 7.5: A frame of the animation "Narrative Line". The synthesis object is the waveform of the audio track.



4. *There are multiple trade-offs in the problem of generating temporal coherent video sequences using an A/S process.*

   Chapter 4 presents different strategies to create temporal coherent animations from video data. There are different trade-offs in any of the algorithms described. Techniques that are highly cohesive (i.e., the dispersion of the local direction is low) are not accurate regarding the quality of representation. In these techniques, synthesis objects are evolved according only to local characteristics of the input image. On the other hand, algorithms that behave according to global criterion, such as the "Minimum Sum" and "Minimum Maximum" have a good quality of representation but are not cohesive. It was also shown in chapter 4 that the application of physical inspired interpolation of the parameters of the synthesis primitives can improve the smoothness of the transition, but always at expense of the quality of representation. Chapter 4 shows that every algorithm has a different behavior in respect to smoothness, cohesion, quality of representation and distribution of jumps. All these differences define a distinct dynamic behavior and the selection of a determinate algorithm or set of algorithms becomes a creative decision.

5. *A/S techniques are suitable for the creation of ambiguous videos.*

   In ambiguous images more than one stable interpretation of the picture is valid. A classic example is Rubin's vase, an image that can be interpreted as a cup or as two faces looking at each other. Video adaptations of ambiguous images are not know by the author of this dissertation. In the examples in chapter 5, the image from the camera is built using synthesis elements that are not abstract geometric objects but rather figurative forms. In this way A/S techniques can be used to create ambiguous videos, videos that convey meaning at the global and local level simultaneously.

7.3 Limitations

As it was pointed out in the previous section, the different results, algorithms, animations and real-time installations presented in this dissertation show that A/S approaches can be considered an important tool for the creative manipulation of video signals. The techniques described in chapter 4 of this dissertation solve the problem of temporal coherence, but when doing so they made the problem of lack of cohesion explicit. A low cohesion value means that synthesis elements that are spatially close to each other can evolve in dissimilar ways over time. This local variance can make the global image more difficult to follow. In addition, in chapter 4 a technique for interpolation in the parameters domain was presented. This interpolation is not guaranteed to work with arbitrary parameter sets. It has not been proven that intermediate values will always produce drawable objects but no pathological cases have been encountered. Even taking into consideration that the cross-synthesis alternatives presented here are an important innovation, the quality of the results is still far from the classic hand made ambiguous images (see figure 5.2). The elements that can be used to build the global images must be very simple and manipulable. However, this is not different than the audio case where good cross-synthesis results also require specific (spectral) characteristics in the carrier signal.

7.4 Extensions

Besides the possible use of the techniques developed here on art related applications, other fields can also benefit from those approaches. Having the possibility of gradually manipulating the number of elements that made up a face, it is inevitable to ask questions related to human perception. These questions include: How many parameters do we need to recognize a face?, Can we detect faces with less synthesis elements than other objects?, and Where is the threshold where we can no longer identify a person on the re-synthesis? A/S processes can also be used for gesture augmentation, people with limited mobility can still have visual feedback for the small motions that they perform. Problems like the linear assignment problem and the bottleneck assignment problem are recurrent in computer science, but they are always presented with abstract examples. Curiously enough, the animations created with these combinatorial optimization algorithms can be used as a way to visualize the algorithms themselves.

7.5 Future Work

The use of the techniques presented here and the adaptation of the examples shown, can lead to richer visual performance experiences (i.e., dance, theater). However, it is challenging to design those kinds of pieces and use the technology as an integral part of the narrative, not just as a visual effect. For instance, pieces like "Slave of Your Words" (section 6.3) have been suggested by critics as a way to represent images (captured from video data) as they are told by a narrator. Another big challenge for the future is to use cross-synthesis to effectively tell two stories simultaneously. In a similar manner as cartoonist Gustave Verbeek did with his "Upside downs strips" in the early 1900s [11], but by using the local versus global as the narrative paths. In terms of algorithms, important aspects to be considered in future works are the exploration of techniques to improve cohesion on the animations created with the matching approaches. This aspect is strongly related to what Bénard et al. [9] call "motion coherence," that states that primitives tend to move in the same direction as the objects they are shaping. However, in many cases, the best places to set the primitives are not good places to track the motion of the scene. If the information of the motion of real objects in the scene is available, it can be used to penalize the cost matrix on the matching approaches as was done for the direction filtering technique described in section 4.1.3. Another aspect that deserves further investigation is the temporal coherence of second level structures. As it was shown in section 3.8, the way in which a set of elements is connected can drastically change the way it looks. However, even if the temporal coherence of the points is guaranteed, the lines connecting them can exhibit appearing and disappearing artifacts. The generative cross-synthesis approaches can have the same problem. Algorithms that simultaneously take care of the two levels (dots and lines) are needed. Finally, a real-time system inspired by the cross-synthesis technique described in section 5.2 is desirable. With this approach, instead of doing the off-line matching, the position and state of the detected objects will be updated from one frame to the next. Having a model of the state transition of the primitives will still be useful to calculate the state that is most likely to be detected on the next frame.

7.6 Chapter Summary

This chapter elaborates upon the examples presented in this dissertation and on how they illustrate the creative possibilities of A/S approaches. The most important contributions and findings of the dissertation are discussed and exemplified. A specific subsection is destined to emphasize the technical developments that can be considered as novel. The limitations and possible uses of the algorithms in different contexts and future works are also examined.

Bibliography

[1] E. Aboufadel, S. Boyenger, and C. Madsen, *Digital creation of chuck close block-style portraits using wavelet filters*, Journal of Mathematics and the Arts, 4 (2010).

[2] A. Agrawal, *Non-photorealistic rendering: Unleashing the artist's imagination [graphically speaking]*, Computer Graphics and Applications, IEEE, 29 (2009), pp. 81–85.

[3] L. Allen, *Taking on a-ha classic*. `http://www.bbc.co.uk/news/entertainment-arts-11485702`, 2010.

[4] M. Anderson, E. McDaniel, and S. Chenney, *Constrained animation of flocks*, in Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation, SCA '03, Aire-la-Ville, Switzerland, Switzerland, 2003, Eurographics Association, pp. 286–297.

[5] E. Bárcenas, M. Díaz, R. Carrillo, R. Solano, C. Soto, L. Valderrama, J. Villegas, and P. Vizcaya, *A coding method for vi-*

sual telephony sequences, in ISCA,Auditory-Visual Speech Processing, ISCA, 2005.

[6] B. Barton, *Radiohead "house of cards"*, in ACM SIGGRAPH 2009 Computer Animation Fesitval, SIGGRAPH '09, New York, NY, USA, 2009, ACM, pp. 118–118.

[7] J. Beardsworth, *Photoshop Fine Art Effects Cookbook: 62 Easy-to-Follow Recipes for Creating the Classic Styles of Great Artists and Photographers*, O'Reilly Media, 1 ed., Feb. 2006.

[8] P. Bénard, A. Bousseau, and J. Thollot, *Dynamic solid textures for real-time coherent stylization*, in ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D), Boston, MA, Etats-Unis, february 2009, ACM, ACM, pp. 121–127.

[9] ——, *State-of-the-art report on temporal coherence for stylized animations*, Computer Graphics Forum, 30 (2011), p. 23672386. DOI: 10.1111/j.1467-8659.2011.02075.x.

[10] P. Bénard, A. Lagae, P. Vangorp, S. Lefebvre, G. Drettakis, and J. Thollot, *A dynamic noise primitive for coherent stylization*, Computer Graphics Forum (Proceedings of the Eurographics Symposium on Rendering 2010), 29 (2010), pp. 1497–1506.

[11] J. R. Block, *Seeing double : over 200 mind-bending illusions / J. Richard Block*, New York : Routledge, 2002. Includes index.

[12] H. Bohnacker, B. Gross, J. Laub, and C. Lazzeroni, *Generative Gestaltung: entwerfen, programmieren, visualisieren*, Schmidt, 2009.

[13] J. D. Bolter and D. Gromala, *Windows and Mirrors: Interaction Design, Digital Art, and the Myth of Transparency*, The MIT Press, Oct. 2005.

[14] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, O'Reilly Media, 2008.

[15] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment problems*, Society for Industrial Mathematics, 2009.

[16] J. Campbell, *Jim campbell, low resolutions works.* `http://www.jimcampbell.tv/portfolio/low_resolution_works/`, 1999-2011.

[17] P. Cook and I. Books24x7, *Real sound synthesis for interactive applications*, AK Peters, 2002.

[18] V. Cosic, *Andreamosaic.* `http://www.ljudmila.org/~vuk/`.

[19] C. Csuri and J. Shaffer, *Art, computers and mathematics*, in Proceedings of the December 9-11, 1968, fall joint computer conference, part II, AFIPS '68 (Fall, part II), New York, NY, USA, 1968, ACM, pp. 1293–1298.

[20] K. Dalal, A. Klein, Y. Liu, and K. Smith, *A spectral approach to npr packing*, in Proceedings of the 4th international symposium on Non-photorealistic animation and rendering, ACM, 2006, pp. 71–78.

[21] A. Denzler, *Andreamosaic*. `http://www.andreaplanet.com/andreamosaic/`.

[22] F. Dietrich, *Visual intelligence: The first decade of computer art (1965-1975)*, Leonardo, 19 (1986), pp. pp. 159–169.

[23] M. Dolson, *The phase vocoder: A tutorial*, Computer Music Journal, 10 (1986), pp. pp. 14–27.

[24] R. O. Duda and P. E. Hart, *Use of the hough transformation to detect lines and curves in pictures*, Commun. ACM, 15 (1972), pp. 11–15.

[25] A. Fabris, *Waldemar cordeiro: Computer art pioneer*, Leonardo, 30 (1997), pp. pp. 27–31.

[26] C. Finch, *Chuck Close: Work*, Prestel Publishing, Oct. 2007.

[27] S. Frere-Jones, *The gerbil's revenge : The new yorker*. `http://www.newyorker.com/arts/critics/musical/2008/06/09/080609crmu_music_frerejones?currentPage=all`, 2008.

[28] P. Galanter, *What is generative art? complexity theory as a context for art theory*, 2003. `http://philipgalanter.com/downloads/ga2003_what_is_genart.pdf`.

[29] M. Gondry, *The work of director michel gondry*. DVD,Palm Pictures, 2003.

[30] H. Gould, J. Tobochnik, and C. Wolfgang, *An introduction to computer simulation methods: Applications to Physical Systems*, Addison-Wesley Longman Publishing Co., Inc., 2005.

[31] P. Haeberli, *Paint by numbers: abstract image representations*, SIGGRAPH Comput. Graph., 24 (1990), pp. 207–214.

[32] L. D. Harmon and B. Julesz, *Masking in visual recognition: Effects of two-dimensional filtered noise*, Science, 180 (1973), pp. pp. 1194–1197.

[33] O. Hershler and S. Hochstein, *At first sight: A high-level pop out effect for faces*, Vision research, 45 (2005), pp. 1707–1724.

[34] A. Hertzmann and K. Perlin, *Painterly rendering for video and interaction.*, in NPAR'00, 2000, pp. 7–12.

[35] R. Hodging, *Robert hodging*. `http://roberthodgin.com/`.

[36] E. Iris Simmons, *Erika Iris Simmons - ghost in the machine*. `http://www.iri5.com/`.

[37] N. Kanopoulos, N. Vasanthavada, and R. Baker, *Design of an image edge detection filter using the sobel operator*, Solid-State Circuits, IEEE Journal of, 23 (1988), pp. 358–367.

[38] F. Keiler, D. Arfib, and U. Zölzer, *Efficient linear prediction for digital audio effects*, DAFX-00, (2000), pp. 19–24.

[39] Knowlton, *Knowlton mosaics - portraits by computer assisted art pioneer ken knowlton*. http://www.knowltonmosaics.com/.

[40] D. Lau and G. Arce, *Modern digital halftoning*, vol. 8, CRC, 2001.

[41] G. Levin, *Floccular portraits - interactive art by golan levin and collaborators*. http://www.flong.com/projects/floccugraph/.

[42] ——, *Segmentation and symptom - interactive art by golan levin and collaborators*. http://www.flong.com/projects/zoo/.

[43] P. Litwinowicz, *Impressions of san francisco*, in ACM SIGGRAPH 97 Visual Proceedings: The art and interdisciplinary programs of SIGGRAPH '97, SIGGRAPH '97, New York, NY, USA, 1997, ACM, pp. 268–.

[44] ——, *Processing images and video for an impressionist effect*, in Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97, New York, NY, USA, 1997, ACM Press/Addison-Wesley Publishing Co., pp. 407–414.

[45] H. Lou, *Implementing the viterbi algorithm*, Signal Processing Magazine, IEEE, 12 (1995), pp. 42–52.

[46] R. Lozano-Hemmer, *Rafael Lozano-Hemmer - project "Eye contact"*. `http://www.lozano-hemmer.com/eye\_contact.php`.

[47] ——, *Rafael Lozano-Hemmer - project "Third person"*. `http://www.lozano-hemmer.com/third\_person.php`.

[48] T.-Q. Luong, A. Seth, A. Klein, and J. Lawrence, *Isoluminant color picking for non-photorealistic rendering*, in Proceedings of Graphics Interface 2005, GI '05, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2005, Canadian Human-Computer Communications Society, pp. 233–240.

[49] R. McAulay and T. Quatieri, *Speech analysis/synthesis based on a sinusoidal representation*, Acoustics, Speech and Signal Processing, IEEE Transactions on, 34 (1986), pp. 744–754.

[50] B. J. Meier, *Painterly rendering for animation*, in Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH '96, New York, NY, USA, 1996, ACM, pp. 477–484.

[51] V. Muniz, *Vik muniz*. `http://www.vikmuniz.net/`.

[52] R. Měch and P. Prusinkiewicz, *Visual models of plants interacting with their environment*, in Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH '96, New York, NY, USA, 1996, ACM, pp. 397–410.

[53] C. Perkins, *The After Effects Illusionist: All the Effects in One Complete Guide*, Focal Press, Pap/Cdr ed., Jan. 2009.

[54] J. Reichardt, *Cybernetic serendipity;: The computer and the arts*, Praeger, 1969.

[55] C. Reynolds, *Flocks, herds and schools: A distributed behavioral model*, in ACM SIGGRAPH Computer Graphics, vol. 21, ACM, 1987, pp. 25–34.

[56] C. Roads, *The Computer Music Tutorial*, MIT Press, Cambridge, MA, USA, 1996.

[57] C. Roads, *Microsound*, The MIT Press, Sept. 2004.

[58] Y. Rodkaew, P. Chongstitvatana, S. Siripant, and C. Lursinsap, *Particle systems for plant modeling*, Plant Growth Modeling and Applications, (2003), pp. 210–217.

[59] X. Serra, *Musical sound modeling with sinusoids plus noise*, Musical Signal Processing, (1997), pp. 497–510.

[60] J. Shi and C. Tomasi, *Good features to track*, in Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on, IEEE, 1994, pp. 593–600.

[61] R. Silvers, *Robert silvers*. `http://www.photomosaic.com/`, 2003.

[62] K. Smith, Y. Liu, and A. Klein, *Animosaics*, in Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation, SCA '05, New York, NY, USA, 2005, ACM, pp. 201–208.

[63] R. SNOWDEN, P. THOMPSON, and T. TROSCIANKO, *Basic vision: an introduction to visual perception*, Oxford University Press, 2012.

[64] T. Strothotte and S. Schlechtweg, *Non-Photorealistic Computer Graphics: Modeling, Rendering and Animation*, Morgan Kaufmann, 1 ed., June 15 2002. ISBN: 1558607870, pages 472.

[65] Translab, *Translab[4]: Algorithm & code visual aesthetics in early computing (1950-80)*. `http://translab.burundi.sk/code/vzx/`.

[66] W. Vasulka and B. O'Reilly, *Scan processor studies (excerpts pt.1) on vimeo*. `http://vimeo.com/7517418`.

[67] J. Villegas, *The autonomous duck: Exploring the possibilities of a markov chain model in animation*, Arts and Technology, (2010), pp. 272–278.

[68] 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, 21 (1995), p. 32.

[69] L. Walker, *Waste land*. O2 Films, 2010.

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

[71] M. Watz, *Marius Watz - grid distortion*. `http://www.unlekker.net/proj/griddistortions/`, 2008.

[72] T. Weise, S. Bouaziz, H. Li, and M. Pauly, *Realtime performance-based facial animation*, ACM Trans. Graph, 30 (2011), pp. 1–77.

[73] J. Xu, X. Jin, Y. Yu, T. Shen, and M. Zhou, *Shape-constrained flock animation*, Comput. Animat. Virtual Worlds, 19 (2008), pp. 319–330.

[74] U. Zölzer, X. Amatriain, D. Arfib, J. Bonada, G. D. Poli, P. Dutilleux, G. Evangelista, F. Keiler, A. Loscos, D. Rocchesso, M. Sandler, X. Serra, and T. Todoroff, *DAFX:Digital Audio Effects*, Wiley, 1st ed., May 2002.