Creative Applications of Microvideos

Angus Forbes and Javier Villegas
School of Information: Science, Technology, and Arts
University of Arizona
angus.forbes@sista.arizona.edu, javier.villegasp@gmail.com

Abstract—**This paper introduces ongoing work on video granular synthesis. The strategies traditionally used in granular synthesis in order to granulate audio signals are extended to streams of video data. We present initial techniques that are made possible through transforming a video signal into a large array of microvideos, or video grains. These involve the dynamic resynthesizing of the video based on spatial and temporal elements of the video grains. Initial explorations in the creative manipulation of videos using these methods are described. We show that video granular synthesis strategies facilitate novel video processing techniques that could lead to new creative effects.**

Keywords - Video granulation; Video processing; Granular synthesis

I. INTRODUCTION TO VIDEO GRANULAR SYNTHESIS

Granular synthesis is a common method for creating new sonic textures [1]. For instance, it is one of the preferred strategies to manipulate the duration of existing sounds without changing their pitch, or changing the pitch without affecting their length [2]. The fundamental elements of a granular synthesizer are small acoustic objects, sounds of short duration that can barely be perceived as individual sonic events. By manipulating the position in time of the grain, the overlapping factor of adjacent grains, or individual characteristics of the grain (e.g., frequency), a composer can create different sonic atmospheres. Interesting transformations can also be obtained with granular techniques if the sound grains are captured from real-world signals, rather than computationally generated. Grains extracted from the source signal can be re-arranged, eliminated, repeated, or otherwise manipulated in order to create compelling effects. This approach is known as micromontage or *granulation* [1], [2]. Granular synthesis, based on granulation, and applied to real-world signals, is an extremely successful technique; many of the most popular audio manipulation software suites now include tools for granular synthesis and transformation (see [3] for a extensive list of software tools).

Creative approaches using granular synthesis strategies are not as common in the video domain. However, some previous works do present spatial-temporal manipulations of video signals which are related to the ones we present here. For instance, one popular technique known as *slit-scan* is also based on a transmutation of the time and space axes. A repository of artworks based on the slit-scan technique has been put together by Golan Levin [4]. Our video granulator software, described below, can be used to produce visual outputs similar to the ones obtained with an slit-scan (see Fig. 7), but it can also be generalized to include geometric manipulations, for example, to allow the signal to be perceived from unusual, i.e., not front-facing, points of view. In many image-based Non-Photorealistic (NPR) effects, sets of pixels are grouped and replaced by synthesis elements that can, for instance, simulate brushstrokes or provide other kinds of creative manipulations. But with most NPR techniques the internal pixel information is usually "smoothed over" and does not remain part of the output [5]. Fels et al. [6] showed a re-interpretation of the space-time cube of a video signal and presented different alternatives to shuffle these two domains. Our technique differs in that it manipulates a larger perceptual entity, the video grain, rather than individual pixels. Alvaro Cassinelli also created a pixel-based interactive piece that allows navigation in time and space using a tangible surface [7], and his examples and experiments with moving objects are relevant to our investigation. A combination of NPR synthesis with space-time analysis is presented by Klein et al. [8]. It uses a set of different "rendering solids" to recreate a NPR version of the input. In some sense, their rendering solids are similar to the time-varying envelope that we use to create spatial grains, but their method is not intended as an extension of granulation techniques. The work described in this paper also relates to previous work by the authors on video processing and analysis, including [9] and [10].

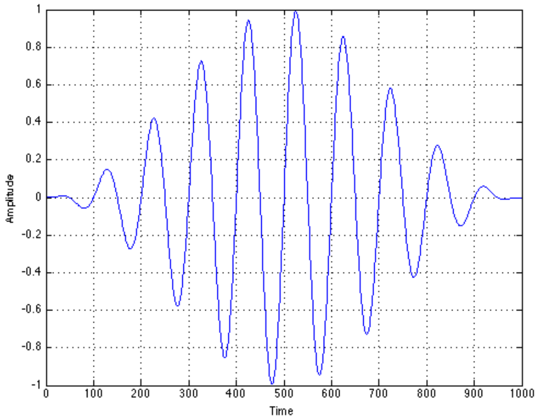

Figure 1: An audio grain

Human perception of audio and video streams has strong

Figure 2: A video grain, the windowing function is applied on the spatial and temporal dimensions.

differences. Extending the concept of granulation to video domains demands a new exploration of the creative possibilities of such techniques. Below we present our ongoing work on the exploration of such alternatives. We present our initial implementation of a *video granulator* and show how basic audio techniques like cloning, skipping, or grain-shuffling can also be creatively applied to video signals. Finally, we provide examples how how the spatial-temporal organization of grains can be manipulated and demonstrate content-dependent manipulations on video signals.

II. SYSTEM OVERVIEW

Similar to an *audio grain*, a *video grain* is a portion of an input video signal windowed by an envelope. In audio granular synthesis, different envelopes are used to overlap regions of the input signal, and can be chosen by a composer for particular effects [1]. In adapting granular synthesis to the video domain, we applied a Hann window envelope to video signals since they create grains with a uniform overlap characteristic [11]. A audio grain is depicted in Fig. 1, and Fig. 2 shows the spatial-temporal windowing of a video grain.

Fig. 3 provides an overview diagram of our video granulator system, transmuting an input video signal into a creatively manipulated output video signal. An input video is interpreted as a *video cube* of three dimensional data (step 1), made up of video frames – the x and y coordinates – extended through time – the z coordinate. According to parameters that define the input window size and various factors that define the amount of overlap, the starting position of each grain in each of the three dimension is calculated and stored in a *grain map* (step 2). This grain map contains information about how the position of each grain created from the input video cube is mapped to an output signal. It is in the construction of the grain map where different manipulations such as cloning, skipping, shuffling grains, or changing overlapping factors can be generated. A scheduler component looks at the output of the grain map to determine what portion of each grain should be used to build the current output frame (step 3).

III. CREATIVE MANIPULATIONS

In this section we will present some of the video manipulations that can be obtained with our granular approach.

A. Grain-Based Manipulations

Specific grains can be *cloned* (added multiple times to the grain map) or *skipped* (not included at all in the grain map). We explored examples where we created video grains from the input video cube using a 50% overlapping factor. On the creation of the output, we either chose some grains to be skipped or cloned, effectively changing the size of the image while preserving the spatial-temporal frequencies. Fig. 4 shows a frame of the video granulator after a cloning operation was performed. The image has a curious resemblance to the "op art" artwork created by Julio Le Parc [12].

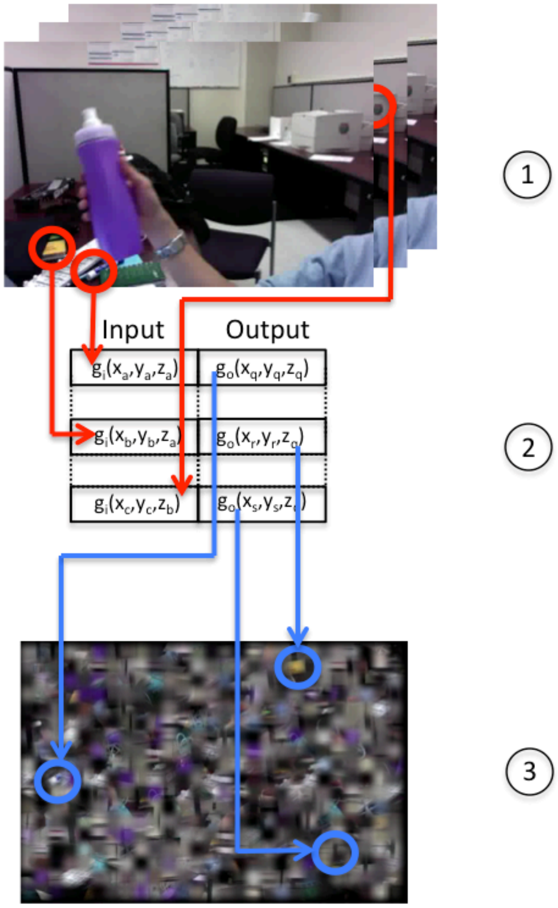

Figure 3: A overview of our *video granulator* system. Step 1 shows the input video cube; step 2 shows the grain map; and finally step 3 shows one frame of the final manipulated video output. The red arrows show example grains from the video cube being added to the grain map, and the blue arrows show how these grains are repositioned temporally and spatially into an output frame.

We also explored randomizing the position of the grains. That is, we altered the spatial and temporal aspects of the grains in different ways. In Fig. 5, we show an arbitrary permutation along all axes. Fig. 6 shows an example frame from an output video which used a grain map that shuffled gains only along the temporal axis. We plan to explore more controlled manipulations of the position of the grains that should lead to interesting visual effects.

B. Spatio-Temporal Reinterpretation

By considering the input video as a cubic array of grains, new interpretations of the data can be created simply by relocating the *point of view* of the array. That is, we can imagine the video being played from a different direction. Fig. 7 shows a

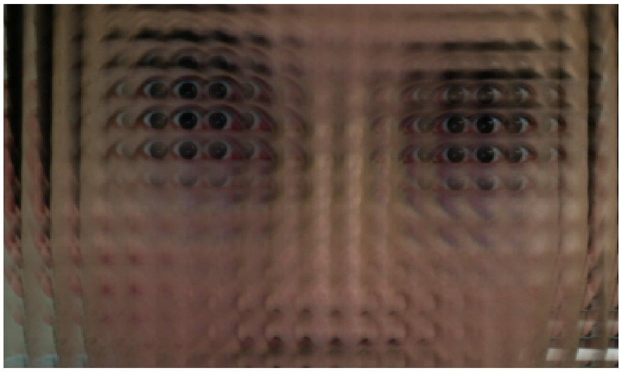

Figure 4: A frame showing the cloning of video grains.

Figure 5: A frame showing arbitrary permutation of grains along the spatial and temporal axes.

Figure 6: Shuffling the time position of some of the grains.

frame obtained while looking at the video array from one side (where a space axis and the time axis, the x and z axes, are interchanged). Fig. 8 shows a view from one of the corners.

C. Image Dependent Manipulations

The position of grains can also be modified by other, non-procedural strategies. For example, higher level information from the video stream can be used to determine the behavior of the grains. In one of our explorations we changed the spatial

Figure 7: The video cube of grains viewed from the side.

Figure 8: The video cube of grains viewed from one corner.

position of the grains according to its temporal variance. The squared root of the temporal variance for a grain spatially placed at x, y is:

$$\bar{\sigma_G}\left(x_0, y_0\right) = \frac{1}{G_s^2} \sum_{x=x_0}^{x_0+G_s} \sum_{y=y_0}^{y_0+G_s} \sqrt{\mathrm{VAR}\left[p_{x,y}\left(t\right)\right]} \qquad (1)$$

Where G_s is the grain size (assumed to be equal in all dimensions), $p_{x,y}(t)$ is the gray-scaled pixel value at position x, y and time frame t, and the variance is calculated over time. The position of the grains that have a variance greater than a predefined threshold is altered in a random (but pre-calculated) direction by an amount proportional to the square of the averaged temporal variance of the grain. Fig. 9 shows a frame of a video of person's hand waving back and forth, illustrating this dynamic manipulation.

IV. CONCLUSION AND FUTURE WORK

This initial work shows that creative manipulations with a granular approach are also possible with video signals. Although there are strong differences in the way in which human perception works in the two domains, some of the strategies can be extended in a very straightforward way. This is the case when we clone or skip grains. In the audio domain, this strategy is commonly used to modify the length of an audio signal without changing its pitch. But this trade-off is not as meaningful when processing visual information. While, for instance, relative small changes on the reproduction rate of a voice signal can immediately sound artificial, we are used to seeing faces at different scales. However, although this pitch-preserving time-modification is not as important

Figure 9: The position of grains in the regions with greater temporal variance is modified.

in images (an exception might involve periodic textures), even a straightforward application incorporating cloning and/or skipping grains produces visually interesting results.

Randomizing grain placements is another strategy that is also used in audio. Our initial experiments lead us to believe that applying this strategy to video signals has significant narrative potential. Different video atmospheres can be created by controlling the amount of randomization and the dimensions (spatial or temporal) to which it is applied. Through presenting the video cube of grains from different directions, diverse interpretations of the same block of visual information can be generated. The spatial and temporal dimension can interchange roles reveal interesting differences about the perception of time-evolving time versus space-changing data.

Our last example illustrates the versatility of our proposal. The dynamics of the grains can be controlled with high level features that introduce many possibilities for *interactive* systems. In this example we used the temporal variance of the grains, but in future explorations we will use more sophisticated measurements, such as optical flow, to control video manipulation parameters. Other image characteristics that can be used to condition the grain behavior include luminance, chrominance, frequency content, spatial position, or even features that measure the size of a face or how close it is to the camera. It is our intention to continue exploring other creative possibilities, such as, for instance, using a variable grain size, filtering the grains in various ways, utilizing a swarming or flocking algorithm on the grains, among others. We also believe that these techniques would enhance some of our existing research into creative video processing techniques and applications [13]–[15].

We have discussed the differences and similarities of granulation in the audio and video domains, but an interesting field for future exploration is the interaction *between* both streams. Some of the effects explained on this document, including cloning and randomization, can be applied to audio and video signals. Highly coupled audio-visual pieces can be generated if both streams undergo similar operations simultaneously. Moreover, an audio granular synthesizer can be used to generate the soundtrack of a granular video, and audio events can be triggered by the video grains. Characteristics such as the density or frequency content of the audio grains can be mapped to other features in the video grains. Finally, spatialized sound and 3D audio compositions that place audio grains at different positions within a virtual space could be complemented by video grains moving with similar dynamics throughout an immersive environment.

One outcome we are currently pursuing is the creation of a more versatile graphical tool similar to the ones that exist for audio granulators and granular synthesizers [3]. A tool for the real-time granular manipulation of video streams would be useful for promoting novel techniques, and would enable their use in a variety of creative situations, including the creation of musical videos, or in live performances by VJs and other visual experimenters.

REFERENCES

[1] C. Roads, Microsound. The MIT Press, Sep. 2004.

[2] U. Zölzer, et al, DAFX: Digital audio effects. Wiley, May 2002.

[3] T. Opie, “Granular synthesis: A granular synthesis resource website,” Available: http://granularsynthesis.com/software.php, Retrieved: December, 2013.

[4] G. Levin, “An informal catalogue of slit-scan video artworks and research,” Available: http://www.flong.com/texts/lists/slit_scan/, Retrieved: December, 2013.

[5] T. Strothotte and S. Schlechtweg, Non-photorealistic computer graphics: Modeling, rendering and animation. Morgan Kaufmann, 2002.

[6] S. Fels, E. Lee, and K. Mase, “Techniques for interactive video cubism (poster session),” in Proceedings of the eighth ACM international conference on Multimedia (MM). ACM, 2000, pp. 368–370.

[7] A. Cassinelli. and M. Ishikawa, “Khronos projector,” in Emerging Technologies SIGGRAPH 2005. Los Angeles, CA: ACM, 2005.

[8] A. W. Klein, P.-P. J. Sloan, A. Finkelstein, and M. F. Cohen, “Stylized video cubes,” in Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA). New York, NY, USA: ACM, 2002, pp. 15–22.

[9] J. Villegas and A. G. Forbes, “Interactive non-photorealistic video synthesis for artistic user experience on mobile devices,” in Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Scottsdale, Arizona, January 2014.

[10] A. G. Forbes, C. Jette, and A. Predoehl, “Analyzing intrinsic motion textures created from naturalistic video captures,” in Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP), Lisbon, Portugal, January 2014.

[11] G. Heinzel, A. Rudiger, and R. Schilling, “Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows,” Max-Planck-Institut fur Gravitationsphysik, Tech. Rep. 395068.0, February 2002.

[12] E. Guigon and A. Pierre, L’oeil moteur: Art optique et cinétique, 1950-1976. Strasbourg, Germany: Muses de Strasbourg, 2005.

[13] C. Roberts, A. G. Forbes, and T. Höllerer, “Enabling multimodal mobile interfaces for musical performance,” in Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), Daejeon, Korea, May 2013.

[14] A. G. Forbes, T. Höllerer, and G. Legrady, “Generative fluid profiles for interactive media arts projects,” in Proceedings of the International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging (CAe), Anaheim, California, July 2013, pp. 123–129.

[15] J. Villegas and A. G. Forbes, “Double-meaning: Interactive animations with simultaneous global and local narrative,” in Proceedings of the Re-new Digital Arts Festival, Copenhagen, Denmark, October-November 2013, pp. 300–304.