Adding Dynamic Smoothing to Mixture Mosaicing Synthesis

Graham Coleman*, Jordi Bonada*, Esteban Maestre*†
*Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain
Email:graham.coleman@upf.edu, jordi.bonada@upf.edu
†Center for Computer Research in Music and Acoustics
Stanford University, Stanford, USA
Email: esteban@ccrma.stanford.edu

***Abstract*—Recent works in sound mosaicing synthesis [1], [2] have proposed algorithms that permit instantaneous mixtures of several sources atoms, based on sparse signal representation techniques. We propose combining l_1 regularization with linear dynamical smoothing as in the Kalman filter (also in [3], [4]) to promote desired transitions between atoms, while adapting the generic approach to the mixture mosaicing context. Furthermore, we modify the dynamics cost slightly to further promote sparse scores in the case of non-negativity. This is a work in progress in which we can present some sound examples, but for which the proposal is not fully validated.**

I. Introduction

Mosaicing, a form of sample-based sound synthesis, consists in transforming and compositing disparate source sound segments from a database so that the result will match perceptual features (descriptors) of a target sequence. Classical methods [5], [6] considered matching a single source segment to a given target context (frame), while more recent methods [1], [2] consider sparse mixtures of multiple source segments at once.

Several criteria for these systems concern the dynamics–the changes from frame to frame–of the composition, or more abstractly, the sampling process that generates it. Do the descriptors of the source units change too much from frame to frame (continuity)? Do the transformation parameters applied to the sources change rapidly from frame to frame (transformation continuity)? Does the sampling process maintain a steady context within the source material by choosing contiguous blocks of material from the original source context, or does it jump around (contiguity)? By modeling dynamics we can search or sample sequences or mixtures that have desired properties above.

II. Proposal

Given a matrix or linear operator D describing favored atom transitions from state to state, one way of generalizing it to mixtures of atoms is simply considering a form of linear dynamics where: $x_{t+1} = Dx_t + w_t$. x_t and x_{t+1} are mixture vectors for time steps t and $t+1$, w_t represents innovation, or deviance from expected dynamics.

Combining the smoothing version of the Kalman filter with an l_1 regularization term as in Basis Pursuit Denoising (BPDN) would give us the following program:

$$\min_x \sum_{t=1}^{T} \|Ax_t - b_t\|_2^2 + \lambda_1 \sum_{t=2}^{T} \|Dx_{t-1} - x_t\|_2^2 + \lambda_2 \sum_{t=1}^{T} \|x_t\|_1 \quad (1)$$

Under the scheme given by Problem 2, if the transition matrix D gives a number of possibilities for a given atom, the most likely successor state (where the innovation cost is zero) will include nonzero weights on all of those possibilities. Therefore, when D includes many alternatives for transitions between atoms, the innovation cost and the sparsity cost are working against each other.

In our application, where weights are constrained to be non-negative, we propose modeling *alternatives* using an innovation cost where only positive innovation is penalized, that is having no cost when weights decrease (state is closer to sparsity than deterministic dynamics). We implement this by introducing a non-negative dummy variable y:

$$\min_x \sum_{t=1}^{T} \|Ax_t - b_t\|_2^2 + \lambda_1 \sum_{t=2}^{T} \|Dx_{t-1} - x_t - y_t\|_2^2 + \lambda_2 \sum_{t=1}^{T} \|x_t\|_1 \quad (2)$$

where both x and y are constrained to be elementwise non-negative. In this scheme, successor states are not penalized for atom transition alternatives not taken, only for unlikely transitions that are taken.

III. Other Approaches

Several other approaches are also likely feasible for encouraging dynamics in synthesis. For one, we could extend the Kalman filter objective with a nonlinear model, which may render the objective function non-convex. In this case, heuristic methods based on convex relaxation such as DC Algorithms (DCA) could be used to find heuristic solutions quickly.

Sampling or Monte-Carlo approaches are also feasible. In particular, particle filters (Sequential Monte Carlo) have been used for tracking, and allow both nonlinear dynamics, and use non-parametric estimates for the states.

Finally, so called greedy signal decomposition methods could perhaps be adapted to account for dynamics. In practice this would be analogous to sampling in many ways. Perhaps a good example of this in image synthesis would be Ashikhmin [7].

References

[1] M. D. Hoffman, P. R. Cook, and D. M. Blei, "Bayesian spectral matching: Turning young MC into MC hammer via MCMC sampling."
[2] G. Coleman, E. Maestre, and J. Bonada, "Augmenting sound mosaicing with Descriptor-Driven transformation," in *Proceedings of DAFx 2010*, Graz, Austria.
[3] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation."
[4] M. S. Asif, A. Charles, J. Romberg, and C. Rozell, "Estimation and dynamic updating of time-varying signals with sparse variations," 2011.
[5] A. Zils and F. Pachet, "Musical mosaicing," *Digital Audio Effects (DAFx)*, 2001.
[6] D. Schwarz, "Data-Driven concatenative sound synthesis," Ph.D. dissertation, 2004. [Online]. Available: http://recherche.ircam.fr/equipes/analyse-synthese/schwarz/thesis/
[7] M. Ashikhmin, "Synthesizing natural textures," in *Proceedings of the 2001 symposium on Interactive 3D graphics*, ser. I3D '01, 2001, p. 217–226, ACM ID: 364405.