Combining Sampling and Autoregression for Motion Synthesis

David Oziem, Neill Campbell, Colin Dalton, David Gibson and Barry Thomas
University of Bristol
United Kingdom
oziem@cs.bris.ac.uk

Abstract

We present a novel approach to motion synthesis. We show that by splitting sequences into segments we can create new sequences with a similar look and feel to the original. Copying segments of the original data generates a sequence which maintains detailed characteristics. By modelling each segment using an autoregressive process we can introduce new segments and therefore unseen motions. These statistical models allow a potentially infinite number of new segments to be generated. We show that this system can model complicated nonstationary sequences which a single ARP is unable to do.

1 Introduction

In this paper we present a novel approach to motion synthesis. Our method is based on recent texture synthesis and video texture research. We use a copying based approach where the sequence is divided into temporal segments and the output is generated from these segments. Individually this method works well but does not provide any new unseen information. Using a statistical model called an autoregressive process (ARP), we generate models of the segments to create new motions of similar look and feel. This gives us a robust method for generating new sequences where the worst case is a completely copied output. The difficulty in manually animating characters is replicating the idiosyncrasies of natural movement. A slight stumble or change in stride length can add much more life to a character than a perfect walk. An example of which is the angle of a cranes wing in flight. The wing sweeps down slowly and rises quickly with slightly different phase and amplitude each time. These idiosyncrasies add a more realistic feel to an animation and adding random noise to a sine wave does not produce this effect. Motion capture systems have been around for several years and many thousands of motions have been accumulated. Successful methods of analysing the data and recreating motion are hugely beneficial saving both time and money. Motion synthesis provides an efficient and cost effective tool for the film and games industry. Thousands of background characters can be synthesised using a single extracted motion.

2 Related Work

This section outlines the selected previous research in the area of synthesis. Synthesis here is defined as generating new copies of a dataset which follow the original pattern. Synthesis is used in many areas of research and is often referred to as modelling. One the targets of our present research is to generate new video footage, which is longer than the original input data. A previous method used to model video textures is an autoregressive processes. An *autoregressive process* (ARP) is a parametric modelling technique. The main principle is that every point in a sequence is a linear combination of *n* previous values. The number of "lagged" values, known as the order, effects the fit of the model and also increases the required amount of training data, see "System Identification" [11] for details on calculating an ARP. ARP's have been used in tracking and for synthesising video textures (a temporal texture). Initial attempts at modelling video textures by Schödl et al. [13] used method of reorganising the frames of the video texture such that a new clip could be generated which was potentially infinitely long. Unfortunately, if the sequence does not contain frames which are close in pixel values and preferably temporally far apart then the sequences quickly reach a dead end or get stuck in a loop. Campbell et al. [5] and Reissell [12] used various forms of ARP to model and recreate video textures. In particular it was shown by Campbell et al. [5] that if a data set has a Gaussian distribution it is likely that an ARP will produce a good model. Although we are looking at motion other synthesis techniques are equally relevent. Texture synthesis involves slightly different issues to motion synthesis but many of the techniques can be transfered. The main theme of the most successful techniques is sampling i.e. the selection and reordering of values or blocks from the original data. Multi-scale sampling was first proposed by Heeger [7] and then later by DeBonet [2]. The concept is that a pyramid hierarchy of the texture is created with low to high resolution copies of the original. Sampling can then be done following the hierarchical structure from low to high resolutions. Efros and Lueng [6] presented work in which they modelled the texture as a Markov chain i.e. each pixel is associated to the surrounding pixels. Wei and Levoy [15] and Hertzmann et al. [8] integrated DeBonet's [2] multiscale paradigm and Efros and Lueng's Markov chain, to produce a superior texture synthesis process. Other sampling techniques have been presetned more recently which copy blocks of texture rather than individual pixels [10, 17]. The main aims of recent work into motion synthesis of motion capture data can be split into two categories. The first is to combine motion signals to produce longer sequences of motion [1, 9]. This process involves a user inputting the desired effect and then a search to attempt to fit a motion to these constraints. The second grouping are those papers which attempt to generate new motions from a single motion sequence. The applications include extending clips or generating a new copy of the original [3, 4, 16].

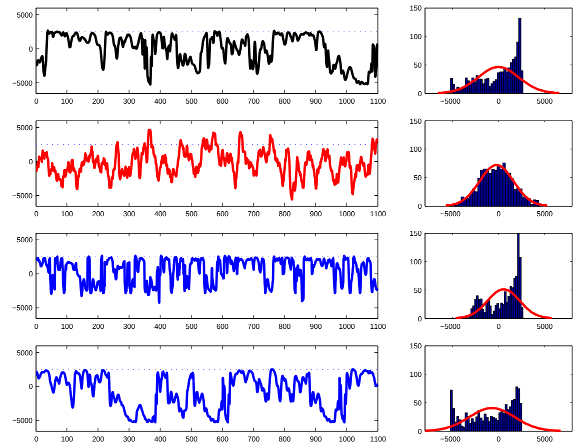

Figure 1. Unnormalised first mode and histogram of the original candle flame sequence (First row), an ARP generated version (Second row), a multi-scale sampled version (Third row), a tiled version (Fourth row).

3 Problem Analysis

The current techniques for synthesis tasks include many difficulties when applied to motion synthesis. The high dimensionality of motion signatures is one of the major causes as well as non-Gaussian distributions and nonstationary signals. We present a new combined technique which integrates two known techniques to produce a hybrid system. Firstly an analysis of current techniques substantiates the choices made when designing our hybrid system. When modelling with an ARP the original dataset must have a Gaussian distribution. Figure 1 shows a non-Gaussian sequence which is the first mode of the eigen decomposition of a video sequence of a flickering candle flame. A version generated using an ARP is included which has a Gaussian distribution as will always be the case over large samples. In Figure 1 it can be seen that the nonstationarity of the sequence is lost in the ARP simulation. In the original data the sequence maintains a level around two thousand five hundred then dips at irregular time intervals. This is lost in the ARP generated version with the sequence peaking at a variety of values. However, it has been shown [5, 12] that when given correctly distributed data an ARP provides an excellent method of producing new unseen data. A multi-scale approach provides an excellent model of the low level structures within the input sequence. By modelling from low resolution to high, it maintains the low level structures and therefore nonstationarities are well modelled. Unfortunately, we have found that the technique does not extend well to high dimension data. We produced a pyramid for each dimension of the data and the cost function for each new point is calculated over all dimensions and n previous values. When trying to find the best next point over a large number of dimensions unless the dimensions have a very constrained relationship, the match found is unlikely to be a good fit every time. This introduces a large amount of high frequency noise in the generated sequence. Copying tiles or segments from the original sequence maintains high resolution features. Copying segments guarantees that most adjacent values will be appropriate to the original. Each segment is chosen from only the previous segment, which causes certain global structures to be lost. This is shown in the bottom plot in Figure 1. The sequence includes more of the dipped segments than the original. Running the generation process many times may produce a well structured sequence but this isn't guaranteed. The search process takes no account of the global structure and so repeated segments appear.

4 Method

This paper relates to any multi dimensional temporal sequence. In particular, we have concentrated on video textures and model animation. A video is a series of pixel based frames and model animation refers to joint rotations. With videos we concatenate each row of pixels to produce a 1xM array for each frame. This then creates a very high dimension sequence and this is a major difficulty when trying to carry out any statistical modelling. For example for a multivariate ARP the size of the parameter vector is related to the number of dimensions squared times the order of the

ARP. Hence the number of unknowns increases exponentially with the number of dimensions therefore we use Principal Components Analysis (PCA) to reduce the complexity of the sequences. With model animation the above is still relevant as models may include tens even hundreds of joints each potentially involving x, y and z rotations. Next the sequence must be separated into a set of overlapping temporal segments. For the purposes of this paper we have kept a constant segment size but there may be benefits in varying their size. A larger segment which is stationary and Gaussian has a much higher chance of being modelled correctly using an ARP. The segment size is one of the inputs to the system. Now that we have a smaller subset of the sequence it may not contain the same non-lineararities as the whole sequence. Using PCA on each segment we can further reduce the dimensions of the segments without introducing significantly more error in reconstruction. This also introduces constraints on the generated segments which helps maintains their look and feel and reduces the number of unknowns helping to produce better models. ARPs are trained for each segment. The parameter estimation phase chooses the "best" model from the given dataset. Here an important question is whether the "best" possible model is good enough to model the segment correctly. We look to validate the model, especially with small segments of motion where a perfect fit is unlikely. To address this problem we have two solutions firstly we generate sequences using the ARP's and then compare range, mean and variance. Our second filter is Schwarz's Bayesian Criterion (SBC) [14]. The fundamental use of SBC is for model order selection. We extend this so we can use it as an initial estimate of the segments which will be modelled well using their respective ARP's. The synthesis phase starts by chosing a random segment. Following segments are chosen such that the final values of the last segment correspond to the beginning of the new one using the the root mean squared difference. Finding the lowest three values provides us with the best matches from which segments are then randomly selected. We use a single frame overlap and find the average between the layers to help hide the seams. Using SBC to select which segments to model using an ARP introduces a trade-off, since setting the threshold too high can include badly modelled segments and setting it too low will result in tiling with no new segments. A number of segments are generated and from these segments ones which exceed the range of the original segment are removed. Tests on the variance and mean are also used to eliminate incorrectly structured segments.

5 Results

We have found that the segmented ARP process can produce results which fit our original aim. High resolution characteristics are maintained and new unseen data is synthesised and integrated well into the sequence. Figure 2 shows one of our synthetic signals which is a sum of three sine waves. The results from the ARP model show a sine wave with a upward trend whereas our model provides us with a similar sequence to the original and includes new segments. The number of modelled segments varies greatly between output sequences. Figure 3 shows two randomly chosen output sequences and an ARP sequence for the candle flame sequence. The original candle flame is a short video clip of about 1100 frames. Using PCA we transformed the sequence into a 20D space with a reconstruction error of 6.8%. The first and second modes are shown in Figure 3 and the first mode in Figure 5. It is a particularly hard sequence to model because it is non-linear and nonstationary. The sequence does not follow a periodic pattern, however, both of the generated sequences are good matches to the original and have the same concentrated curve along the bottom right and randomly cut across into the centre of the plot. Approximately 20% of the generated sequences are new unseen segments. The barman sequence originated as a manually animated model. Here we show the application of our method on a sequence of joint rotations. The barman sequence is similar to the candle flame in the complexity of the first two modes (Figure 4). In the original clip the barman mostly wipes a small area in front of him but takes one large sweep of the surface. This can be seen in the concentrated area in the top left and then the long sweeping curve which steams from it. An ARP generated version of the sequence does not maintain any of the original characteristics. Our generated versions maintain the low and high resolution characteristics and they include approximately 15-25% of segments which have been generated using an ARP. The new sequences are particularly appealing as they maintain the general feel of the original but include a large perceptual

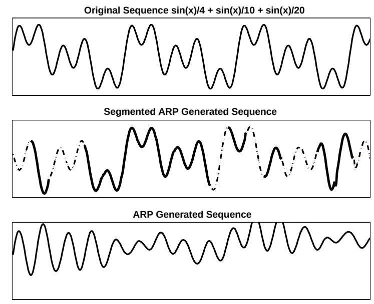

Figure 2. (Top) Original sequence. (Centre) Segmented ARP generated sequence, dotted portions represented copied segments and solid lines represent ARP generated segments. (Bottom) ARP Generated sequence.

variation.

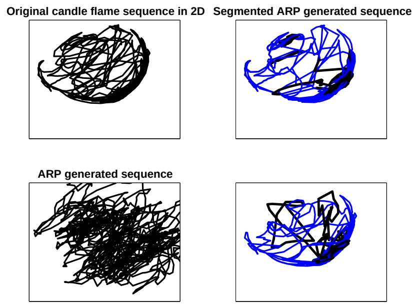

Figure 3. (Top left) Original candle flame sequence, reduced to 2D using the first two modes from PCA. (Bottom left) ARP generated sequence. (Right) Two randomly picked sequences generated with our process. The black lines represent segments generated using an ARP and the rest were copied.

6 Conclusions and Further Work

We have shown that it is possible to produce new sequences using a segment copying based scheme, and by modelling the segments using an ARP we can generate new unseen data. Our method allows us to model nonstationary sequences and produce new unseen data within the generated sequences which has not been previously addressed. Around 20-30% of the generated sequences in the examples shown have been generated using an ARP. One limit of our system is that segment copying has no hierarchical structure. Each segment is only selected on the basis of the previous n frames. There is no consideration of the global structures and so sequences may be produced which are locally correct but globally invalid. In some cases, if the sequence has a single peak which is only reached once, the generation process can get stuck copying the same segments repeatedly. Our technique currently has little user interaction other than a starting point and a level of randomness which is allowed in the sequence. Future work could benefit from allowing variable segment lengths and dimensionality and more advanced crossover techniques. The need for a global structure could be addressed by clustering the segments using a reduced eigenspace and then learn a model of the changes in state.

References

[1] O. Arikan, D. A. Forsyth, and J. F. O'Brien. Motion synthesis from annotations. *ACM Transactions on Graphics (TOG)*, 22(3):402–408, 2003.

[2] J. S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. *Computer Graphics*, 31(Annual Conference Series):361–368, 1997.

[3] M. Brand and A. Hertzmann. Style machines. In Kurt Akeley, editor, *Siggraph, Computer Graphics Proceedings*, pages 183–192. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[4] A. Bruderlin and L. Williams. Motion signal processing. *Computer Graphics*, 29(Annual Conference Series):97–104, 1995.

[5] N. W. Campbell, C. Dalton, D. Gibson, and B. Thomas. Practical generation of video textures using the auto-regressive process. In Paul L. Rosin and David Marshall, editors, *British Machive Vision Conference*, pages 434–443, Cardiff, UK, September 2002. British Machine Vision Association.

[6] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *ICCV (2)*, pages 1033–1038, 1999.

[7] D. J. Heeger and J. R. Bergen. Pyramid-based texture analysis/synthesis. In *Siggraph*, pages 229–238, 1995.

[8] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In Eugene Fiume, editor, *Siggraph, Computer Graphics Proceedings*, pages 327–340. ACM Press / ACM SIGGRAPH, 2001.

[9] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 473–482. ACM Press, 2002.

[10] V. Kwatra, A. Schdl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. *ACM Transactions on Graphics (TOG)*, 22(3):277–286, 2003.

[11] L. Ljung. *System Identification, Theory for the user*. Prentice Hall, 2nd edition, 1999.

[12] L.-M. Reissell and D.K. Pai. Modeling stochastic dynamical systems for interactive simulation. In *Computer Graphics Forum*, volume 20, pages 339–348. Blackwell Publishing, 2001.

[13] A. Schödl, R. Szeliski, D. H. Salesin, and I. Essa. Video textures. In Kurt Akeley, editor, *Siggraph, Computer Graphics Proceedings*, pages 489–498. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[14] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.

[15] L. Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In Kurt Akeley, editor, *Siggraph*, *Computer Graphics Proceedings*, pages 479–488. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[16] A. Witkin and Z. Popović. Motion warping. *Computer Graphics*, 29(Annual Conference Series):105–108, 1995.

[17] B. Guo; H. Shum; Y.-Q. Xu. Chaos mosaic: Fast and memory efficient texture synthesis. Technical report, Microsoft Research, April 2000.

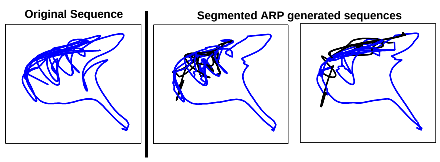

Figure 4. (Left) Original barman sequence, reduced to 2D using first two modes from PCA. (Right) Two randomly picked sequences generated with our process. The black lines represent segments generated using an ARP and the rest were copied.

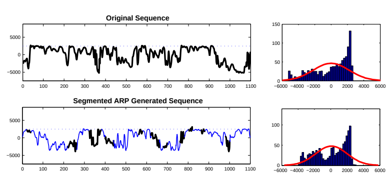

Figure 5. (Top left) Original candle flame sequence, reduced to 1D using the first mode from PCA. (Bottom) Randomly picked sequence generated with our process. The black lines represent segments generated using an ARP and the rest were copied.