Synthesizing Transient Traffic of Temporal Visual Signals for Discrete Event Simulation

Mauritz Panggabean and Leif Arne Rønningen

Department of Telematics
Norwegian University of Science and Technology (NTNU)
N-7491, Trondheim, Norway
{panggabean, leifarne}@item.ntnu.no

Abstract—This paper presents the analysis and modeling of transient traffic source of temporal visual signals for discrete event simulation. Transient nature implies the inability of applying established traffic theory that assumes stochastic processes. The focus is on the transient traffic of foreground object due to its motions as well as camera panning and zoom. We show that such transient traffic can be well modeled as power functions. Implemented for discrete event simulation in DEMOS, the model with a number of parameters can synthesize transient traffic of temporal visual signal in piecewise manner. Our simulation results reveal that arbitrary form of transient traffic can be synthesized in discrete event simulation as a series of the power function model with different parameter values. An exemplary application of the model in the simulation of a complex queuing system is given within our current research in futuristic multi-camera tele-immersive collaboration system.

I. INTRODUCTION

The audiovisual traffic generated as an input to a queuing system often has non-linear and transient behavior. This means that established stochastic theory assuming stationary processes cannot be applied in the study of such phenomena. However, aggregated control packet streams may be modeled using established queuing theory [1].

In this work we assume that an input video consists of scene with important objects that attract the most viewer’s attention through eye gazes. We are interested in the data traffic that represents only one particular important object and not the entire content of the video. Thus, instead of the term video in general meaning, we use the term temporal visual signals to refer to such visual information.

Seen from the sources, audiovisual data are generated as packets which rate can be constant for some time, and then instantly step to another value. We are not generally interested in mean values, but rather the transient slopes and their variations, and extreme values and their duration. Therefore, it is best to analyze the traffic in piecewise manner where each piece of traffic will be either stationary or transient, as illustrated in Figure 1. The pictured traffic can be segmented into nine sub-traffic which mostly appear with transient properties. Each transient piece can be analyzed and modeled.

The objective of this work is investigate a correct model that in general can represent any transient traffic with just a few parameters. The model will be implemented in a tool for discrete event simulation to synthesize arbitrary transient traffic of temporal visual signals. It must have real physical meaning, for example, a person entering and disappearing from the scene, the act of camera zooming in an object, or an object performing particular motions in the scene.

This paper adopts the following structure. Section II presents the analysis and modeling phases of the work. The results from implementing the model in discrete event simulation are covered and discussed in Section III. An application of the model in discrete event simulation of a complex queuing system is exemplified in Section IV which include some previous work that form the basis for this paper. Final remarks come in Section V that concludes the paper.

II. FROM TRANSIENT SIGNALS TO SIMULATION MODELS

We focus on the transient traffic from uncompressed visual signals of a person as the object in three video clips with the following scenarios where the camera is always static.

1) Video clip PANNING shows the object entering the scene from the left side until disappearing on the right side, equivalent to the panning motion of the camera.
2) In the video clip ZOOM, the object gradually moves closer to the camera, analogous to a zooming-in camera.
3) The object performs some motions with the limbs at the center of the scene in video clip MOTION.

The three video clips are recorded using an off-the-shelf camcorder with 1920×1080-pixel resolution and 30Hz frame rate. A person is featured with the upper-half body in front of a background with the same color to ease the object segmentation. The clips are transcoded into de-interlaced color.
AVI format at 1280×720-pixel resolution to simplify the color-based object segmentation. Matlab is used on a PC with a 2.99GHz processor and 8.00GB RAM that processes 10-second input video clips.

From each video clip, a number of transient parts are selected for analysis and modeling. This yields sequences PANNING, ZOOM, and MOTION which consist of 94, 293 and 301 frames, respectively. Figure 2 depicts exemplary frames of the three cut-out sequences and the resulting traffic from the sequences are shown in Figure 3. The actual traffic assumes temporal uncompressed three-channel color visual signals. The object consists of arrays of 8×8-pixel blocks and each pixel value is encoded with eight bits.

Let us analyze some transient parts from the actual traffic in Figure 3. Figure 4 plots them together with the results of curve fitting to the data points using Matlab, each with two functions: power function \( f(x) = ax^b \) and linear function \( f(x) = cx + d \). Note that the frame numbers and the data sizes per frame are normalized to be between 0 and 1 prior to curve fitting to simplify the analysis. The stair-case shapes present in the transient traffic of sequence PANNING and MOTION imply the necessity of much higher frame rates to capture more data points with higher accuracy, due to motions of high speed.

The parameter values and the root mean square errors (RMSEs) for the goodness of fit for the fitted curves in Figure 4 are as follows. The numbers in brackets are the low and high bounds with 95% confidence.

- **Sequence PANNING**
  - \( a = 1.108 \) (1.007, 1.21), \( b = 0.8518 \) (0.6703, 1.033), RMSE = 0.09655
  - \( c = 1.116 \) (0.9418, 1.291), \( d = 0.03048 \) (-0.07434, 0.1353), RMSE = 0.104

- **Sequence ZOOM**
  - \( a = 0.9533 \) (0.942, 0.9646), \( b = 1.339 \) (1.308, 1.371), RMSE = 0.03234
  - \( c = 0.9854 \) (0.9665, 1.004), \( d = -0.08577 \) (-0.09675, -0.07479), RMSE = 0.04202

- **Sequence MOTION**
  - \( a = 1.134 \) (0.8431, 1.425), \( b = 0.8762 \) (0.3309, 1.421), RMSE = 0.1677
  - \( c = 1.145 \) (0.6255, 1.665), \( d = 0.02072 \) (-0.3072, 0.3487), RMSE = 0.172
Evidently the power function \( f(x) = ax^b \) gives smaller RMSE values than those of the linear polynomial function, although the differences are small. In fact, linear functions are special cases of power functions where \( b = 1 \). Our goal is to construct a correct model to generate synthetic transient traffic of temporal visual signals for discrete event simulation. As presented in the next section, our results show that the power function is a good candidate for the goal.

III. SIMULATION RESULTS AND DISCUSSION

The power function \( f(x) = ax^b + d + e \) is implemented to synthesize temporal visual signals for discrete event simulation by using DEMOS [2]. A fast and powerful system for discrete event simulation, DEMOS gives much freedom to users to improve and improve, including changing the built-in functions or adding new features, if necessary. The following are the parameters required by the model for synthetic traffic generation in increasing manner as in Figure 4.

1) Parameter \( b \) to define the bending of the curve, where \( 0.5 < b < 1 \).

2) Assuming that the synthetic transient part to generate has \( D_{\text{min}} \) and \( D_{\text{max}} \) as the respective minimal and maximal data size, set parameters \( a \) and \( d \) as \( a = D_{\text{max}} - D_{\text{min}} \) and \( d = D_{\text{min}} \).

3) The values of \( e \) define the random smoothness of the curve at generated points at \( x \), where \( e \) is a random real value that is uniformly distributed between parameters \( e_{\text{min}} \) and \( e_{\text{max}} \). The generation of random value with well-known distributions is a built-in feature in DEMOS. Note that \( e < a \).

4) Frame rate \( F \) and simulation time \( S \) in seconds.

Given all the parameters above, the model can generate the synthetic data rate per frame \( f(x) \) with increasing trend every \( t = 1/F \) seconds where \( x = t/S \). For synthetic transient traffic with decreasing trend, simply use the modified power function \( f(x) = a(1-x^b) + d + e \).

We employ the model to synthesize transient traffic of temporal visual signals that resemble some transient parts of the actual traffic shown in Figure 3. The synthetic traffic of the signals are depicted in Figure 5 where \( e_{\text{min}} = 0 \), \( e_{\text{max}} = 0.5 \) and \( S = 1 \) in all cases. The other parameter values for the traffic generation for the three sequences are as follows: \( a = 4 \), \( b = 0.65 \), \( c = 0 \) and \( F = 19 \) (sequence PANNING); \( a = 12 \), \( b = 1.34 \), \( c = 8 \) and \( F = 230 \) (sequence ZOOM); \( a = 3 \), \( b = 0.75 \), \( c = 7 \) and \( F = 9 \) (sequence MOTION, increasing part); \( a = 2.5 \), \( b = 2.5 \), \( c = 7.5 \) and \( F = 11 \) (sequence MOTION, decreasing part).
One can see from the actual and synthetic traffic that the power function performs very well as the model for generating synthetic transient traffic of temporal uncompressed visual signals. It will be very useful for discrete event simulation that needs such signals as the input. The model serves as the building block to generate any piecewise synthetic transient traffic for simulation. The next section presents an exemplary application of the model in our research for futuristic tele-immersive collaboration systems. It includes extensive discrete-event simulation of a novel complex queuing system.

IV. AN EXEMPLARY APPLICATION

NTNU has recently started an ambitious project to build advanced environments that will support various real-time artistic collaborations such as music, singing, dancing and drama [3]. Using cutting edge technologies such as auto-stereoscopic multi-view 3D displays in all the surfaces, the environments will function also as laboratories in which interesting multi-disciplinary research questions in science, technology and arts will be studied. Table I lists the main technical requirements for the envisioned collaborations related to visual aspects.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Main technical requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Guaranteed maximum end-to-end delay ≤ 10-20ms (cf. 400ms for videoconference [4])</td>
</tr>
<tr>
<td>2.</td>
<td>Near-natural video quality</td>
</tr>
<tr>
<td>3.</td>
<td>Auto-stereoscopic multiview 3D vision</td>
</tr>
<tr>
<td>4.</td>
<td>High spatial and temporal resolution due to life-size dimension of objects i.e. mostly humans</td>
</tr>
<tr>
<td>5.</td>
<td>Accurate representation of physical presence cues e.g. eye contact and gesture</td>
</tr>
<tr>
<td>7.</td>
<td>Quality allowed to vary with time due to different technical specifications among collaboration spaces</td>
</tr>
<tr>
<td>9.</td>
<td>Graceful quality degradation due to traffic overload or failure</td>
</tr>
</tbody>
</table>

Guaranteeing maximal end-to-end delay is essential to facilitate good synchronization in real-time artistic collaborations from distributed places, as shown in, for example, [5] and [6]. The delay can be as low as 11.5ms, including propagation and all processing delays. We estimate the same need in other unexplored forms of artistic collaborations, such as real-time singing practice of several locally distributed singers and remote conducting between a conductor and a choir in different locations.

The Internet and the source-coding approach are not designed to address and deliver such a guarantee for video data. Thus we argue that network nodes should be designed such that they are able to intelligently drop video packets due to immediate traffic conditions with graceful degradation of visual quality. Given the basic definition of data compression as data reduction necessary for storage or transmission of an information, we call this approach compression-by-network (CbN). Figure 6 depicts how it differs from the source-coding approach with raw input digital signals.

CbN is designed to operate on the Distributed Multimedia Plays (DMP) architecture [3], which involves novel queuing systems with advanced control feedback mechanisms. The input to CbN will be uncompressed temporal visual signals that are assumed to be produced by multi-array of cameras in a collaboration space. Simulation will be the most appropriate approach to study and investigate such a system of high complexity, before the physically built system allows extensive measurements. The model proposed in this paper will certainly be useful for such simulations. More details and results from this research are available as reported work, for example, in [7]-[10]. Unfortunately, to our best knowledge, we have not found other reported work on modeling and synthesis of input transient traffic for discrete event simulation.

V. CONCLUSION

We have presented our analysis of transient traffic of temporal visual signals that represents a person in the video scene as the important object. Our study covers three main cases, i.e. panning and zooming modes of the camera, as well as motions of the object. Our objective is to construct a correct model of such transient traffic. The model implementation will be used for synthesizing such traffic in discrete event simulation. Our study reveals that power function can perform as such a model. The results from implementing the model in DEMOS show that the model can be easily employed as a building block to synthesize any transient traffic in piecewise manner. Finally, an example of the model application for discrete event simulation of a complex queuing system is given.

REFERENCES