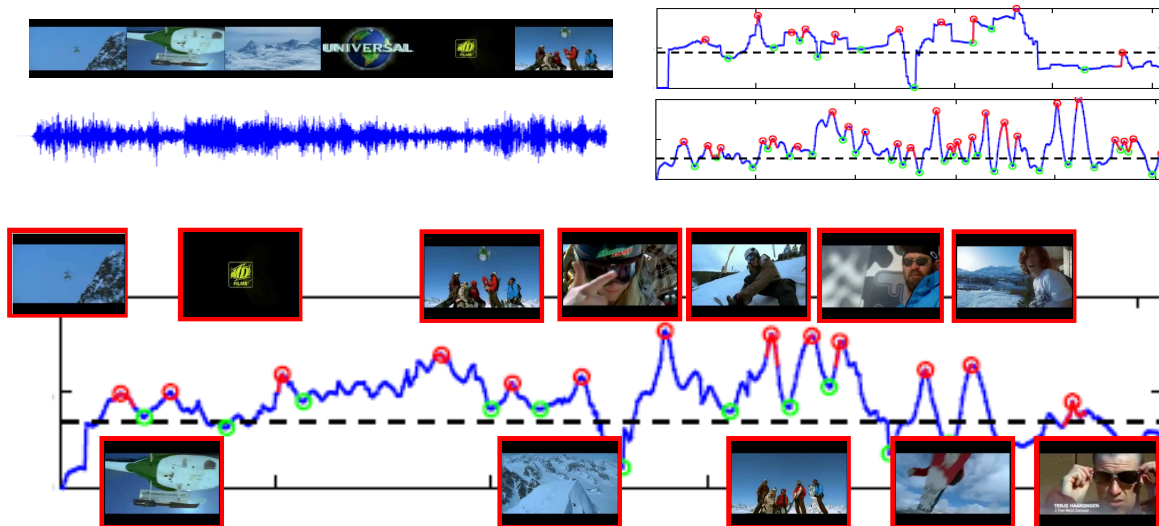


Contributing Partner: ICCS-NTUA (Muscle Participant No 23)

Title: “AudioVisual Saliency Curves (AVsal) & Movie Key-frame Selection Tool”

Type: source code/demo

Author(s) /Developer(s): G. Evangelopoulos, K. Rapantzikos, P. Maragos



Overview

AVsal is a Matlab-based implementation of the core of the movie summarization system, and in specific audio, visual and integrated audiovisual saliency curve extraction. Saliency curves model viewer attention and importance of information by a single ‘saliency’ value per time instance. Attention is modeled separately for each stream (audio and visual) and the potential of inter-module fusion is explored for audiovisual event detection.

Audio attention is captured by signal modulations and related multifrequency band features, extracted through nonlinear operators and energy tracking. The features are fused (linearly or non-linearly) in a single audio saliency curve. Visual saliency is measured by means of a spatiotemporal attention model driven by various feature cues (intensity, color, motion). Features from both modules mapped to one-dimensional, time-varying saliency curves, from which important audio or visual events can be detected through adaptive, threshold-based mechanisms. Audio and video curves are integrated linearly in a single attention curve, where events (i.e. geometrical features) may be enhanced, suppressed or vanished.

As a part of the overall movie summarization system, the software is incorporated in a movie key-frame selection tool, where the most important frames (aurally and visually) are used to form a static video summary (or static storyboard). Given any video input the tool returns a set of key-frames that correspond to the most salient movie units.

Software Description

Full Matlab source-code is distributed along with a demo script that implements the complete system and demonstrates the use of the various functions, parameter definition files (*defineParams*, *asalParams*, *vsalParams*) that can be easily modified by the user and an input function/utility (*mmread*) capable of reading any multimedia video file.

The included *readme.pdf* file provides documentation, file-structure and example usage information.

Software Architecture

AVsal has been designed modularly and provides functions in three different categories:

- Audio feature extraction and integration (linear or nonlinear): Short-time modulation-based features and fusion, linear or non-linear (energy-weighted), to a single saliency curve.
- Video features (intensity, color, motion) and frame saliency map construction.
- Integration: audio-to-video rate mapping, feature normalization and linear (dynamic, user-defined) audiovisual fusion.
- Key-frame detection based on geometric characteristics of the audiovisual curve (edges, peaks, lobes, levels).

The objective is for the system to be able to read, split and process any media file and for the user to easily modify, test and tune the basic system parameters (processing parameters, fusion schemes, fusion weights) externally.

Input and Output

Videos are processed from a data folder (*datadir*) given their name with the extension (e.g. **myvideo.mpeg**) and desired segment duration in frames. For memory-management reasons, the streams are processed in segments of user-defined length and the intermediate result structures are stored and loaded from the hard-drive. For that, a file/directory structure is created in a user-defined results folder (*resdir*). Intermediate results are stored in the *resdir/myvideo/audio* and *resdir/myvideo/visual* folders, while final results are stored in .mat and .txt files in the *resdir/myvideo* folder. For example, the processing of 512 frames of the input 'myvideo.mpeg' results in the following output mat files:

myvideo_saliency_frames_1_512.mat

```
saliency =
    framerate: [1 512]
    audiosalcurve: [512x1 double]
    audiosalcurveN: [512x1 double]
    audioParams: [1x1 struct]
    visualscurve: [512x1 double]
```

myvideo_avsaliency_frames_1_512.mat

```
AVsal =
    framerate: [1 512]
    methodA2A: 'linear'
```

```

weightsA2A: [0.3333 0.3333 0.3333]
weightsV2V: [0.2000 0.4000 0.4000]
methodA2V: 'linear'
weightsA2V: [0.5000 0.5000]
salcurve: [512x1 double]

```

Detected key-frames are stored as variables in .mat files and .tiff images in the *resdir/myvideo/key* directory, e.g. the first key-frame that corresponds to frame 12 is the image **myvideo_keyframe_1_frame_12.tiff**.

Software Package Organization

The source code is organized in a file structure hierarchy which separates the distinct system modules.

Example Usage

- Set root, data and results directories by modifying the m-files in the *dirs* folder.
- Define system parameters from *defineParams*, *asalParams*, *vsalParams* files (default values are given).
- In the *demo_AVsal* script, give the name of the input video and desirable duration in frames.

System Requirements and Installation

System Requirements

The software has been developed and tested in Matlab (Version 7.0 and larger) and Windows XP operating system. It requires Matlab's Image Processing Toolbox and the following functions /mex files/codecs:

1. [mmread.m](#) by Micah Richert, distributed on the Mathworks file exchange site. This contains binaries for Windows as well as source code. This serves as the interface that reads the audio and visual streams of any input media file.
2. Depending on the installed audio/video codecs, a codec package may be required like for example [ffdshow](#).

Software Installation

Unzip and add folder in the Matlab path directory. For this from the Matlab command window go to 'File/Set Path/Add with Subfolders' and choose the folder 'AVsal'. Set the directories *rootdir* (necessary! the root of the software package) and *datadir*, *resdir* (optional! define where you keep your data and where you want to store your results) from the m-files in the *dirs* folder.

Contact Person(s) / Download Details

Developers/authors: G. Evangelopoulos and K. Rapantzikos

Contact emails : gevag@cs.ntua.gr, krap@image.ntua.gr

Software webpage: <http://cvsp.cs.ntua.gr/software/AVsal/>

Related Publications

This software is related to the following publications

1. G. Evangelopoulos, K. Rapantzikos, P. Maragos, Y. Avrithis and A. Potamianos, “Audio-visual attention modeling and salient event detection”, in *Multimodal Processing and Interaction: Audio, Video, Text*, edited by P. Maragos, A. Potamianos and P. Gros, Springer, 2008.
2. K. Rapantzikos, G. Evangelopoulos, P. Maragos and Y. Avrithis., “An audiovisual saliency model for movie summarization”, *IEEE Intl’ Workshop on Multimedia Signal Processing (MMSP-07)*, Chania, Greece, 2007.
3. G. Evangelopoulos, K. Rapantzikos, A. Potamianos, P. Maragos, A. Zlatintsi, Y. Avrithis Movie summarization based on audio-visual saliency detection, *IEEE Intl’ Conf. on Image Proc. (ICIP-08)*, San Diego, 2008.

Licensing Conditions:

If you use this software or part of it in your research, please cite articles [1] and [3] as the most up-to-date references to the system’s functionality.