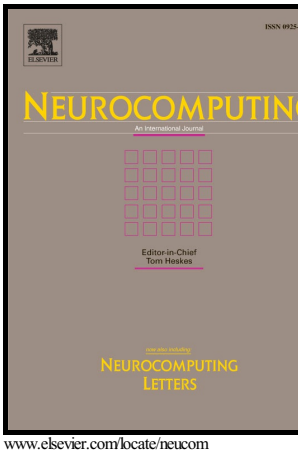
Summarizing video sequence using a graph-based hierarchical approach

Luciana do Santos Belo, Carlos Antônio Caetano Jr., Zenilton Kleber Gonçalves do Patrocínio Jr., Silvio Jamil Ferzoli Guimarães



www.elsevier.com/locate/neucom

PII: S0925-2312(15)01229-1
DOI: http://dx.doi.org/10.1016/j.neucom.2015.08.057
Reference: NEUCOM15989

To appear in: *Neurocomputing*

Received date: 24 April 2015
Revised date: 5 August 2015
Accepted date: 16 August 2015

Cite this article as: Luciana do Santos Belo, Carlos Antônio Caetano Jr., Zenilton Kleber Gonçalves do Patrocínio Jr. and Silvio Jamil Ferzoli Guimarães Summarizing video sequence using a graph-based hierarchical approach *Neurocomputing,* http://dx.doi.org/10.1016/j.neucom.2015.08.057

Summarizing video sequence using a graph-based hierarchical approach

Luciana do Santos Belo, Carlos Antônio Caetano Jr., Zenilton Kleber Gonçalves do Patrocínio Jr., Silvio Jamil Ferzoli Guimarães*

Audio-visual Information Processing Laboratory
Computer Science Department
Pontifícia Universidade Católica de Minas Gerais

Abstract

Video summarization is a simplification of video content for compacting the video information. The video summarization problem can be transformed into a clustering problem, in which some frames are selected to saliently represent the video content. In this work, we use a graph-based hierarchical clustering method for computing a video summary. In fact, the proposed approach, called HSUMM, adopts a hierarchical clustering method to generate a weight map from the frame similarity graph in which the clusters (or connected components of the graph) can easily be inferred. Moreover, the use of this strategy allows the application of a similarity measure between clusters during graph partition, instead of considering only the similarity between isolated frames. We also provide an unified framework for video summarization based on minimum spanning tree and weight maps in which HSUMM could be seen as an instance that uses a minimum spanning tree of frames and a weight map based on hierarchical observation scales computed over that tree. Furthermore, a new evaluation measure that assesses the diversity of opinions among users when they produce a summary for the same video, called *Covering*, is also proposed. During tests, different strategies for the identification of summary size and for the selection of keyframes were analyzed. Experimental results provide quantitative and qualitative comparison between the new approach and other popular algorithms from the literature, showing that the new algorithm is robust. Concerning quality measures, HSUMM outperforms the compared methods regardless of the visual feature used in terms of *F-measure*.

Keywords: Graph-based hierarchical video summarization, Covering, Global descriptors, Observation

*Corresponding author

Email addresses: lucyannasbelo@gmail.com (Luciana do Santos Belo), carlos.a.caetano.jr@gmail.com (Carlos Antônio Caetano Jr.), zenilton@pucminas.br (Zenilton Kleber Gonçalves do Patrocínio Jr.), sjamil@pucminas.br (Silvio Jamil Ferzoli Guimarães)