Accepted Manuscript

# Multimodal Data Fusion for Sensitive Scene Localization

Daniel Moreira, Sandra Avila, Mauricio Perez, Daniel Moraes, Vanessa Testoni, Eduardo Valle, Siome Goldenstein, Anderson Rocha

PII: S1566-2535(17)30286-5
DOI: 10.1016/j.inffus.2018.03.001
Reference: INFFUS 964

To appear in: *Information Fusion*

Received date: 1 May 2017
Revised date: 28 February 2018
Accepted date: 5 March 2018

Please cite this article as: Daniel Moreira, Sandra Avila, Mauricio Perez, Daniel Moraes, Vanessa Testoni, Eduardo Valle, Siome Goldenstein, Anderson Rocha, Multimodal Data Fusion for Sensitive Scene Localization, *Information Fusion* (2018), doi: 10.1016/j.inffus.2018.03.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

# Multimodal Data Fusion for Sensitive Scene Localization

Daniel Moreira[a], Sandra Avila[a,b], Mauricio Perez[a], Daniel Moraes[a], Vanessa Testoni[c], Eduardo Valle[b], Siome Goldenstein[a], Anderson Rocha[a,*]

[a] *Institute of Computing, University of Campinas, Brazil*
[b] *School of Electrical and Computing Engineering, University of Campinas, Brazil*
[c] *Samsung Research Institute Brazil, Brazil*

**Abstract**

The very idea of hiring humans to avoid the indiscriminate spread of inappropriate sensitive content online (e.g., child pornography and violence) is daunting. The inherent data deluge and the tediousness of the task call for more adequate approaches, and set the stage for computer-aided methods. If running in the background, such methods could readily cut the stream flow at the very moment of inadequate content exhibition, being invaluable for protecting unwary spectators. Except for the particular case of violence detection, related work to sensitive video analysis has mostly focused on deciding whether or not a given stream is sensitive, leaving the localization task largely untapped. Identifying when a stream starts and ceases to display inappropriate content is key for live streams and video on demand. In this work, we propose a novel multimodal fusion approach to sensitive scene localization. The solution can be applied to diverse types of sensitive content, without the need for step modifications (general purpose). We leverage the multimodality data nature of videos (e.g., still frames, video space-time, audio stream, etc.) to effectively single out frames of interest. To validate the solution, we perform localization experiments on pornographic and violent video streams, two of the commonest types of sensitive content, and report quantitative and qualitative results. The results show, for instance, that the proposed method only misses about five minutes in every hour of streamed pornographic content. Finally, for the particular task of pornography localization, we also introduce the first frame-level annotated pornographic video dataset to date, which comprises 140 hours of video, freely available for downloading.

*Keywords:* multimodal data fusion, sensitive scene localization, pornography localization, violence localization

## 1. Introduction

We define a *sensitive scene* as a motion picture excerpt whose content may inflict harm (e.g., trauma, shock, or fear) to particular audiences (e.g., children or unwary spectators), due to the

*Corresponding author
*Email address:* anderson.rocha@ic.unicamp.br (Anderson Rocha)

Download English Version:

https://daneshyari.com/en/article/6937915

Download Persian Version:

https://daneshyari.com/article/6937915

Daneshyari.com