A New Deep Wavefront Based Model for Text Localization in 3D Video

Publisher: IEEE

Cite This PDF

Lokesh Nandanwar ; Palaiahnakote Shivakumara ; Raghavendra Ramachandra ; … All Authors

Abstract: With the evolution of electronic devices, such as 3D cameras, addressing the challenges of text localization in 3D video (e.g., for indexing) is increasingly drawing the ... **View more**

Metadata

Abstract:

With the evolution of electronic devices, such as 3D cameras, addressing the challenges of text localization in 3D video (e.g., for indexing) is increasingly drawing the attention of the multimedia and video processing community. Existing methods focus on 2D video and their performance in the presence of the challenges in 3D video, such as shadow areas associated with text and irregularly sized and shaped text, degrades. This paper proposes the first approach that successfully addresses the challenges of 3D video in addition to those of 2D. It employs a number of innovations, among which, the first is the Generalized Gradient Vector Flow (GGVF) for dominant points detection. The second is the Wavefront concept for text candidate point detection from those dominant points. In addition, an Adaptive B-Spline Polygon Curve Network (ABS-Net) is proposed for accurate text localization in 3D videos by constructing tight fitting bounding polygons using text candidate points. Extensive experiments on custom (3D video) and standard datasets (2D video

and scene text) show that the proposed method is practical and useful, and overall outperforms existing state-of-the-art methods.

Published in: IEEE Transactions on Circuits and Systems for Video Technology (Volume: 32 , Issue: 6, June 2022)

Page(s): 3375 - 3389

Date of Publication: 07 September 2021

INSPEC Accession Number: 21762835

DOI: 10.1109/TCSVT.2021.3110990

Publisher: IEEE

ISSN Information:

Funding Agency:

Contents

Text localization in 3D video is an important topic for content-based video retrieval, particularly for annotating video based on semantics [1], [2]. It has attracted considerable research attention due to explosive growth of multimedia content which includes 2D and 3D video data, available [1], [2]. As a result, there is an increasing number of large repositories containing 2D, 3D video/images and multimedia content [1]. To ensure the robustness and accuracy of retrieval systems, text localization is vital as it provides significant semantic information for annotating video [3], [4]. Existing models focus on text localization in 2D video but not 3D video [3], [4]. Therefore, there is a need for a model that can work for both 2D and 3D video. Example of retrieval cases can include events extraction from 3D sports video, choosing a particular scene in a 3D movie, tracking and watching person behavior and interaction captured by 3D camera during exhibitions, processions, celebrations, etc. These situations motivated the authors to introduce the problem of text localization in 3D video in this work.