Abstract

We present an implementation of a search engine that searches videos based on its textual content. The system consists of four parts video processing, spell correction, indexing and searching. The video processing is done by dividing the video into frames and extracting text out of it. Lecture videos, news having some textual content in it show good results.

References

- Huizhong Chen, Sam S. Tsai, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, "Robust Text Detection In Natural Images With Edge-Enhanced Maximally
- B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with
- Luís Gomes, Dimosthenis Karatzas, "MSER-based Real-Time Text Detection and
- Marc Davis, "Media Streams: Representing Video for Retrieval and Repurposing.
- K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and
- C. Merino and M. Mirmehdi, "A framework towards real-time detection and
- M. Donoser and H. Bischof, "Efficient maximally stable extremal region (mscr)
- R. Minetto, N. Thome, M. Cord, N. J. Leite, and J. Stolfi, "Text detection and
- Liu, T. Choudhary, "Content Extraction and Summarization of Instructional
  Videos," in IEEE, 2006
- Haojin Yang, "Lecture Video Indexing and Analysis Using Video OCR
  Technology," in IEEE, 2011
- Zi Huang, Yijun Li, Jie Shao, Shen1 Heng, Tao, Wang1 Liping, Wang1, Danqing, Zhang3
  Xiangmin, Zhou1, Xiaofang, Zhou1, "Content-Based Video Search: is there a need, and is
  it possible," in IEEE, 2008
- Julien Law-To, Rémi Landais, Gregory Grefenstette, "VOXALEADNEWS: A
  Scalable Content Based Video Search Engine," in IEEE, 2012
- P. Geetha Vasumathi Narayanan, "An Effective Video Search Re-Ranking for
  Content Based Video Retrieval," in IEEE, 2011

Index Terms

Computer Science

Pattern Recognition
Keywords

Indexing  Retrieval  Video  OCR  Searching  Search Engine  Apache Solr.