Celebrity Face-Name Association in Web Videos using Unsupervised Approach

Shweta Tadge[1], Prof. Ranjana Dahake[2]

[1] PG Student, Department of Computer Engineering, MET BKC, Nashik.
[2] Professor, Department of Computer Engineering, MET BKC, Nashik.

ABSTRACT

This paper explores the problem of missing name and missing faces in unconstrained videos with user provided metadata. Rather than depending upon supervised learning, a better relationship built from the content of a video, those relationship includes the arrival of faces in different spatio-temporal contexts and visual similarities between faces. The knowledge base consists of tagged images along with a set of names and celebrity social networks. Celebrity social network is built based on the co-occurrence statistics of celebrities in video metadata. Merging of relationship along with knowledge base is carried out via conditional random field. Two types of face-name association are investigated: within video face labeling and between video face labeling. The within video labeling takes care of noisy as well as incomplete labels in metadata, in which null assignment for the labels is permitted. Furthermore Between video face labeling addresses the flaws within metadata, particularly to correct incorrect names and label faces having no available names in metadata of a video. To do so it considers a gathering of socially associated videos for combined name inference. The experimental result analysis on web video dataset shows that proposed approach is very much powerful for handling the issue of missing names and incorrect names in face labeling problem than existing approaches.

Keyword: - *Celebrity face labeling, social networks, unsupervised learning, unconstrained videos.*

---

1. Introduction

Individuals do upload large number of videos, in which 80\% are related to people. In those videos 75\% are related to celebrities. The tremendous growth of video on the internet and the rising insufficiency in metadata associated with video forces us to look at the content from the video content for search retrieval and browsing based exposure. A large chunk of users' browsing patterns are centered around people present in the video. With the massive growth of digital videos in the Internet, recognizing and understanding the visual content is becoming an increasingly important problem. These days identification of characters in web videos is a challenging task because of huge deviation in the approach of person or celebrities within web videos. Naming celebrities in the user-generated videos provides great help for both indexing and browsing. In all the top video search engines like YouTube, indexing of these videos is based on user-provided text data like title of videos or description, which found to be noisy and incomplete most of the times. Often a mentioned celebrity may not exist within video, and a celebrity which actually exist within a video is not mentioned in user-provided text. There is no correlation present between video content and metadata associated with video due to these issues, people-related video search results into non satisfactory retrieval of videos. Identifying the direct relation between faces and names can help in rectifying the potential errors in user provided metadata hence it is provided as an initial processing step for indexing of videos. Rich context information cannot be applied directly for face naming in unrestricted videos due to lack of prior knowledge and the context cues.

Fig -1 explores problem using a web video example. In this video nineteen faces are detected in which only four having their names mentioned in metadata. Also, among the five mentioned celebrities in metadata, only four of them are actually present in the video. It concludes that faces are missing within videos and names are missing in text provided with video. Furthermore, faces can appear wildly different because of motion blurriness, resolution and lighting changes. In other words, the problem of face name association can be specific to inadequate metadata, false metadata and visual appearance suggestion. That means it consider the reality that the text provided with video