

# Unsupervised Video Tag Correction System

Hoang-Tung Tran\*, Elisa Fromont\*, Francois Jacquenet\*,  
Baptiste Jeudy\*, Adrien Martins\*

\*Laboratoire Hubert Curien, UMR CNRS 5516  
18 Rue du Professeur Benoît Luras, 42000 Saint-Etienne  
{hoang.tung.tran, elisa.fromont, francois.jacquenet}@univ-st-etienne.fr  
{baptiste.jeudy, adrien.martins}@univ-st-etienne.fr

**Abstract.** We present a new system for video auto tagging which aims at correcting and completing the tags provided by users for videos uploaded on the Internet. Unlike most existing systems, we do not learn any tag classifiers or use the questionable textual information to compare our videos. We propose to compare directly the visual content of the videos described by different sets of features such as Bag-of-visual-Words or frequent patterns built from them. Then, we propagate tags between visually similar videos according to the frequency of these tags in a given video neighborhood. We also propose a controlled experimental set up to evaluate such a system. Experiments show that with suitable features, we are able to correct a reasonable amount of tags in Web videos.

## 1 Introduction

Classic text-based search engines already offer a good access to multimedia contents in the online world. However, they cannot index the extensive number of online videos unless these videos are carefully annotated before being put on the Web. However, user-provided annotations are often incorrect, i.e. irrelevant to the video (e.g. to increase the video's number of views), and incomplete. To overcome these drawbacks, we will focus on the task of setting up an automatic system to improve annotations of web videos. There have already been many efforts to automatically annotate videos (e.g (Morsillo et al., 2010), (Shen et al., 2011)). However, most of the proposed systems use limited concepts (tags) and some supervised information to learn one or many classifiers to tag a video dataset. These approaches thus seem inappropriate for any video on a large website such as Youtube where the number of possible tags is unlimited and where the true labels are inaccessible a priori. We thus would like to propose an unsupervised approach based on the comparison of the visual content of the videos to propagate the tags from the neighbor videos based on their textual frequency. In this approach the main scientific locks reside i) in the choice of the features that will be used to make relevant unsupervised comparisons, ii) in the comparison method itself, iii) in the propagation process and iv) in the evaluation of the entire system. A review of related works concerning the above mentioned problems is briefly given in Section 2. In Section 3, we describe in details how to apply data mining techniques as well as our proposed method to compare videos. The experiments done so far are presented in Section 4 and we conclude in Section 5.