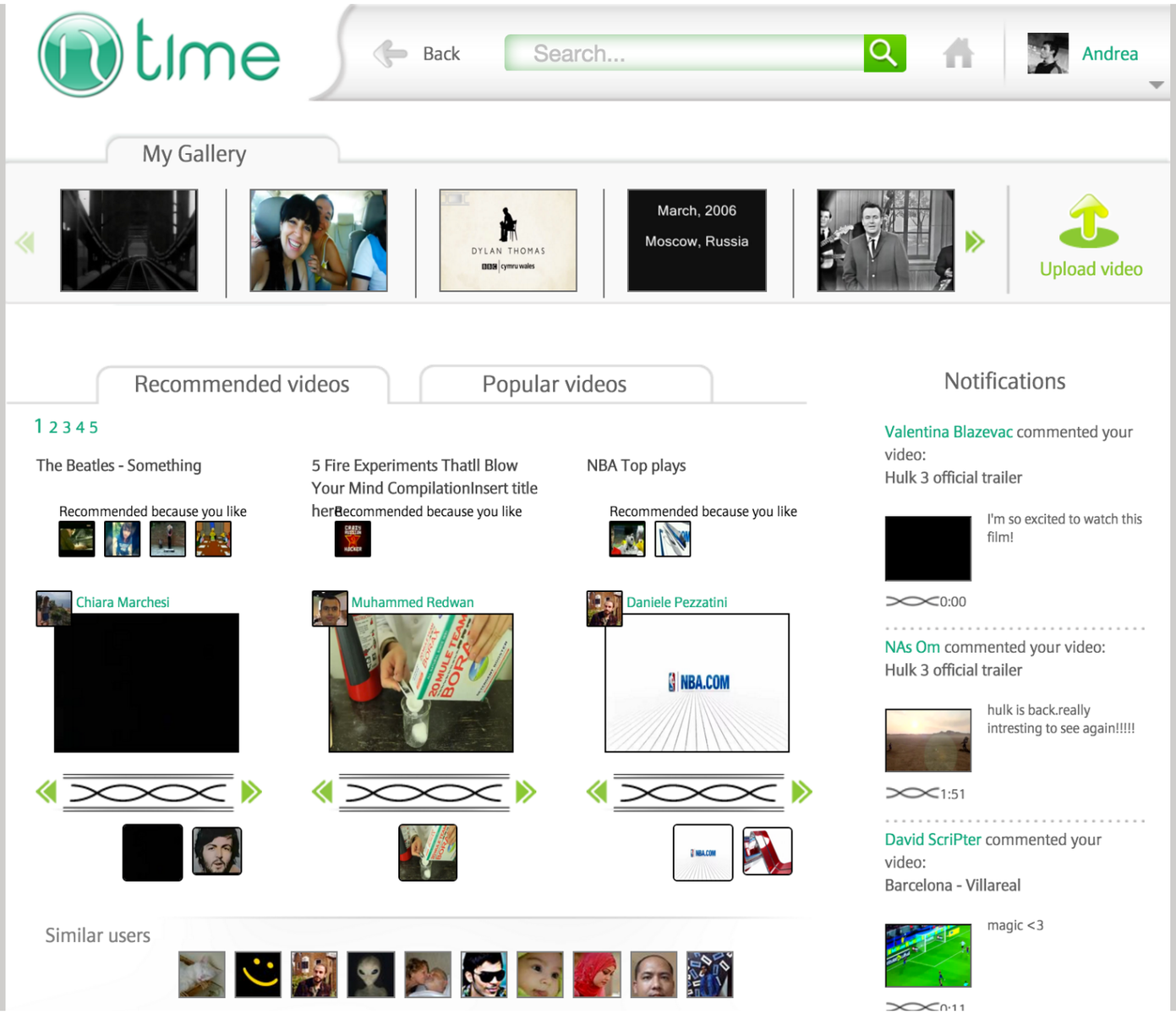
ACM MULTIMEDIA 2015
26 - 30 October 2015, Brisbane, Australia

# A System for Video Recommendation using Visual Saliency,Crowdsourced and Automatic Annotations

A. Ferracani, D. Pezzatini, Marco Bertini, Saverio Meucci, Alberto Del Bimbo

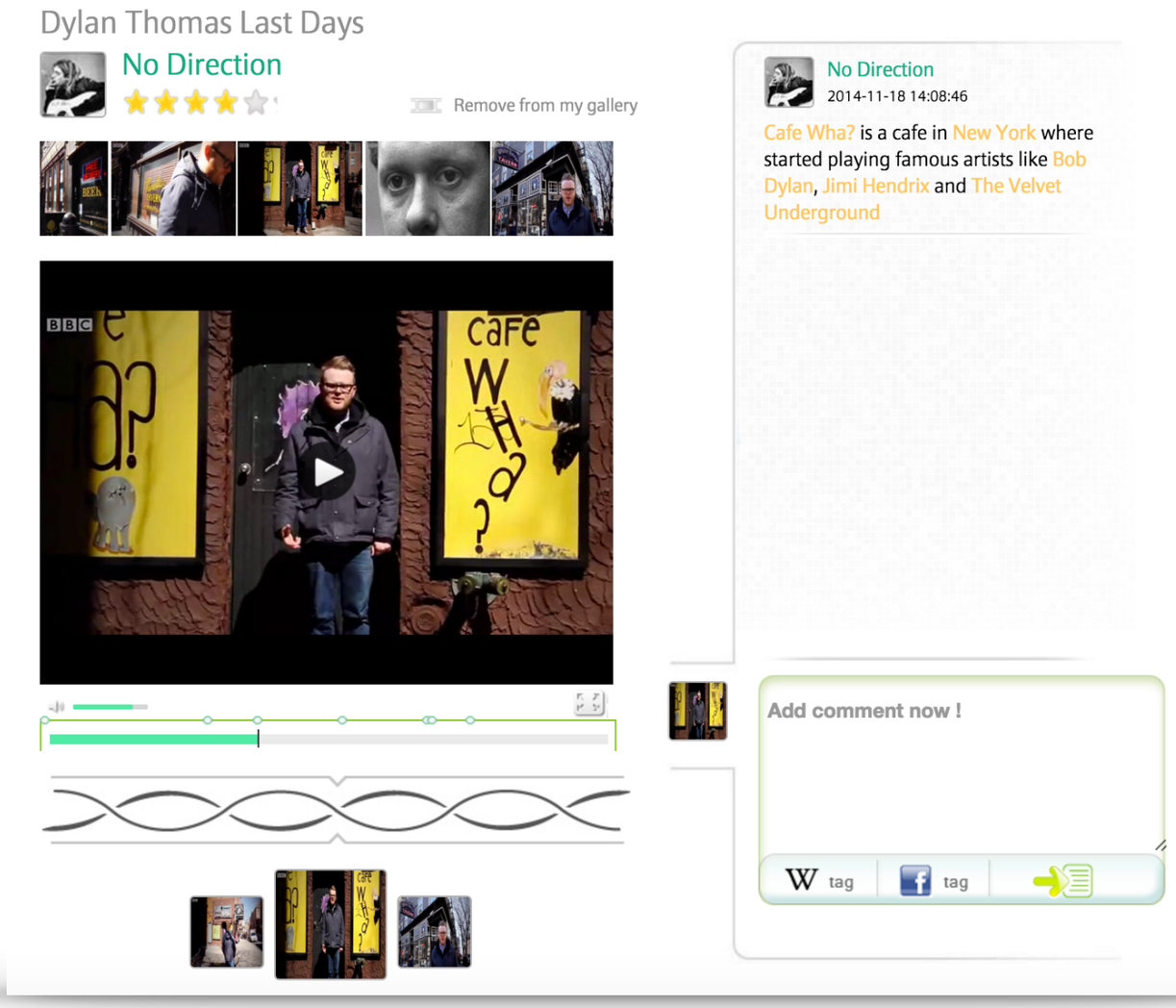

http://www.micc.unifi.it

We propose a system **to improve content-based video recommendation** that exploits **visual saliency** to better represent video features and content.



Visual saliency is used

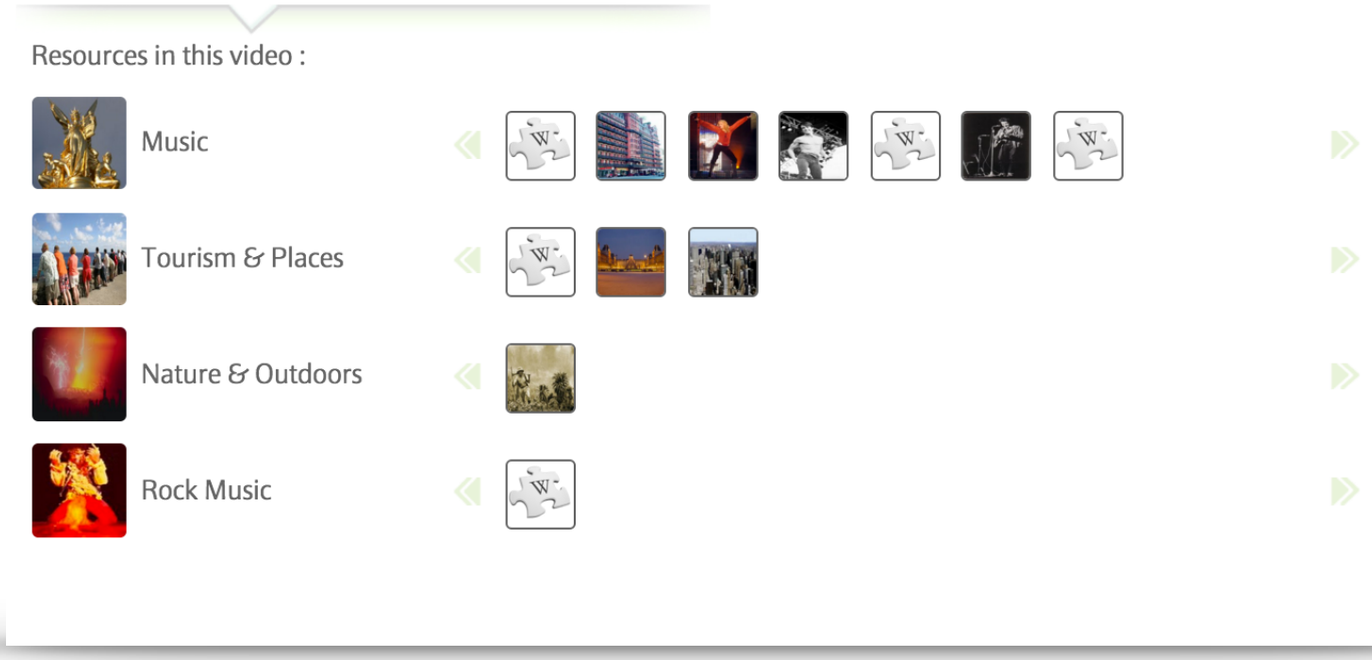
- to **choose relevant frames** to be presented in a web-based interface to tag and annotate video frames in a social network;

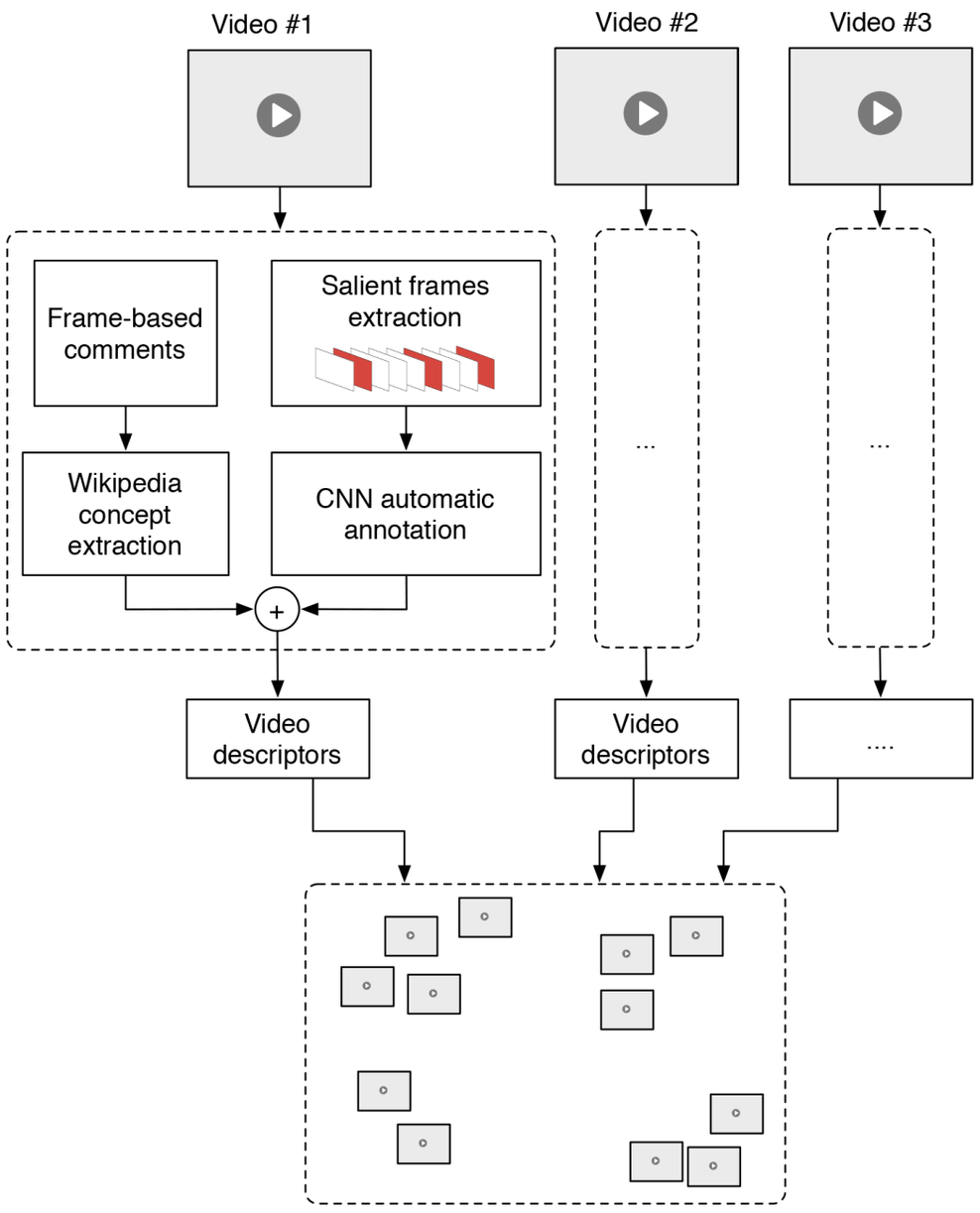


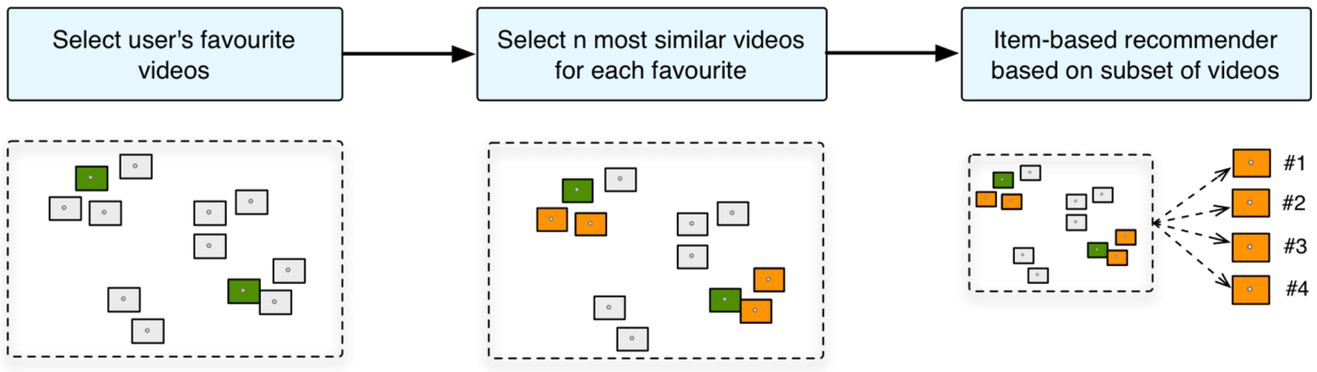

- to **summarize video content** to create a more effective video representation used in the recommender system.

The system exploits **automatic annotations** from CNN-based classifiers **on salient frames** and user generated annotations.

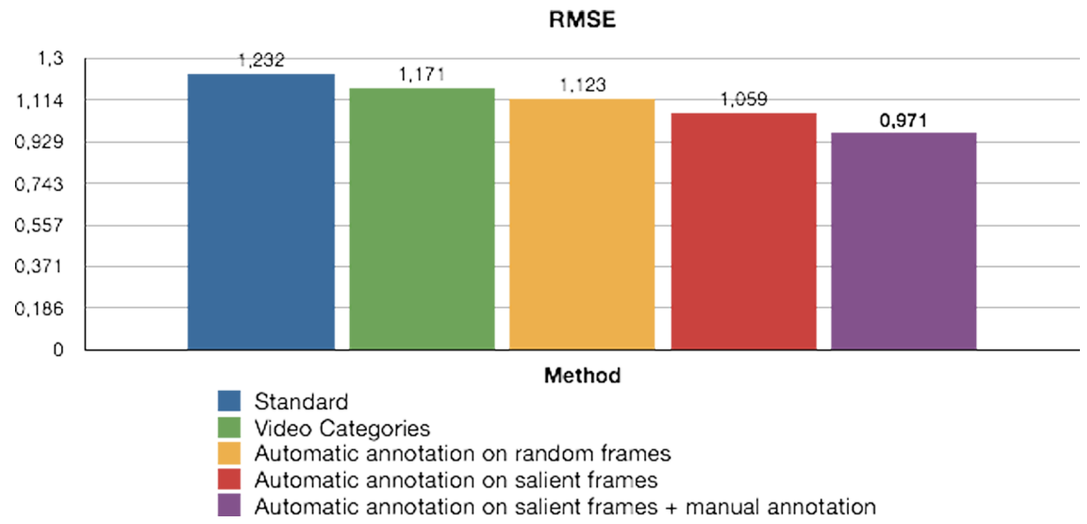
Users can share and annotate videos at frame level using **concepts derived from Wikipedia**. All these concepts are clustered in 54 categories using Fuzzy K-Means in a two-levels taxonomy of interests and classified using a semantic distance with a nearest neighbour approach.




Video #1
Video #2
Video #3
Frame-based comments
Salient frames extraction
Wikipedia concept extraction
CNN automatic annotation
+
...
...
Video descriptors
Video descriptors
....


Select user's favourite videos
Select n most similar videos for each favourite
Item-based recommender based on subset of videos
#1
#2
#3
#4

We evaluate the performance of the proposed recommender, in terms of **RMSE**, comparing it to several baselines:




1) **standard** item-based recommender

2) recommender working on a selection of items, based on similarity computed **using system categories** only (no BoW content description);

3) recommender working on a selection of items, based on content similarity (i.e. automatic annotations) computed on **randomly selected frames**;

4) recommender working on a selection of items, based on content similarity computed on n **frames with visual saliency score above the average**;

5) recommender working on a selection of items, based on content similarity computed on a) n frames with **visual saliency score above the average** and b) crowdsourced manual annotations

# More Info

MICC
http://www.micc.unifi.it

MICC Visual Information and Media Lab:
http://www.micc.unifi.it/vim/

MICC blog
http://www.micc.unifi.it/blog