Sentiment Analysis of Weakly Labeled Social Media First Person Vision streams

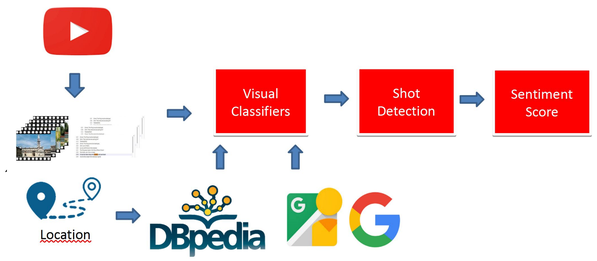
Patrizia Varini
Regione Emilia-Romagna
Università degli studi di Modena e Reggio Emilia

Goal

This work proposes a novel approach to draw dominant sentiment maps related to main sites of interest in Art Cities, analyzing egocentric or hand held captured streams extracted from Social Media repositories jointly with their eventual descriptions and tags and trascription of their associated audios.

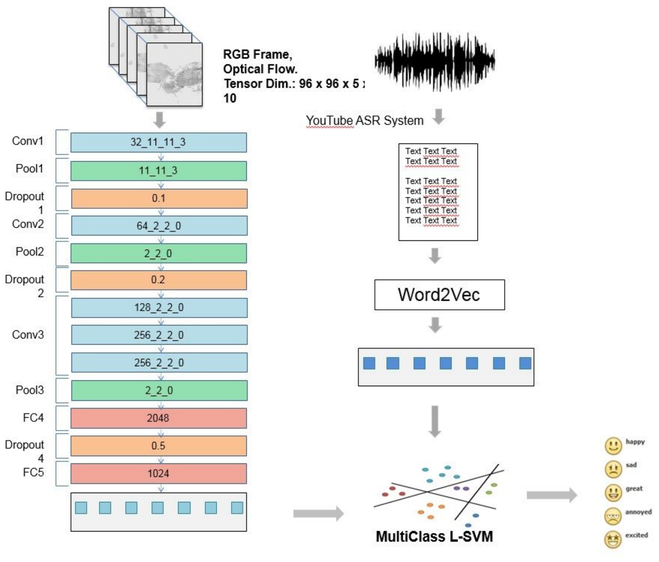
General approach

We first get from a specific location of interest, a list of preminent places of interest, such as monuments, churches, buildings, museums, galleries, famous store, and build specialized visual classifierss. Subsequently, we extract from YouTube videos captured in the specific location and their ASR, using a pre-trained CNN to filter only FPV streams, with ASR in english, detect Points of Interest and perform shot detection.



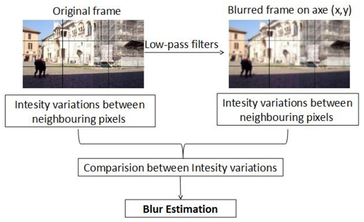
A dataset is built, starting from extracted and filtered FPV videos, and fully annotated according to the 5 classes of interest, corresponding to the sentiments we want to detect: “sad”, “annoyed”, “happy”, “great”, “excited”.

A Sentiment classifier is thus built, using a multiclass L-SVM approach on features jointly extracted respectively from streams using a 10 layers 3D CNN and from text using a Word2Vec based approach.
Below we show the sentiment classifier schema.



Visual and semantic Features

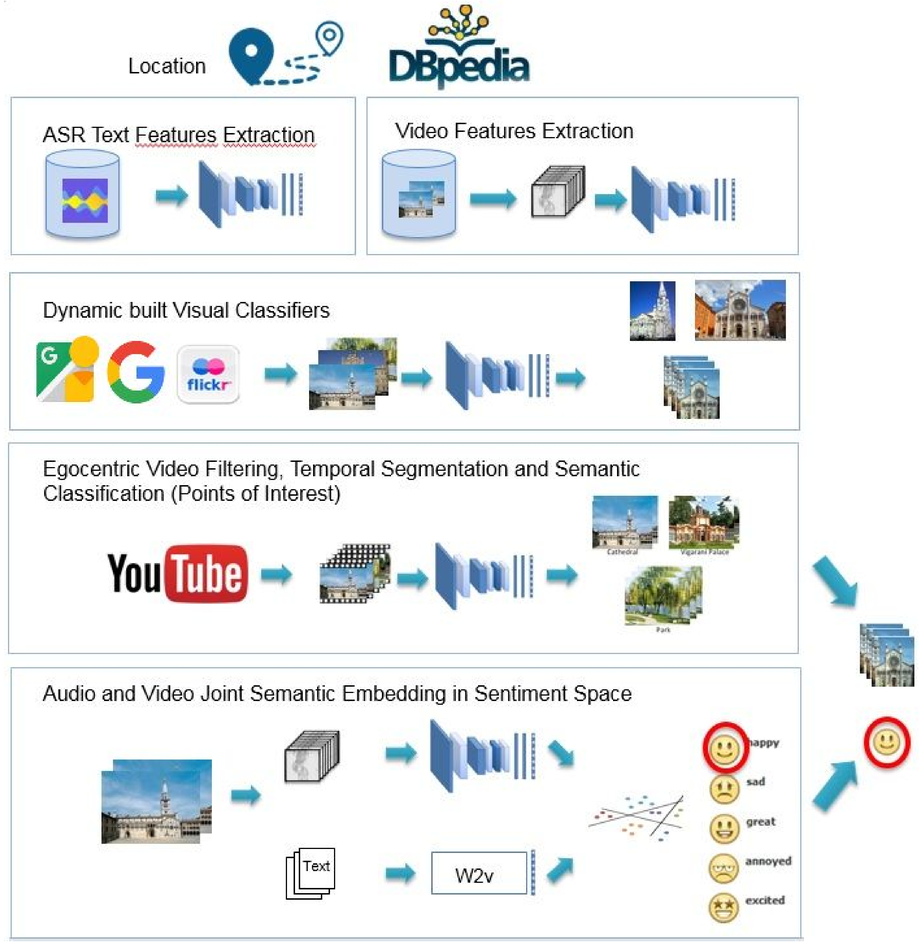
Visual Features: We assume that the pattern of visitors motion and the visual relevance of what he sees is a relevant information to classify a certain sentiment of the camera’s wearer. We use apparent motion (Optical Flow and Motion Boundary derivative) and Blur



Semantic Features: We train Word2Vec model on DB-Pedia english abstracts, setting the dimensionality of our feature vector to 300.

The final feature vector that feeds L-SVM has dimension of 1324.

Overall Schema of the Proposed Method



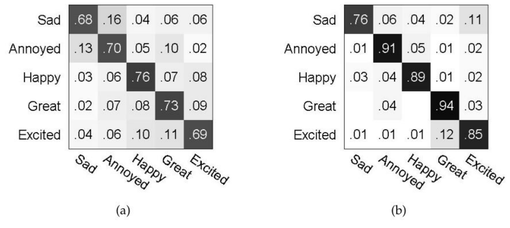
Sentiment maps

Aggregated map, showing sentiments associated to each of the most popular sites or items in **Florence**. To obtain the resultant map, which georeferentiates the results described so far, we leverage Google Maps API. For each item or place of interest, the emoticon related to the predominant sentiment is shown at its GPS coordinate location

Conversely, one can aggregate on sentiment dimension and obtains a graph which show items according to dominant sentiments detected

Experimental results

Dataset: 48 FP videos extracted by YouTube captured in four italian art cities, Florence, Naples, Venice, Rome. Below it is shown the classification accuracy using different descriptors: a) CM for Modern Art Gallery, Florence ; b) CM for Cathedral, Florence



Method	Accuracy
Visual Features only	61.40
Text Features Only	56.35
Visual and Text Feat.	71.67