



# UNIVERSITA' POLITECNICA DELLE MARCHE

DII – Dipartimento di Ingegneria dell'Informazione

## Automatic Annotation of Corpora For Emotion Recognition Through Facial Expressions Analysis

---

Claudia Diamantini, Alex Mircoli, Domenico Potena, Emanuele Storti

{c.diamantini, a.mircoli, d.potena, e.storti}@univpm.it

# Scenario

- Explosion of publicly available User-Generated Content (UGC)
  - Social networks (Facebook, Twitter)
  - Movie/product reviews (IMDB)
  - ...
- People express their personal feelings and opinions
- Opportunity: emotion analysis

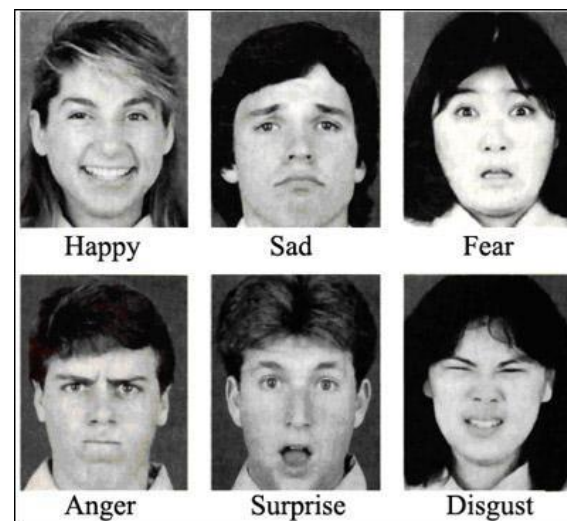


# Issues

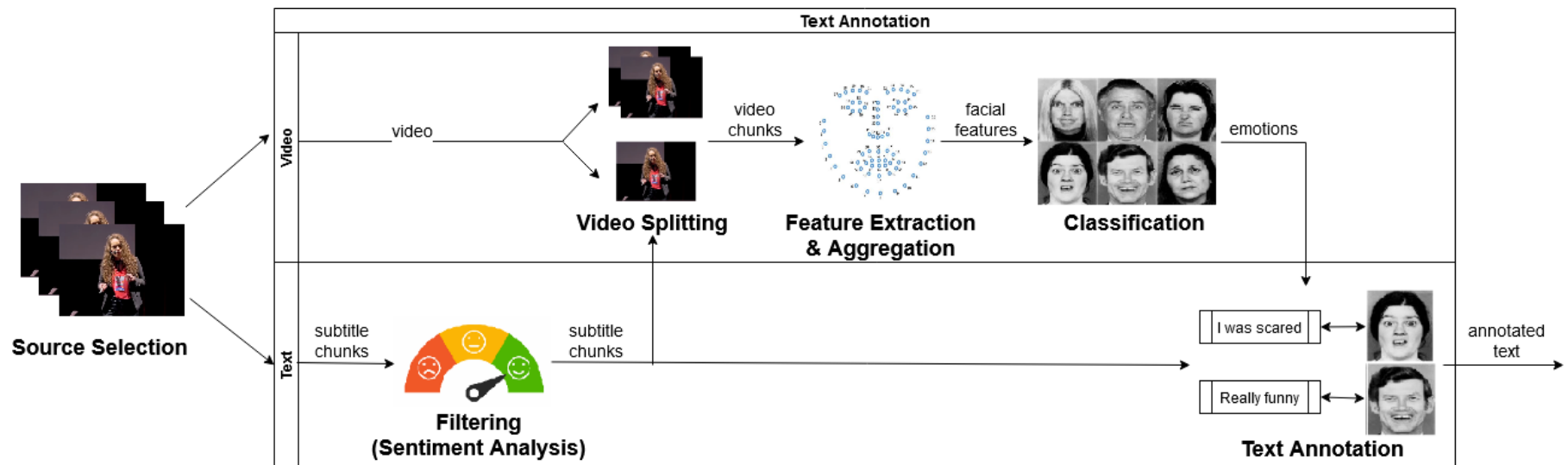
- Deep learning algorithms for emotion recognition need huge amounts of labelled data
- Data need to be manually labelled
  - expensive and time-consuming task
- Data labelled by human annotators can be biased

# Idea

- Idea: **automatically** annotate sentences through **facial expression analysis** in subtitled videos
- Facial expressions are coherent with speech [Livingstone, 2015]
  - Condition: spontaneous speakers
- Ekman's Theory [Ekman, 1969]
  - Facial expressions are **universal**



# Methodology



# Results

- Correlation between opinions and emotions:
  - SemEval 2007 Task #14: 1000 headlines
  - Text annotated w.r.t polarity and emotion
  - Correlation between non-neutral polarity and emotion: 0.92

	Non-neutral polarity	Neutral polarity
Non-neutral emotion	753	136
Neutral emotion	67	21

# Results

- Facial expression recognition (FER)
  - 200 manually labelled YouTube video chunks
  - 4-class emotional annotation (anger, happiness, sadness, neutral)
  - Several classifiers: Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Tree (DT), Random Forests (RF) and Multi-Layer Perceptron (MLP)
  - SVM: 72% accuracy

	Actual Neutral	Actual Happiness	Actual Anger	Actual Sadness
Predicted Neutral	43	8	10	14
Predicted Happiness	5	41	8	5
Predicted Anger	0	0	30	2
Predicted Sadness	2	1	2	29
Recall	0.86	0.82	0.60	0.58
Precision	0.57	0.69	0.94	0.85

# Results

## ■ Text annotation

- Same dataset as FER task
- Accuracy: 64.5%
- Some emotion-bearing texts are classified as neutral
- Some misclassifications are actually wrong human annotations

	Actual Neutral	Actual Happiness	Actual Anger	Actual Sadness
Predicted Neutral	40	9	13	14
Predicted Happiness	7	38	3	9
Predicted Anger	1	2	26	2
Predicted Sadness	2	1	8	25
Recall	0.80	0.76	0.52	0.50
Precision	0.53	0.67	0.84	0.69



# Conclusion & Future Directions

- Methodology for the automatic annotation of corpora
  - analysis of facial expressions in subtitled videos.
- Experiments on a dataset of YouTube videos
  - 4-class emotion recognition
  - Accuracy of facial expression recognition: 72%.
  - Accuracy of text annotation: 64.5%
- Results could be improved by adopting a multi-modal approach
  - speech-based emotion features (e.g., tone of voice).
- Word-level annotation
  - assign frames to each word through speech recognition techniques