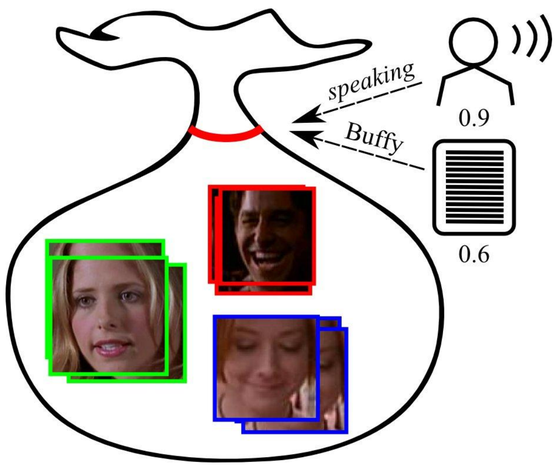
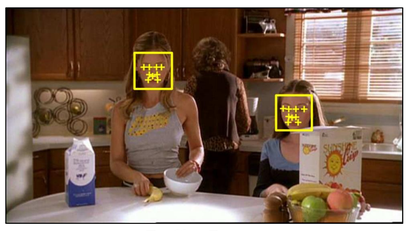
LEARNING FACE RECOGNITION IN VIDEOS FROM ASSOCIATED INFORMATION SOURCES*

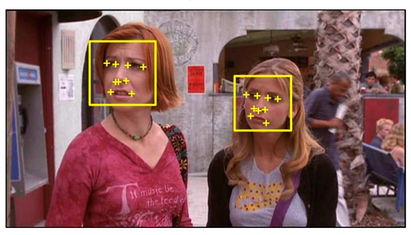
Martin Köstinger
koestinger@icg.tugraz.at
Institute for Computer Graphics and Vision, Graz University of Technology

Abstract Videos are often associated with additional information that could be valuable for interpretation of its content. This especially applies for the recognition of faces within video streams, where often cues such as transcripts and subtitles are available. However, this data is not completely reliable and might be ambiguously labeled. To overcome these limitations, we propose a new semi supervised multiple instance learning algorithm, where the contribution is twofold. First, we can transfer information on labeled bags of instances, thus, enabling us to weaken the prerequisite of knowing the label for each instance. Second, we can integrate unlabeled data, given only probabilistic information in form of priors.

Motivation




Buffy, Dawn


Willow, Buffy

Face recognition in videos: Often valuable information, i.e. extracted from transcript or subtitles, cannot be unambiguously assigned to exactly one person. Further not all information is completely reliable.

Main Contributions

- **Semi-Supervised Multiple Instance Learning (SSMIL)**
  - Fuses Semi-Supervised Learning (SSL) with Multiple Instance Learning (MIL)
  - Makes use of information cues that are **unreliable** and/or **ambiguous**
- **Application to Face Recognition in Videos**
  - Fully autonomous
  - Demonstrated on a publicly available benchmark dataset

Semi-Supervised Multiple Instance Learning

Idea: Joint probabilistic loss function that combines **SSL** and **MIL**

$$\mathcal{L}_l(\mathcal{D}_l) = -\sum_{i=1}^{N_l}\sum_{z\in\mathcal{Y}}[z=y_i]\log(P(y=z|\mathcal{B}_i^l))$$

Labeld loss

$$\mathcal{L}_u(\mathcal{D}_u) = -\sum_{i=1}^{N_u}\sum_{z\in\mathcal{Y}}P_P(y=z|\mathcal{B}_i^u)\log(P(y=z|\mathcal{B}_i^u))$$

Loss over unlabeld bags as deviation of the model from the prior [3]

$$\mathcal{L}(\mathcal{D}_l\cup\mathcal{D}_u) = \mathcal{L}_l(\mathcal{D}_l) + \mathcal{L}_u(\mathcal{D}_u)$$

Sum of both losses

Optimization - Gradient Boosting [4]

- Allows to use **arbitrary (differentiable) loss function**
- **Weights** of the samples
  - derivative of the loss function with respect to the current strong classifier output

$$\frac{\partial\mathcal{L}(\mathcal{D}_l\cup\mathcal{D}_u)}{\partial F(\mathbf{x}_{ij})}$$

- **Weak classifier**
  - approximates inverse direction of the gradients
  - weight determined by line search

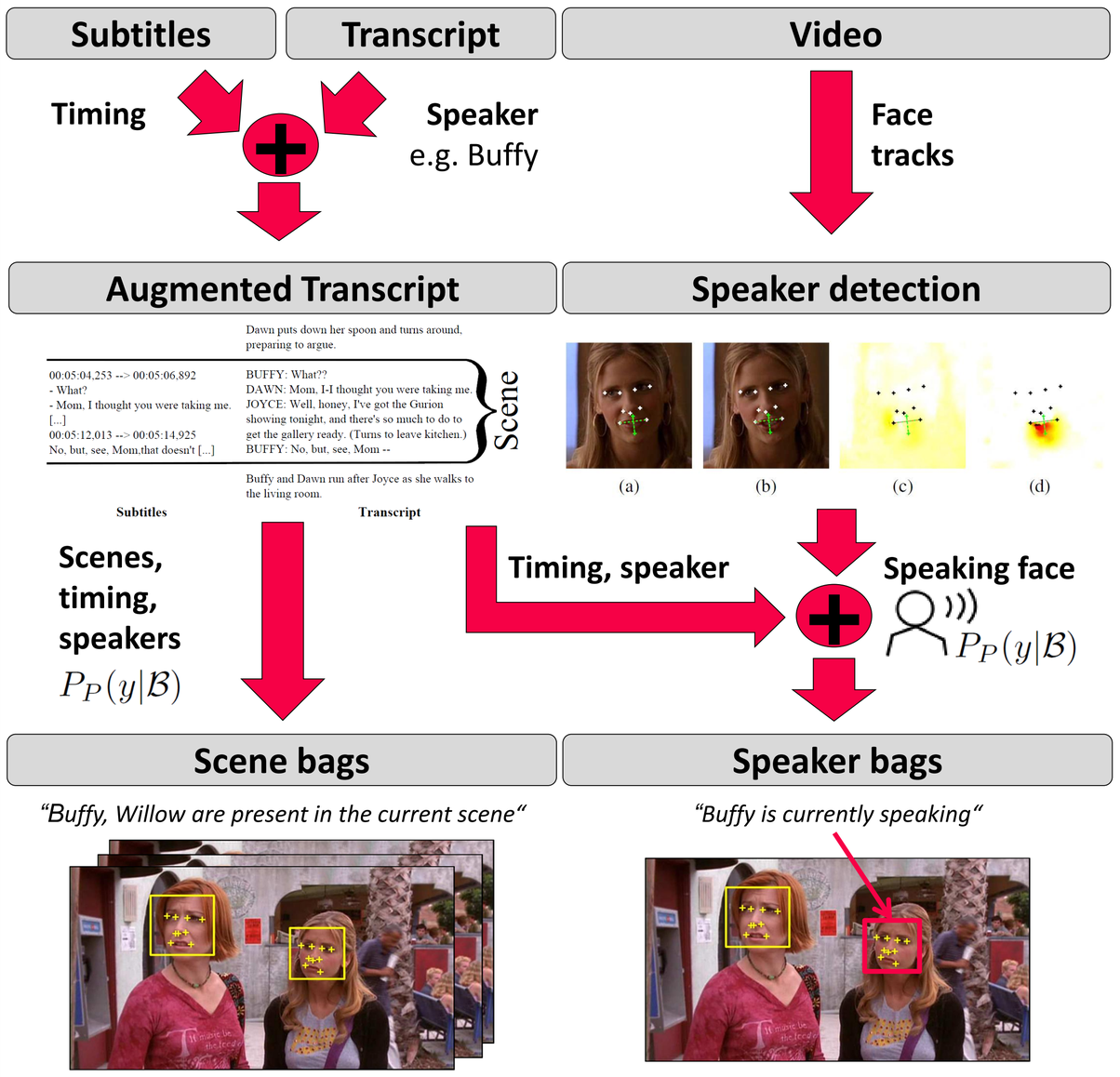
*The work was supported by the FFG projects MDL under the Austrian Security Research Program KIRAS

Face Recognition from Videos

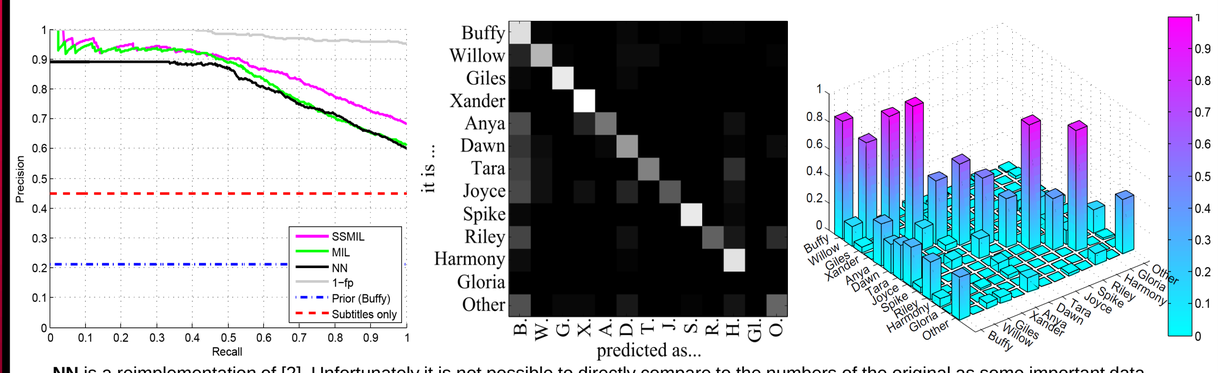
Evaluation on the „**Buffy the Vampire Slayer**„ dataset [2]

- 27504 face detections
- 11 named entities, “other” and false positive.
- **Task:** Label each of the 516 face tracks

How to automatically obtain training data?



Results



NN is a reimplementation of [2]. Unfortunately it is not possible to directly compare to the numbers of the original as some important data (speaker detection, clothing descriptors) is not provided

Conclusion

- **SSMIL** benefits from previously unusable cues (ambiguous and/or not completely reliable)
  - Single weak information source suffices to obtain a reasonable performance gain over existing work!
- Method is easily **extendable**
  - different bag types
  - information sources
  - appearance descriptors

[1] M. Köstinger, P. Wohlhart, P. Roth and H. Bischof, Learning to Recognize Faces from Videos and Weakly Related Information Cues. In Proc. AVSS, 2011.
[2] M. Everingham, J. Sivic, and A. Zisserman. “Hello! My name is... Buffy” – automatic naming of characters in TV video. In Proc. BMVC, 2006.
[3] A. Saffari, H. Grabner, and H. Bischof. SERBoost: Semisupervised boosting with expectation regularization. In Proc. ECCV. 2008.
[4] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. The Annals of Statistics, 28(2):337–374, 2000.