Improved Weak Labels using Contextual Cues for Person Identification in Videos

Makarand Tapaswi, Martin Bäuml, Rainer Stiefelhagen

Computer Vision for Human Computer Interaction, Karlsruhe Institute of Technology, Germany

Contributions

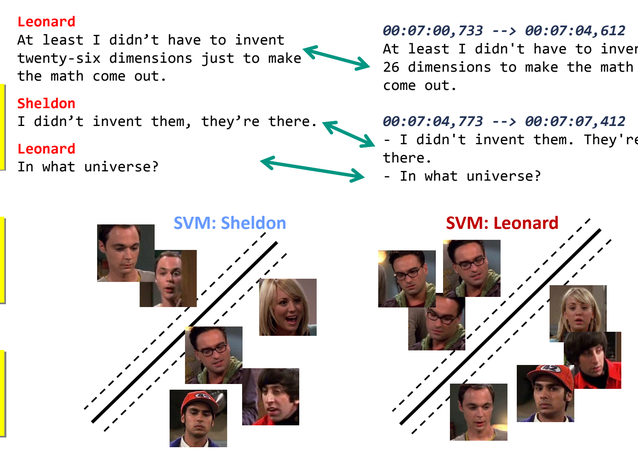
- Detailed error analysis of current weak label assignment methods for face tracks
- Revised weak label assignment approach through an *energy* function incorporating positive and negative constraints
- Evaluation on 2 TV series: number of erroneous weak labels halved, id performance improved

Person ID pipeline

(1) Obtain weak labels through subtitle-transcript alignment

(2) Train person-specific face models using weak labels

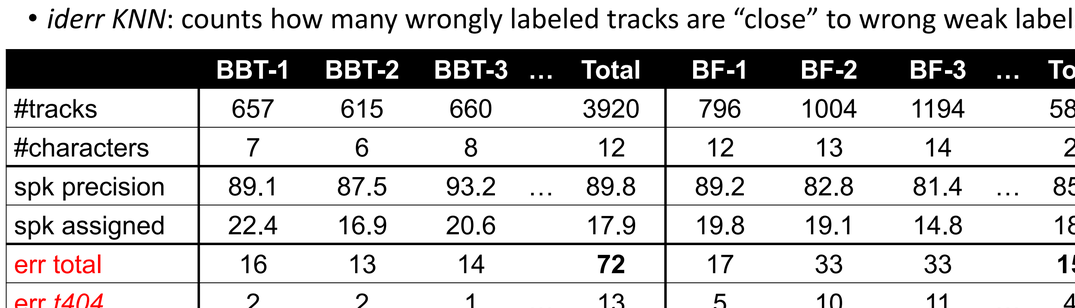
(3) Evaluate each model against all face tracks in the video



M. Everingham, J. Sivic and A. Zisserman. "Hello! My name is ... Buffy" Automatic Naming of Characters in TV Video. In BMVC 2006.

Problem with assignment method

- Assumes tracks are independent
- Introduces avoidable assignment errors
  - *t404* – Track not found
  - *hiscore* – Speaking score higher for wrong person



1 During subtitle S2, Penny does not have a track while Leonard (T3) appears to be speaking. This can result in an error of type ***t404***.

2 During subtitle S4, Sheldon's track (T5) has a higher speech activity than Leonard's (T6). This can result in erroneous assignment of type ***hiscore***.

Weak label assignment analysis of [Bäuml2013]

- *spk assigned*: percentage of tracks assigned a weak label
- *err total*: number of wrong weak labels
- *iderr named ft:* face tracks for the named characters who are identified erroneously
- *iderr KNN*: counts how many wrongly labeled tracks are "close" to wrong weak labels

	BBT-1	BBT-2	BBT-3	…	Total	BF-1	BF-2	BF-3	…	Total
#tracks	657	615	660		3920	796	1004	1194		5861
#characters	7	6	8		12	12	13	14		27
spk precision	89.1	87.5	93.2	…	89.8	89.2	82.8	81.4	…	85.9
spk assigned	22.4	16.9	20.6		17.9	19.8	19.1	14.8		18.7
err total	16	13	14		**72**	17	33	33		**155**
err *t404*	2	2	1	…	13	5	10	11	…	46
err *hiscore*	14	11	13		59	12	23	22		109
id acc.	91.0	91.1	75.0		80.2	81.0	70.8	76.3		74.9
iderr named ft	54	53	67	…	**422**	144	198	281	…	**1176**
iderr KNN=1	22	14	13		87	18	24	39		126
iderr KNN=20	42	35	45		257	47	70	123		462

Please refer to Table 1 of the paper for a more detailed overview.

M. Bäuml, M. Tapaswi and R. Stiefelhagen. Semi-supervised Learning with Constraints for Person Identification in Multimedia Data. In *CVPR* 2013.

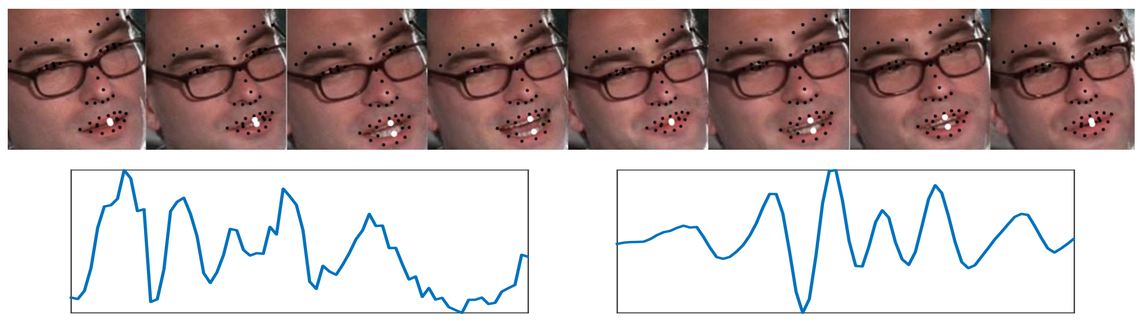
Contact

{tapaswi, baeuml}@kit.edu

Project page (code)

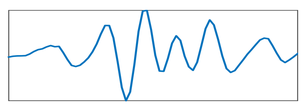
http://cvhci.anthropomatik.kit.edu/projects/mma

Is the face track speaking?



lip motion without / with band-pass filter (BPF) (3.75 – 7.5Hz)

Improved assignment model



SPEAKING SCORE

positively influences assigning label to a track

THREADING

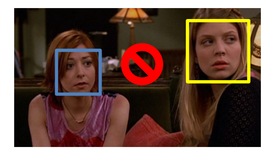
same weak label to tracks in a thread

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \; \underbrace{-w_S \sum_i \boldsymbol{x}_i^T s_i}_{\text{speaking}} \underbrace{-w_T \sum_{(t_i,t_j)\in\mathcal{P}} \boldsymbol{x}_i^T \boldsymbol{x}_j}_{\text{threading}} + \underbrace{w_R \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i}_{\text{regularizer}} \underbrace{+w_U \sum_{(t_i,t_j)\in\mathcal{N}} \boldsymbol{x}_i^T \boldsymbol{x}_j}_{\text{uniqueness}}$$

$$\text{subject to } \sum_{c=1}^{C} x_i^c = 1$$

CONSTRAINED OPTIMIZATION

solve for the weak label assignment to minimize an "energy" considering all terms

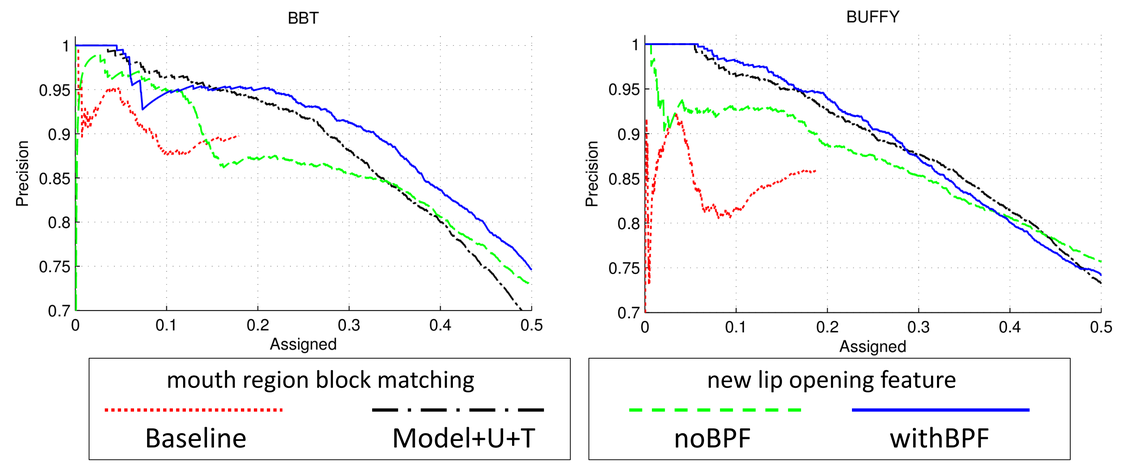


UNIQUENESS

different labels to tracks in the same frame

Results

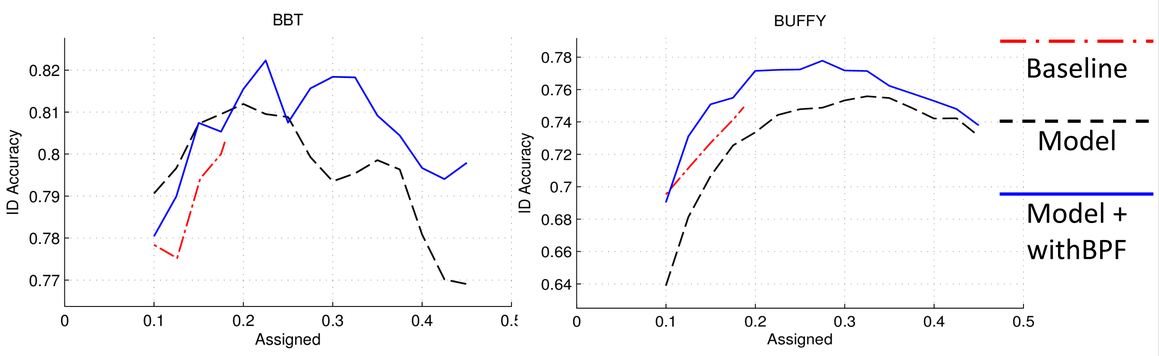
- Weak label assignment performance



- New lip opening feature 200x faster!
- Reduces weak label assignment errors @ same recall

		spk prc.	spk asg.	err total
BBT	Baseline	89.8	17.9	72
	M+U+T	94.3	18.0	40
BUFFY	Baseline	85.9	18.7	155
	M+U+T	93.5	18.8	72

- Face track identification performance



	BBT-1	BBT-2	BBT-3	BBT-4	BBT-5	BBT-6	BF-1	BF-2	BF-3	BF-4	BF-5	BF-6
baseline	91.0	91.1	75.0	80.9	81.1	61.8	81.0	70.8	76.3	76.3	73.8	71.3
model	91.5	92.0	76.7	80.9	83.2	**62.9**	82.0	70.0	77.3	75.8	76.8	**71.6**
model + withBPF	**92.9**	**92.7**	**78.6**	**83.0**	**84.7**	61.5	**83.7**	**75.7**	**79.3**	**77.7**	**80.0**	70.4

Acknowledgment: This work was funded by the German Research Foundation (DFG).