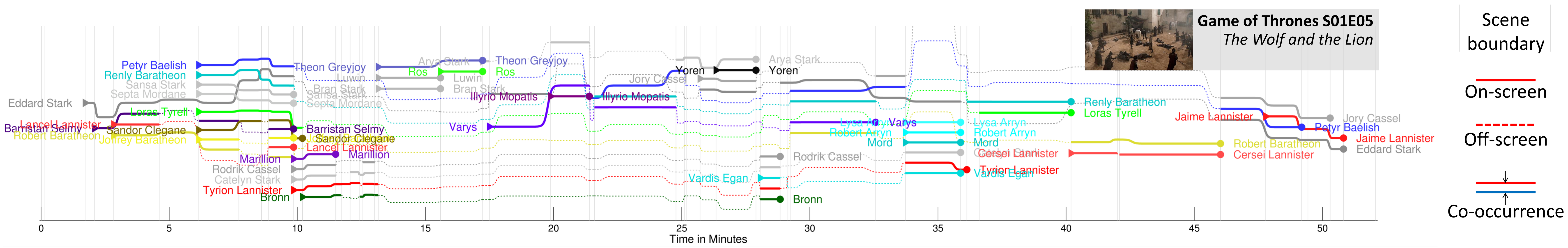
StoryGraphs
Visualizing Character Interactions as a Timeline

Makarand Tapaswi, Martin Bäuml, Rainer Stiefelhagen

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

Contributions

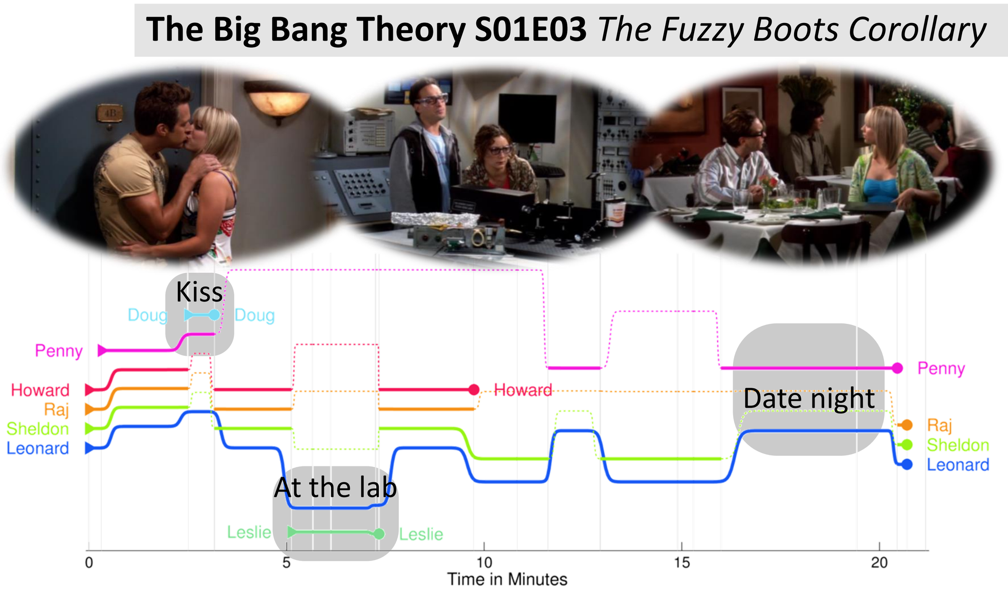
- An automatic approach to visualize character interactions within videos as timeline charts
- Positions of character lines in the *StoryGraph* are formulated as an optimization problem trading off *aesthetics* and *functionality*
- Video scene change detection based on dynamic programming using cues such as shot threading



Motivation

Inspired by the comic strip http://xkcd.com/657 we present a technique to automatically visualize character interactions of a TV episode. StoryGraphs can be used in *smart browsing*, or *video retrieval* such as searching for favorite scenes in the episode. They also see applications in other areas which depend on analysis of people-people interactions such as *crisis control room* analysis, *group meetings*, etc.

The Big Bang Theory S01E03 *The Fuzzy Boots Corollary*



Scene detection

- *Shot threading*

via SIFT matches

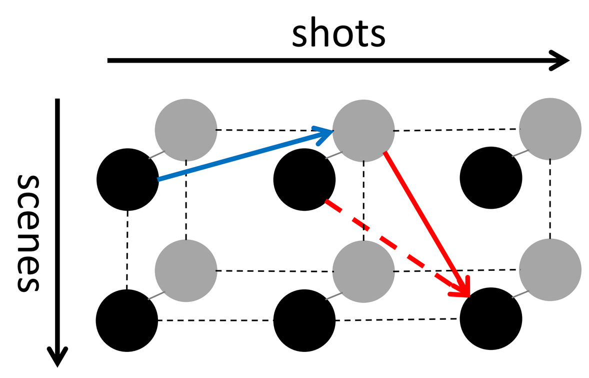
- *Dynamic programming*

Group shots into scenes to maximize internal color-based shot similarity and retain shot-threads within one scene

$$\underbrace{\boxed{\mathcal{S}^*}}_{\text{scenes}} = \underset{S_i}{\operatorname{argmax}} \sum_{i=1}^{N_{sc}} \sum_{s \in S_i} \alpha(|\mathcal{P}|)\big(\underbrace{\boxed{\phi(\mathcal{C}(s,\mathcal{P}))}}_{\text{color-similarity}} + \underbrace{\boxed{T_{s,\mathcal{P}}}}_{\text{threading}}\big)$$

- *Concept*

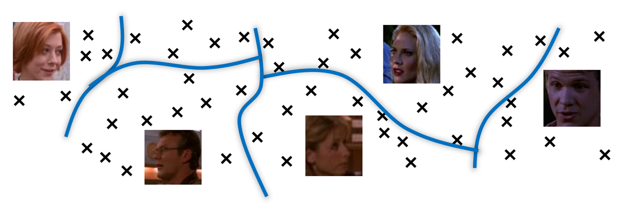
Add shot to	Color similarity	Part of thread
Same scene	high ↑	✓
New scene	low ↓	×



Face tracking and recognition [Bäuml2013]

Multi-pose face detector and particle filter tracker

Multinomial logistic regression classifiers



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 (data and code)
http://cvhci.anthropomatik.kit.edu/projects/mma

StoryGraphs loss functions

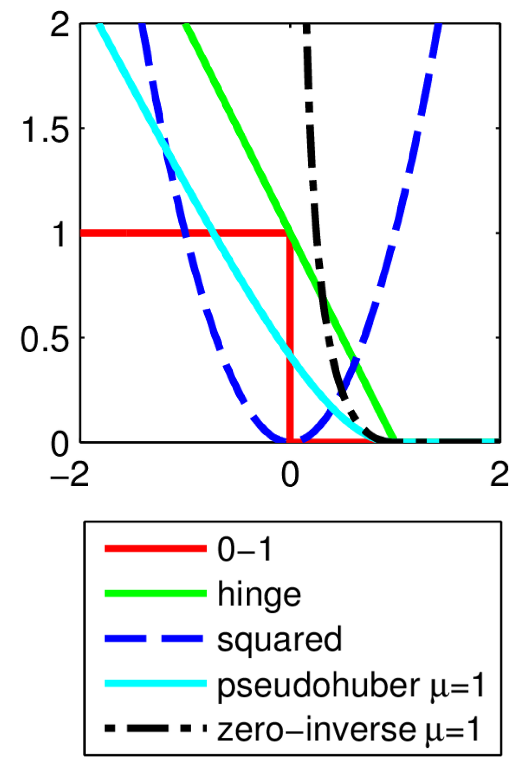
- *proximity*

bring co-occurring characters closer

$$L_p^{(1)} = \frac{1}{N_P N_T} \sum_{c_i,c_j,t} p_{c_i,c_j,t} \cdot (x_t^{c_i} - x_t^{c_j})^2$$

separate characters that do not appear together

$$L_p^{(2)} = \frac{1}{N_P N_T} \sum_{c_i,c_j,t} \mathbb{1}\{p_{c_i,c_j,t} = 0\} \cdot (x_t^{c_i} - x_t^{c_j})^2$$



- prefer *straight lines*

$$L_l = \frac{1}{N_C N_T} \sum_{c,t} (x_t^c - \mu_{\setminus x_t^c})^2$$

- maintain *minimum separation*

$$L_s = \frac{1}{N_P N_T} \sum_{c_i,c_j,t} \mathcal{Z}((x_t^{c_i} - x_t^{c_j})^2, \mu_s)$$

- minimize number of *crossings*

$$L_c = \frac{1}{N_P N_T} \sum_{c_i,c_j,t} \mathcal{H}((x_t^{c_i} - x_t^{c_j})(x_{t+1}^{c_i} - x_{t+1}^{c_j}), \mu_c)$$

- *optimize*

$$\mathcal{L}(\mathbf{x}) = L_p^{(1)}(\mathbf{x}) - L_p^{(2)}(\mathbf{x}) + L_l(\mathbf{x}) + L_s(\mathbf{x}) + L_c(\mathbf{x})$$

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{x})$$

Results

- Scene detection evaluation

Episode	$\sigma(s)$		R_{30}	
	Ours	[15]	Ours	[15]
BBT-1	**18.5**	21.4	**75.0**	62.5
BBT-2	**8.0**	16.2	**90.9**	81.8
BF-1	**13.4**	19.5	**84.9**	72.7
BF-2	**12.0**	16.3	**96.7**	86.7
GOT-1	**12.6**	21.6	**88.6**	80.0
GOT-2	**17.1**	29.5	**87.1**	77.4

[15] Z. Rasheed and M. Shah. Detection and Representation of Scenes in Videos. *IEEE Transactions on Multimedia*, 7(6): 1097–1105, 2005.

- Impact of person identification on StoryGraphs

	BBT (6)	BUFFY (6)	GOT (10)
#Characters	11	27	66
Mean Accuracy	92.36%	78.12%	75.25%
SG Presence Error Count	33	260	680
SG Presence Error Fraction	4.85%	8.08%	4.24%

- Quantitative evaluation

	BBT-1	BBT-2	BF-1	BF-2	GOT-1	GOT-2
Move	0.32	0.34	0.46	0.29	0.30	0.21
MaxSep	0.45	0.67	0.30	0.34	0.34	0.43
Sep > $2\mu_s$	4	12	0	2	1	6
Sep < $0.5\mu_s$	1	1	1	0	1	3
#Cross	0	1	83	43	212	66

Acknowledgment: This work was supported by the Quaero Programme, funded by OSEO.