Story comparison for estimating field of view overlap in a video collection

ICPR 2020 - Oral Presentation

January 11th 2021

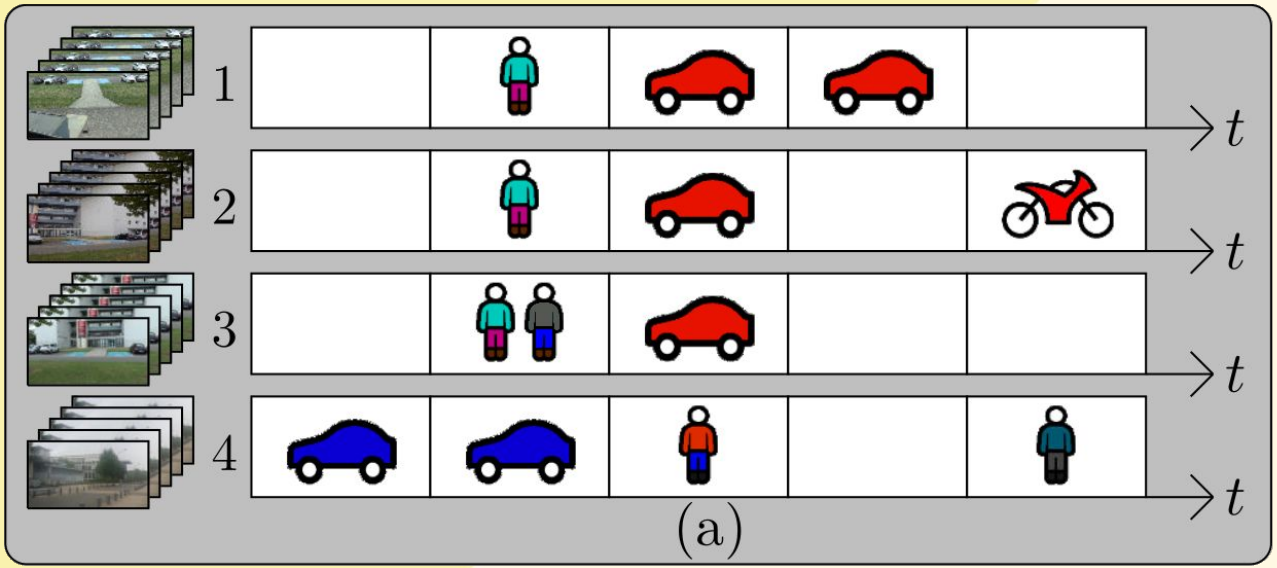
Institut de Recherche en Informatique de Toulouse (IRIT)

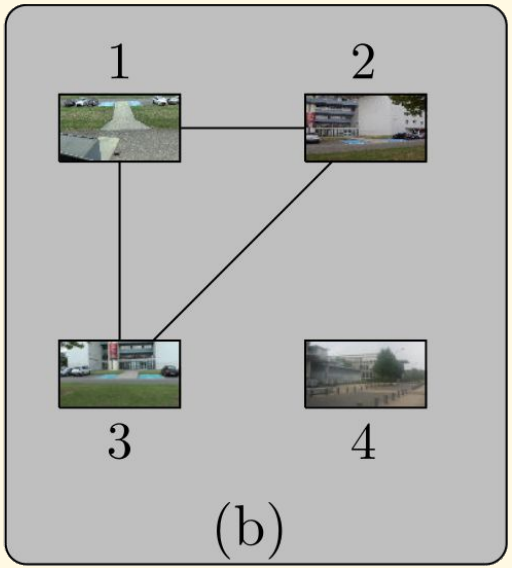
Thierry Malon (speaker)
Sylvie Chambon
Alain Crouzil
Vincent Charvillat

Overview of the proposed approach

Goal: automatically finding videos of a collection that have overlapping fields of view

Hypothesis: static cameras and temporally synchronised videos, no other metadata

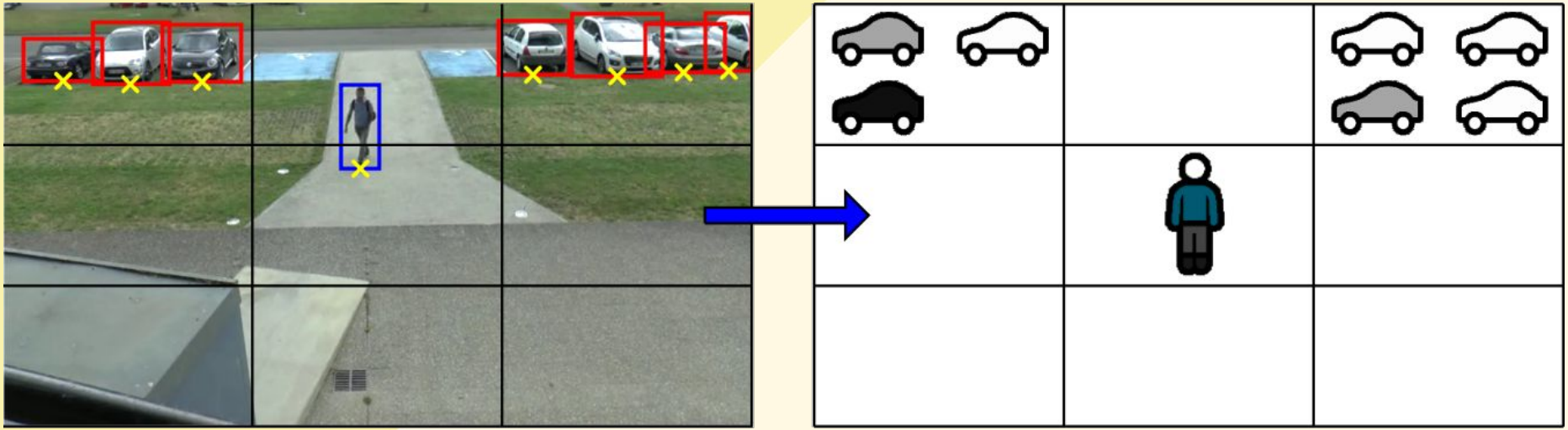




Detection step

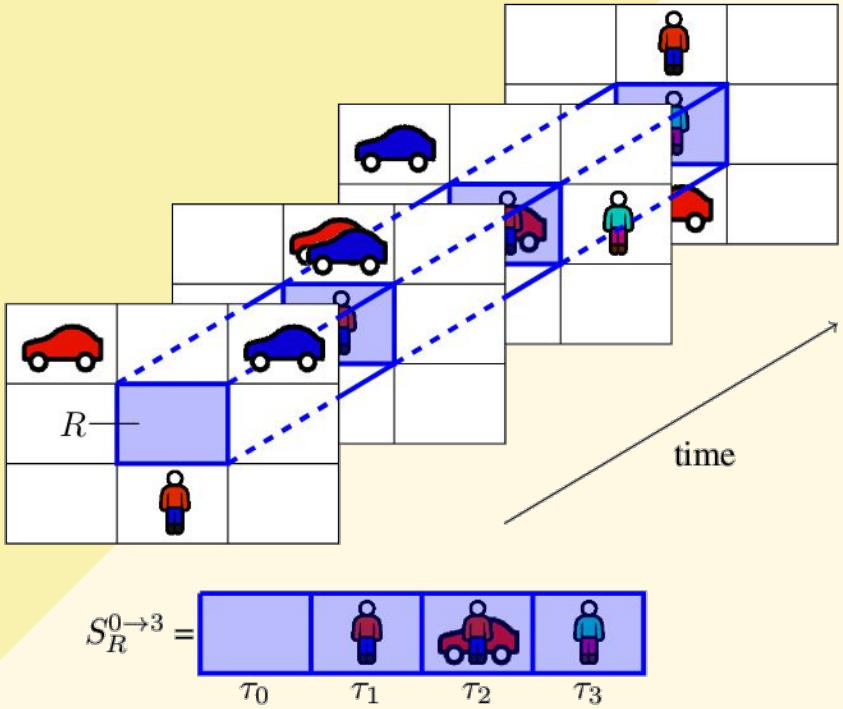
Detection: using 3 different existing CNN approaches
(SSD [Liu, 2016], Mask-RCNN [He, 2017], YOLOv3 [Redmon, 2018])

Descriptors: using 3 different existing appearance descriptors
(HOG [Dalal, 2005], ColorNames [Yang, 2014], Latent representation of Resnet18 [He, 2016])



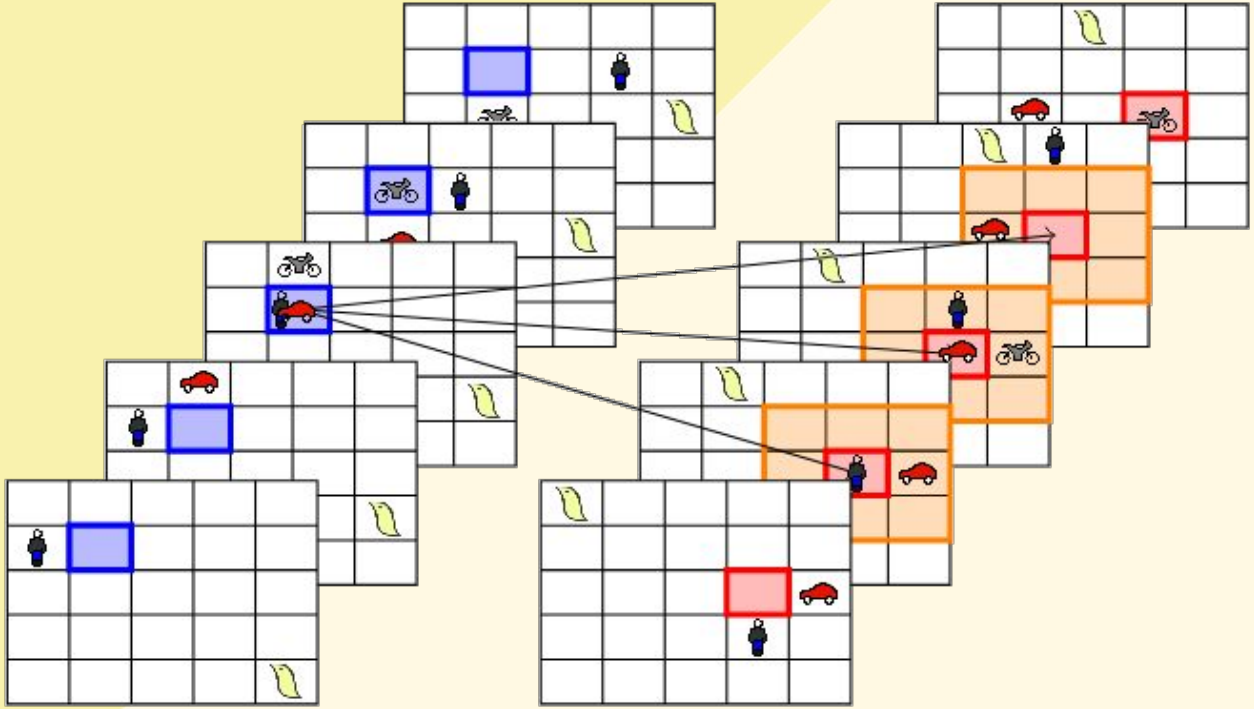
Definition of a region story

Story of a region: list of objects (category + appearance descriptor) detected at regular time steps in the region



Comparison between stories

Distance between stories: proportion of objects from each story which have a correspondent object in the spatiotemporal neighborhood of the other story



Multiresolution

To avoid comparing all possible region pairs, we compare regions at different scales and keep on comparing at a thinner scale if the distance if neither in the accepted link interval (in green) nor in the rejected link interval (in red)

$d(S_R, S_{R'}) = 0.809$

(a) $s = 1$

$d(S_R, S_{R'}) = 0.992$

(b) $s = 2$

$d(S_R, S_{R'}) = 0.573$

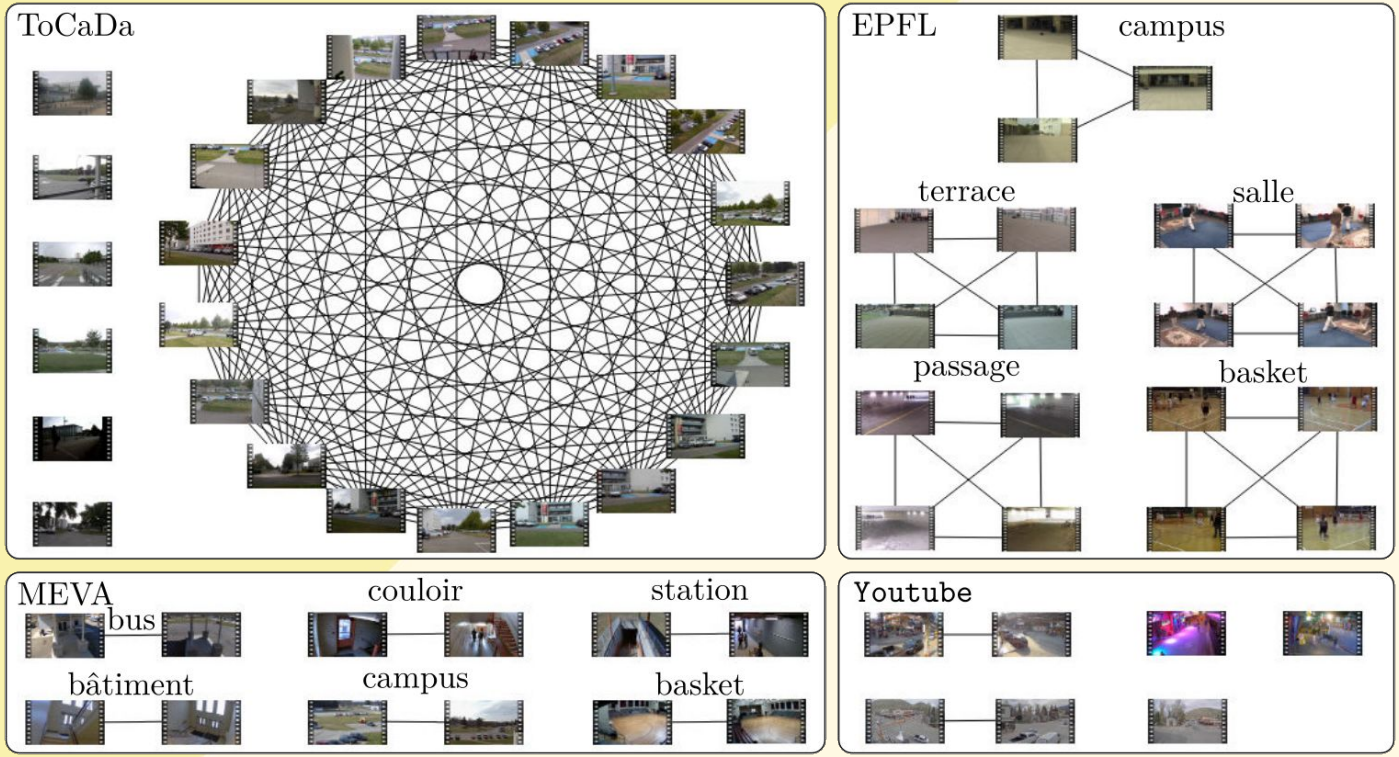
(c) $s = 2$

$d(S_R, S_{R'}) = 0.183$

(d) $s = 3$

Experiments

Datasets: multiview datasets (ToCaDa, EPFL, MEVA, Youtube videos)



Experiments

Evaluation: F1 score of the overlap links between videos

True Positive (TP)

False Negative (FN)

True Negative (TN)

Precision = TP / (TP + TN)

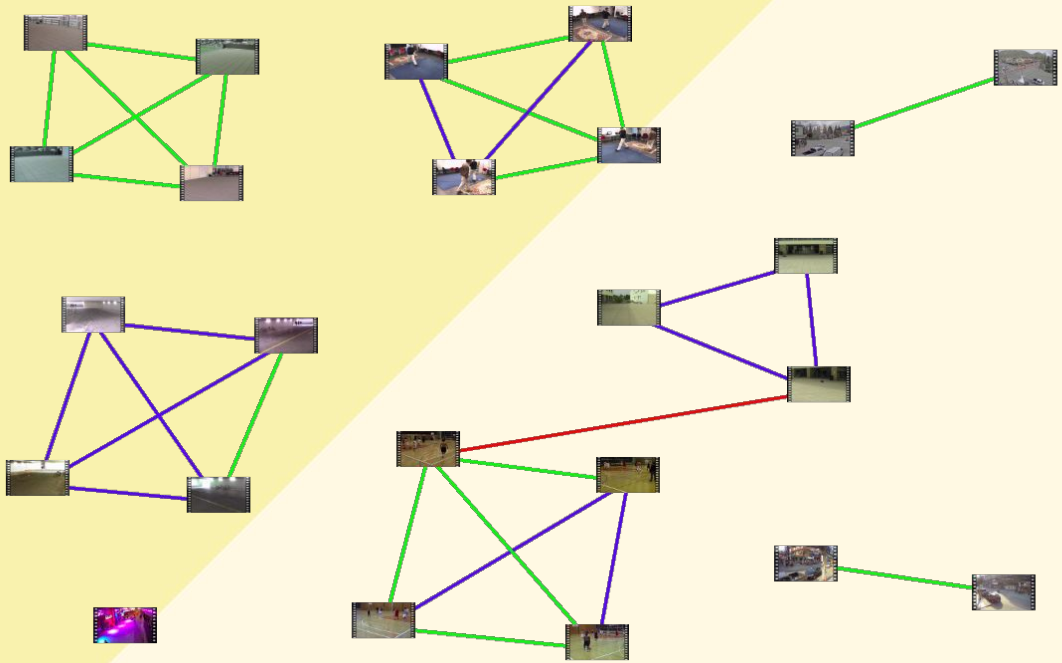
Recall = TP / (TP + FN)

F1 score = 2 × Precision × Recall / (Precision + Recall)

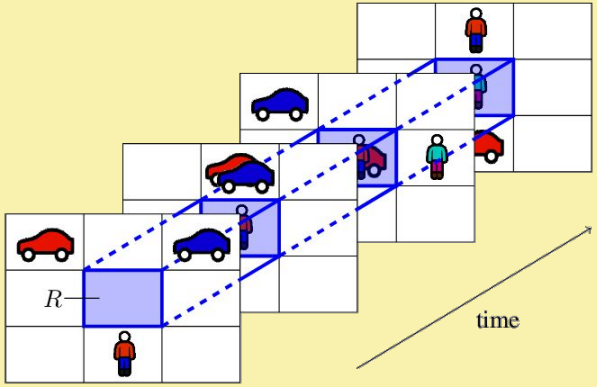
	SSD					Mask-RCNN					YOLOv3				
Dataset	base	ctg	ctgcn	ctghg	ctgld	base	ctg	ctgcn	ctghg	ctgld	base	ctg	ctgcn	ctghg	ctgld
Live Cameras	19	22	27	25	25	80	**100**	**100**	**100**	**100**	80	**100**	**100**	**100**	**100**
MEVA	28	29	32	30	35	59	60	67	62	**80**	59	59	65	68	76
EPFL	56	56	64	59	67	61	61	71	65	79	63	63	74	66	**81**
ToCaDa	25	29	34	32	35	83	87	88	87	88	85	88	89	88	**90**
All real videos	16	19	21	20	23	32	36	43	39	47	35	38	47	44	**51**

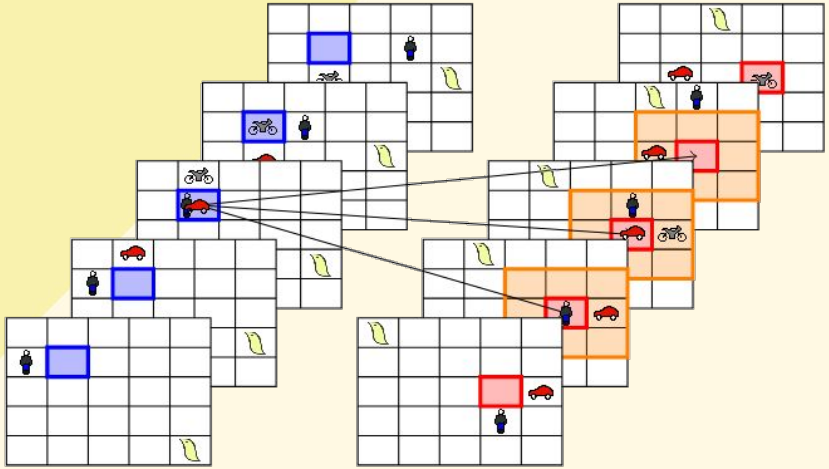
Experiments

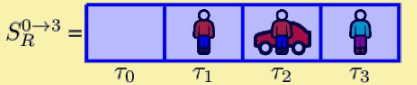
Overlap graph: overlap links found on a subset of the dataset

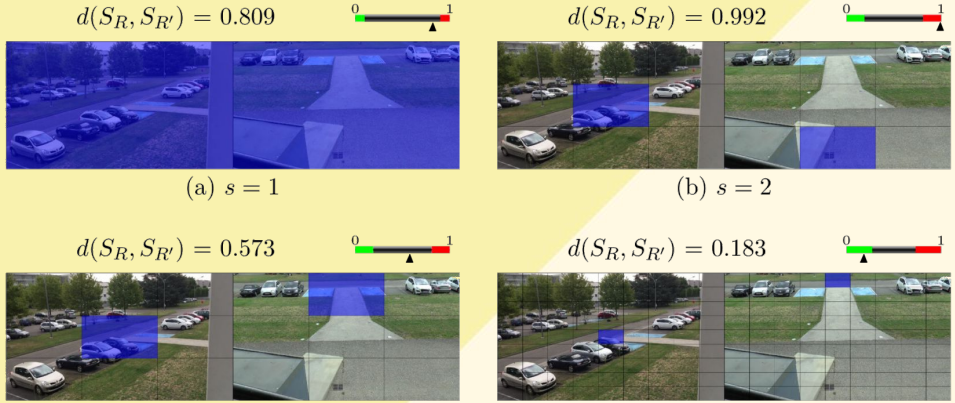


Questions

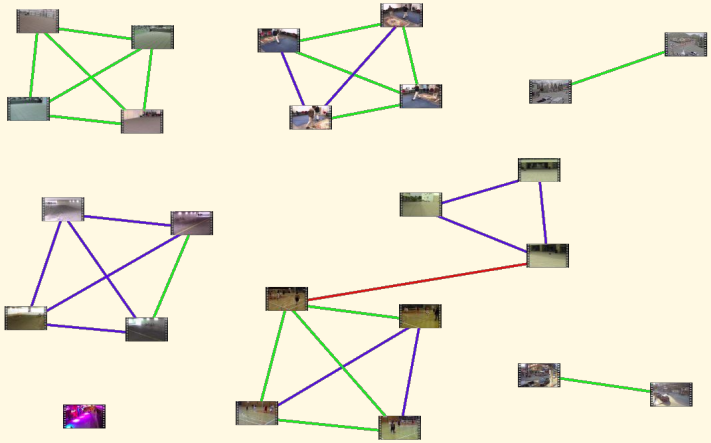








(a) $s = 1$ (b) $s = 2$ (c) $s = 2$ (d) $s = 3$



Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)

Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)

Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)

Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)

Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources).

Problem statement

From a collection of videos, how to automatically