A predictor of moving objects for first-person vision

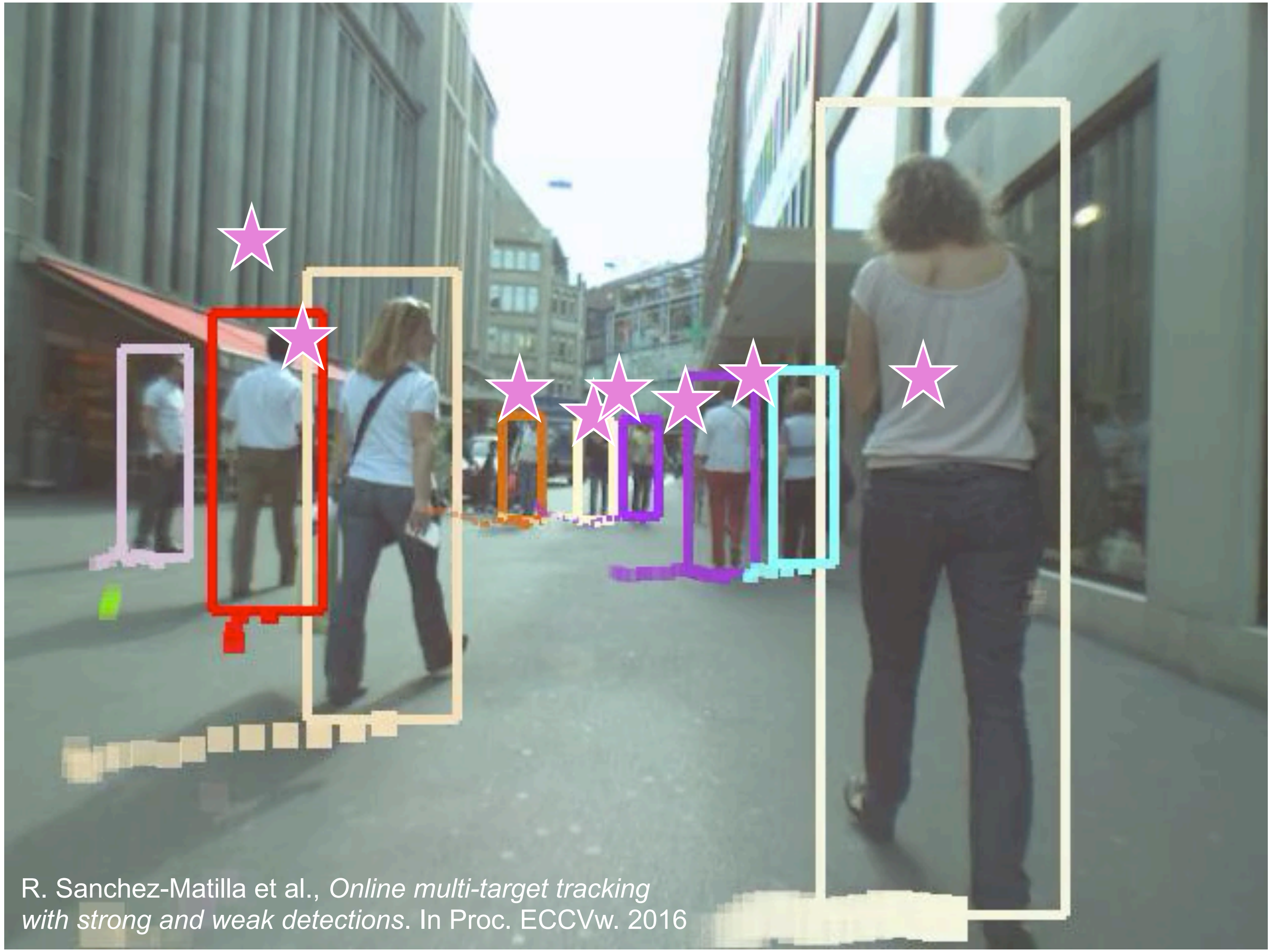
Ricardo Sanchez-Matilla and Andrea Cavallaro

CIS centre for intelligent sensing

Queen Mary University of London

Linear motion prediction for tracking

Drifting tracks



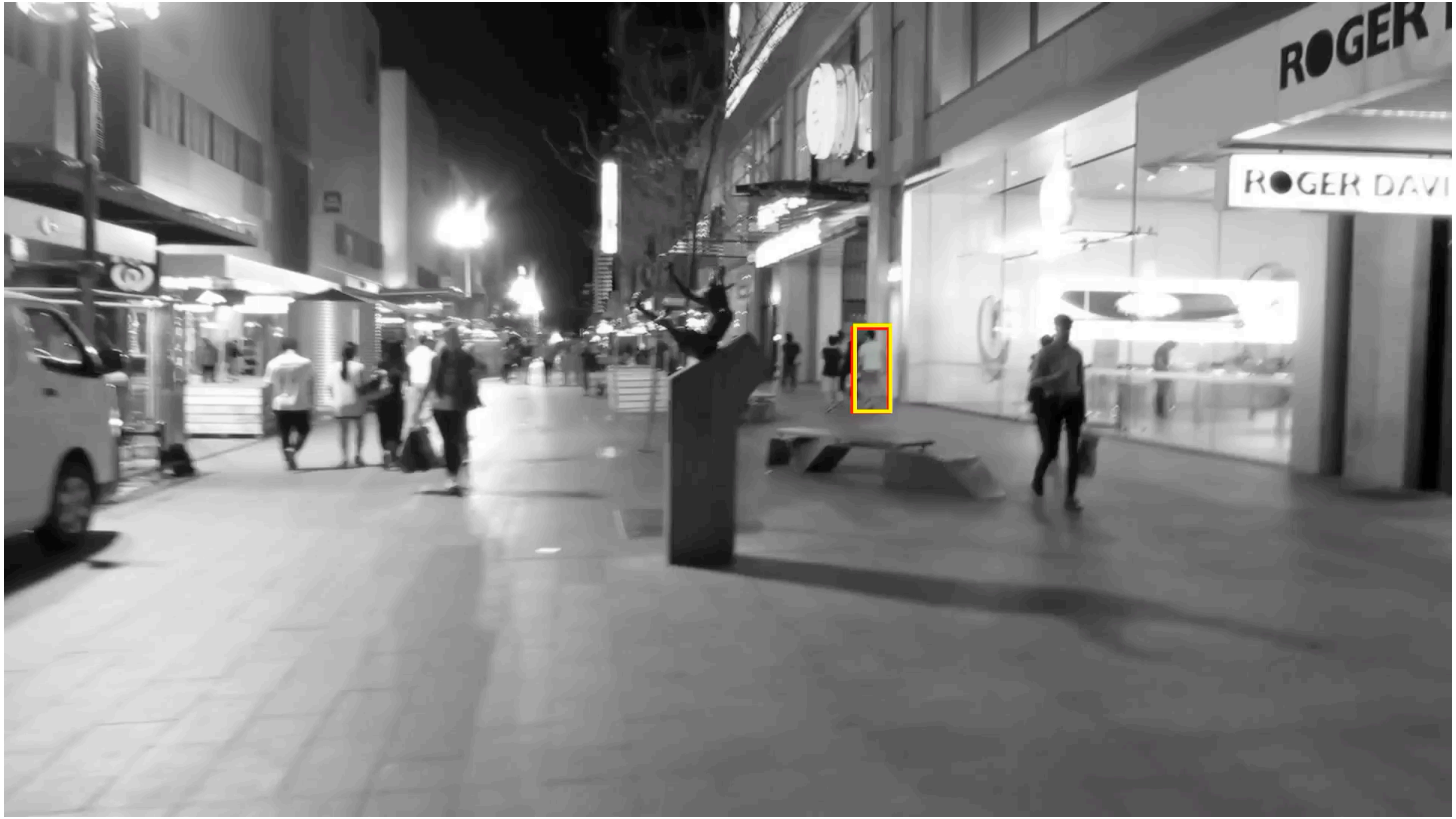
R. Sanchez-Matilla et al., *Online multi-target tracking with strong and weak detections*. In Proc. ECCVw. 2016

Note
tracker does not use appearance features

CIS centre for intelligent sensing

Queen Mary University of London

Linear prediction of a moving object (with a moving camera)



Manual annotation

Linear prediction

Challenges
- unknown camera motion
- unknown object motion

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]			

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]	👎	👎	👎
[2]		👎	👎

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017
[2] G. Mattyus et al., *Multi target tracking on aerial videos*. In Proc. ISPRS Workshop. 2010

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]			
[2]			
[3]			

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017
[2] G. Mattyus et al., *Multi target tracking on aerial videos*. In Proc. ISPRS Workshop. 2010
[3] S. Li et al., *Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models*. In Proc. AAAI *2017*

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]	👎	👎	👎
[2]		👎	👎
[3]		👎	👎
[4]		👎	👎

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017
[2] G. Mattyus et al., *Multi target tracking on aerial videos*. In Proc. ISPRS Workshop. 2010
[3] S. Li et al., *Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models*. In Proc. AAAI *2017*
[4] S. Lankton et al., *Improved tracking by decoupling camera and target motion*. In Proc. SPIE. 2018

What information is needed?

	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]	👎	👎	👎
[2]		👎	👎
[3]		👎	👎
[4]		👎	👎
[5]	👎		

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017
[2] G. Mattyus et al., *Multi target tracking on aerial videos*. In Proc. ISPRS Workshop. 2010
[3] S. Li et al., *Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models*. In Proc. AAAI *2017*
[4] S. Lankton et al., *Improved tracking by decoupling camera and target motion*. In Proc. SPIE. 2018
[5] W. Choi et al., *A general framework for tracking multiple people from a moving camera*. IEEE TPAMI. 2018

What information is needed?

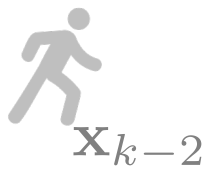
	Assumption		
Reference	Camera calibration	Scene specific	Object-location specific
[1]	👎	👎	👎
[2]		👎	👎
[3]		👎	👎
[4]		👎	👎
[5]	👎		
Proposed	👍	👍	👍

[1] J. Arrospide et al., *Homography-based ground plane detection using a single on-board camera*. In IET ITS. 2017
[2] G. Mattyus et al., *Multi target tracking on aerial videos*. In Proc. ISPRS Workshop. 2010
[3] S. Li et al., *Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models*. In Proc. AAAI *2017*
[4] S. Lankton et al., *Improved tracking by decoupling camera and target motion*. In Proc. SPIE. 2018
[5] W. Choi et al., *A general framework for tracking multiple people from a moving camera*. IEEE TPAMI. 2018

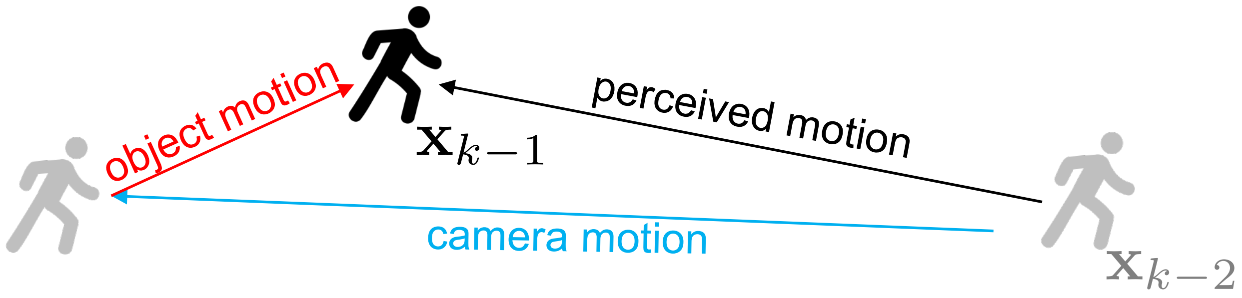
The idea

$\mathbf{x}_{k-2}$

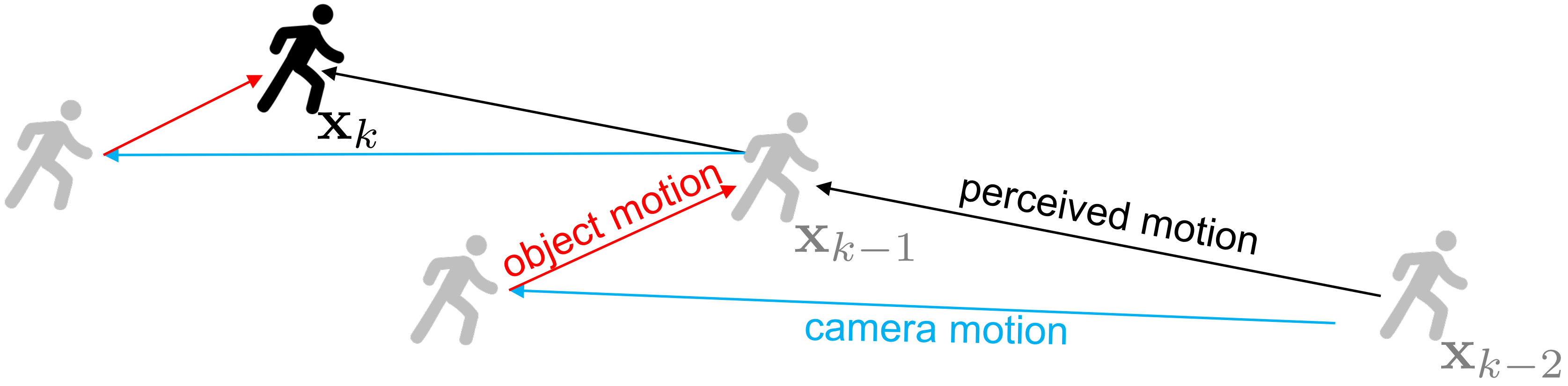
The idea



The idea


object motion
perceived motion
camera motion
$\mathbf{x}_{k-1}$
$\mathbf{x}_{k-2}$

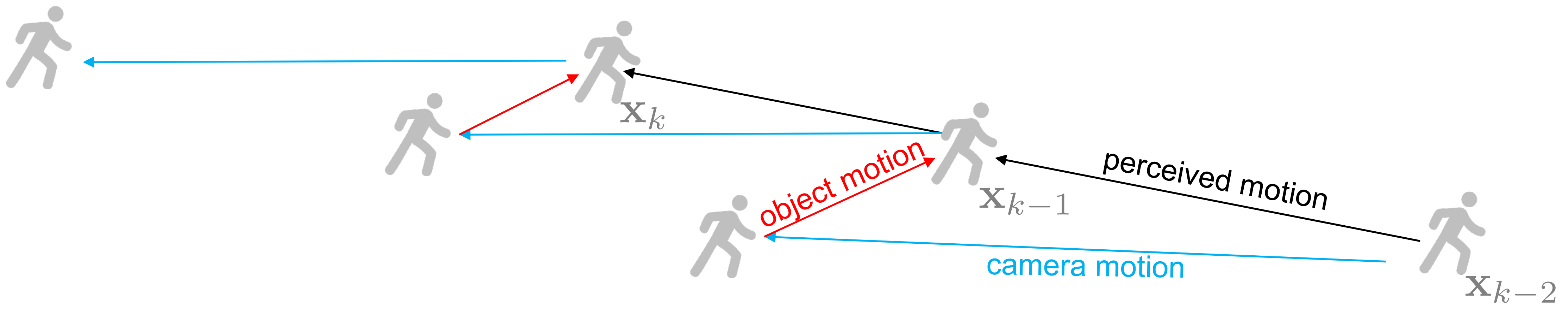
The idea



$\mathbf{x}_k$
object motion
$\mathbf{x}_{k-1}$
perceived motion
camera motion
$\mathbf{x}_{k-2}$

The proposed approach

T_P : observed frames

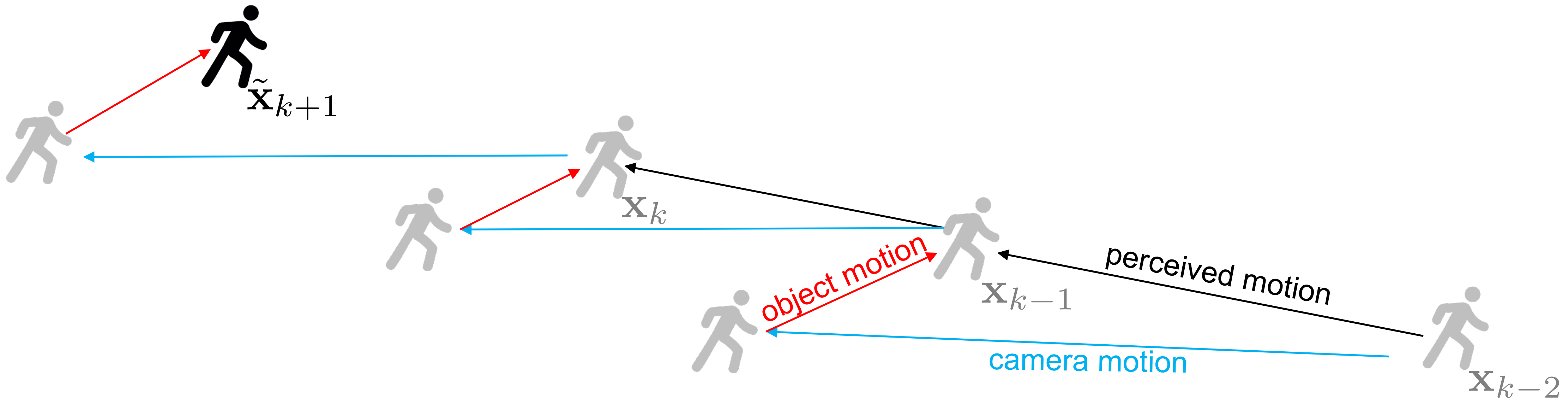


$$\frac{1}{\alpha_k}\mathbf{H}_{k|k-1}\mathbf{x}_k$$

camera motion

The proposed approach

T_P : observed frames

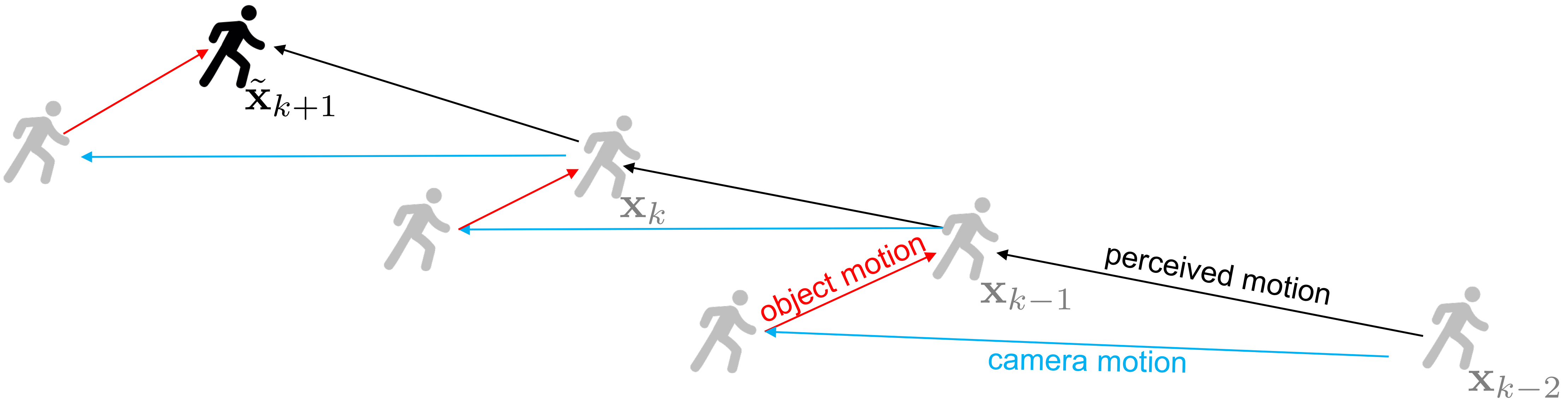


$$\frac{1}{\alpha_k}\mathbf{H}_{k|k-1}\mathbf{x}_k + \dot{\mathbf{x}}_{k|k-T_P+1}$$

camera motion | object motion

The proposed approach

T_P : observed frames



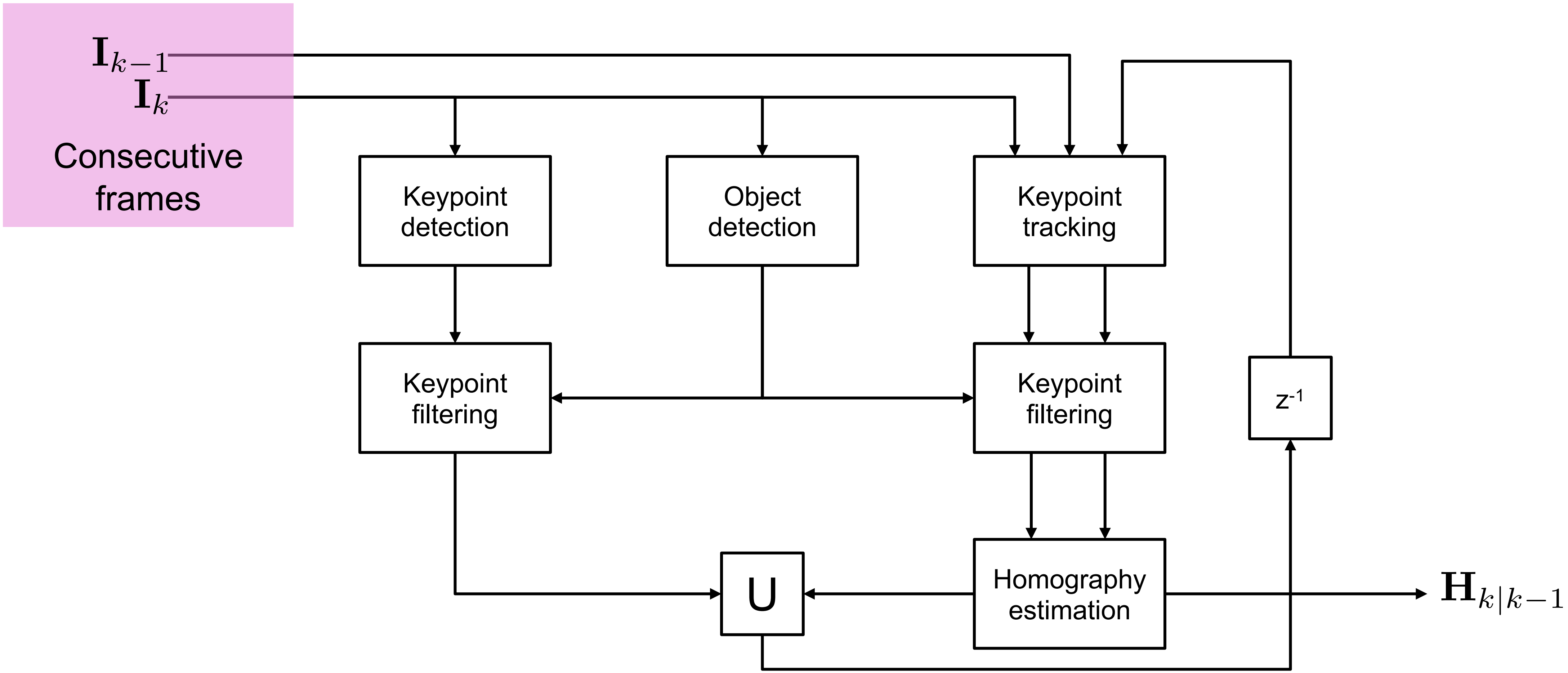
$$\tilde{\mathbf{x}}_{k+1} = \frac{1}{\alpha_k}\mathbf{H}_{k|k-1}\mathbf{x}_k + \dot{\mathbf{x}}_{k|k-T_P+1}$$

object state

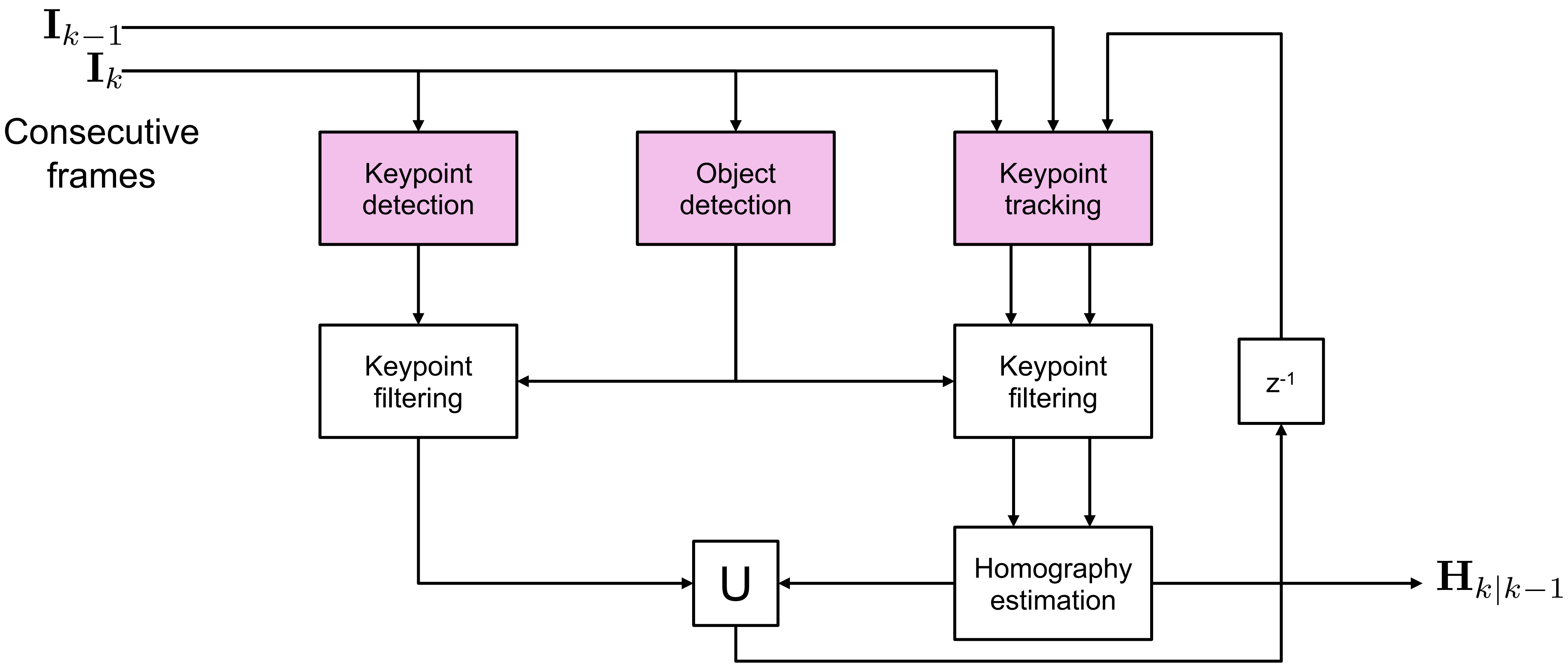
camera motion

object motion

Camera motion estimation

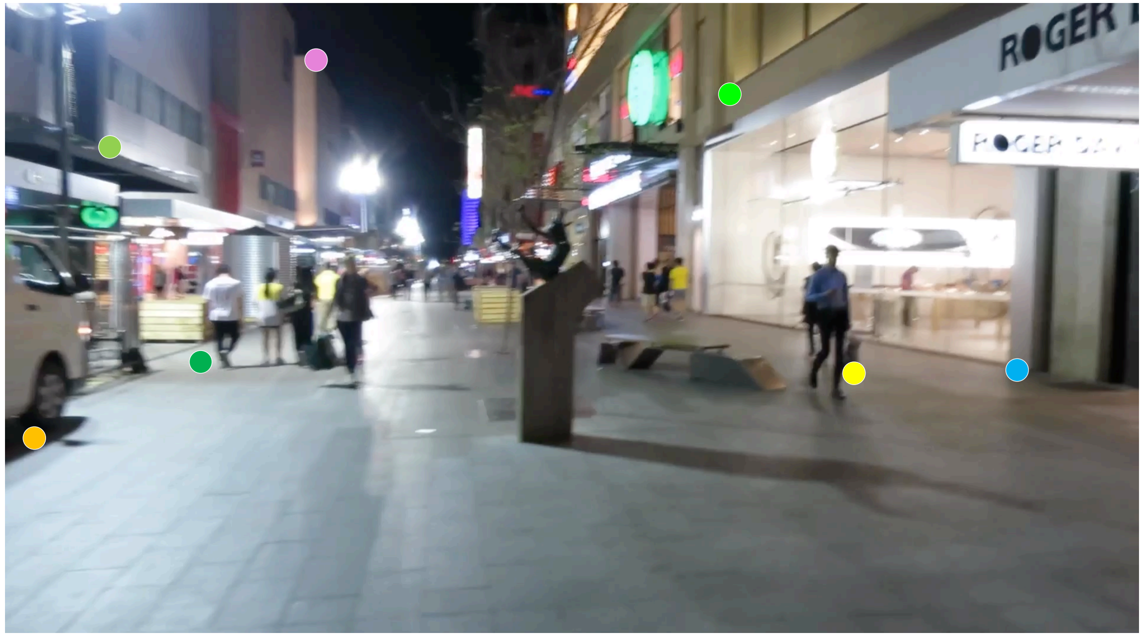


Camera motion estimation



Keypoint detection

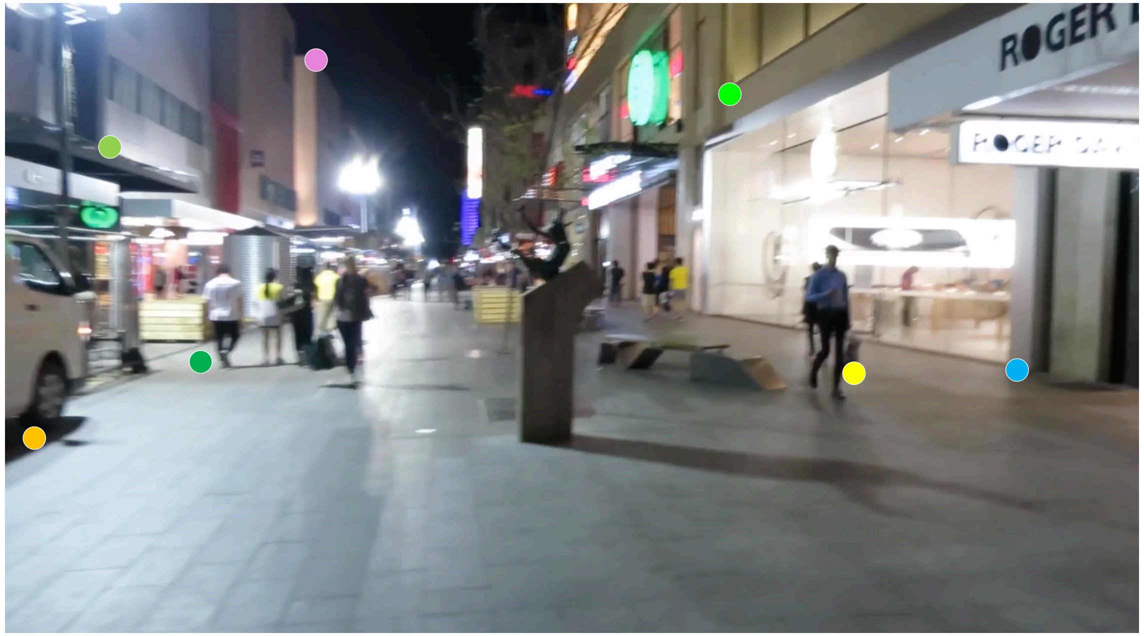
Frame k-2



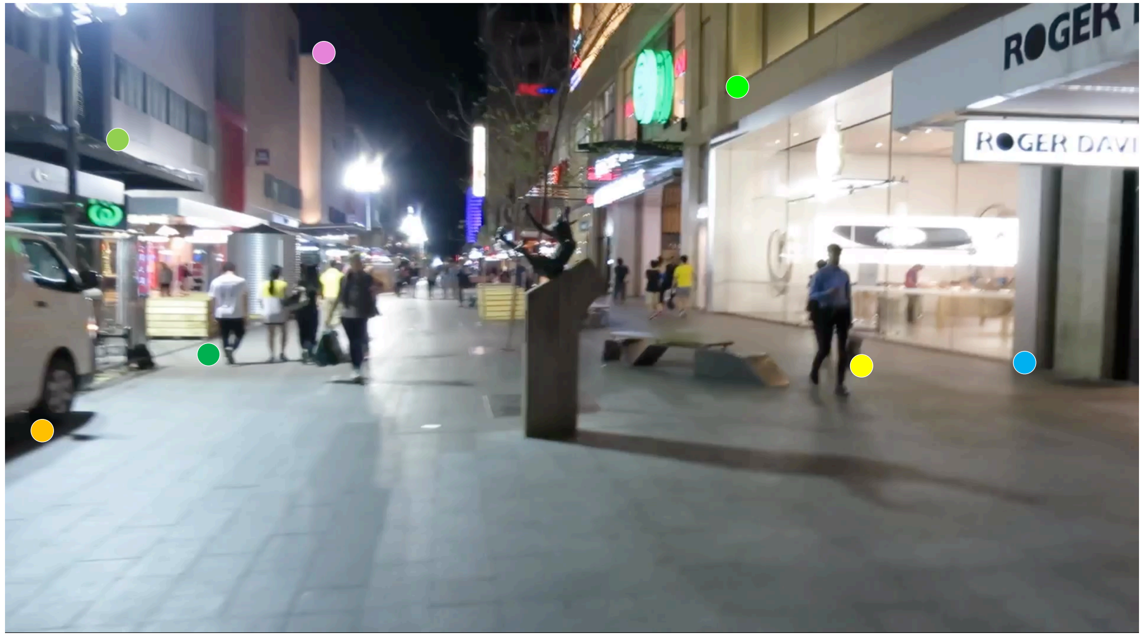
Keypoints

Keypoint tracking

Frame k-2



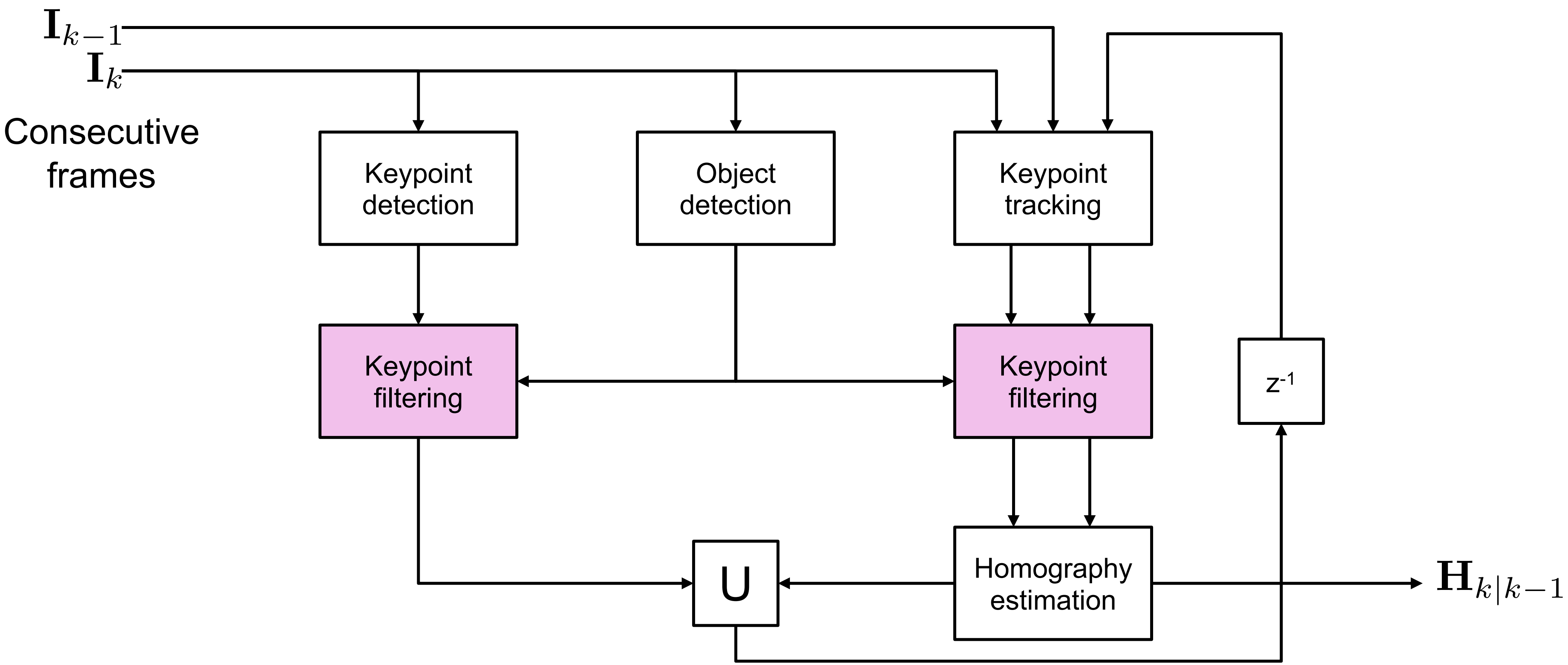
Frame k-1



Keypoints

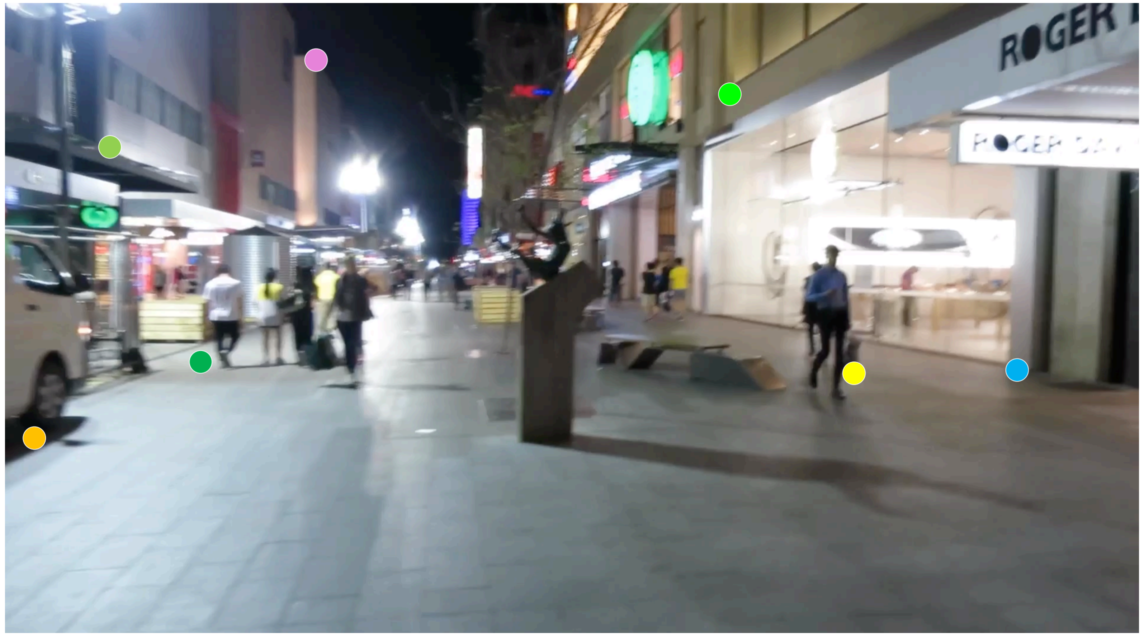
Note
The colours indicate the association between keypoints

Camera motion estimation

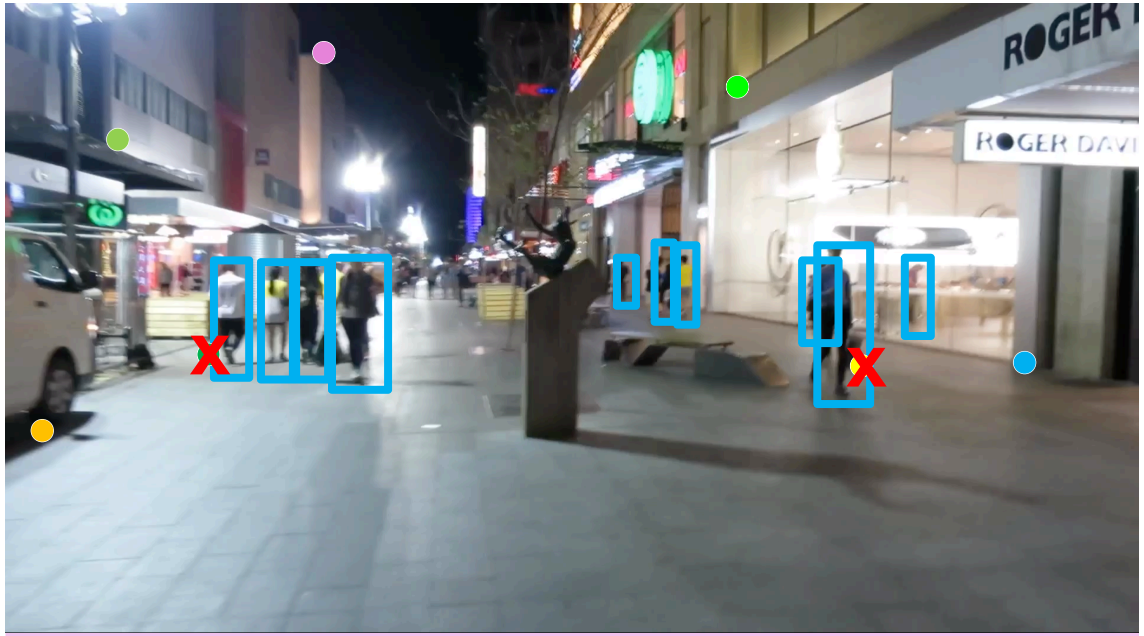


Keypoint filtering

Frame k-2



Frame k-1

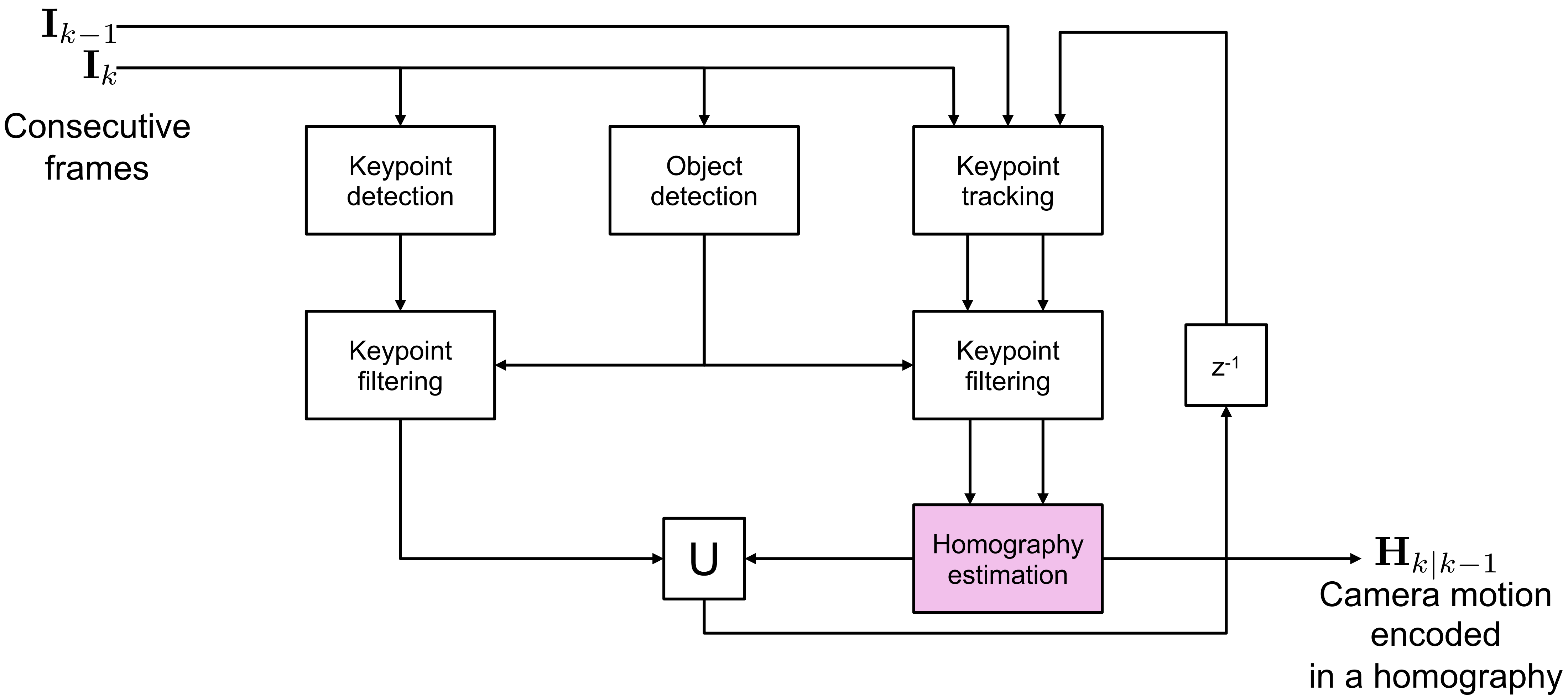


Keypoints within (potential) moving objects are removed

Keypoints

Object detection

Camera motion estimation



Validation

- Dataset: Multiple Object Tracking Challenge 2015, 2016 and 2017
- Prediction accuracy: mean squared error (bounding boxes centres)
- Comparison
  - Location based
    - Static Prediction (SP)
    - Linear Prediction (LP) [6]
    - Exponential Prediction (EM) [7]
    - Linear Regression (LR)
    - LSTM [8]
  - Location and image based
    - Simple Homography-based prediction (SH) [3]
    - Proposed with keypoints on ground plane (GMG)
    - Proposed (GM)

[3] S. Li et al., *Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models*. In Proc. AAAI. *2017*
[6] K. Shafique et al., *A rank con- strained continuous formulation of multi-frame multi-target tracking problem*. In CVPR. 2018
[7] V. Akbarzadeh et al., *Target trajectory prediction in PTZ camera networks*. In CVPR. 2013
[8] M. Babaee et al., *Occlusion handling in tracking multiple people using RNN*. In Proc. ICIP. 2018

Prediction accuracy [pixels]

number of observed frames

number of frames to predict

Proposed

T_P	T_F	SP	LP	EM	LR	LSTM	SH	GMG	GM
2	1								
	10								
	20								
	30								
10	1								
	10								
	20								
	30								
20	1								
	10								
	20								
	30								
30	1								
	10								
	20								
	30								

Prediction accuracy [pixels]

number of observed frames

number of frames to predict

Proposed

T_P	T_F	SP		LP		EM		LR	LSTM	SH	GMG	GM
2	1	7.3	(8.2)	**2.1**	**(4.2)**	**2.1**	**(4.2)**					
	10	35.0	(47.4)	*13.7*	*(19.6)*	*13.7*	*(19.6)*					
	20	60.3	(81.2)	*31.2*	*(40.5)*	*31.2*	*(40.5)*					
	30	80.5	(106.1)	50.7	(62.2)	50.7	(62.2)					
10	1	7.2	(8.2)	*2.8*	*(3.4)*	**2.7**	**(3.4)**					
	10	35.0	(46.8)	15.4	(18.7)	14.9	(18.3)					
	20	59.9	(79.4)	32.0	(38.0)	31.3	(37.5)					
	30	79.4	(101.8)	49.7	(57.6)	49.0	(57.0)					
20	1	7.2	(8.2)	3.6	(3.8)	*3.3*	*(3.7)*					
	10	34.9	(46.2)	18.4	(21.7)	17.1	(20.4)					
	20	59.3	(76.8)	35.7	(41.5)	33.8	(39.7)					
	30	78.8	(99.3)	52.9	(60.7)	50.8	(58.7)					
30	1	7.2	(8.2)	4.0	(4.3)	*3.5*	*(3.9)*					
	10	34.6	(45.3)	20.3	(23.7)	18.1	(21.5)					
	20	59.2	(76.0)	38.0	(44.3)	35.1	(41.1)					
	30	78.8	(99.1)	55.4	(65.0)	52.0	(60.8)					

Prediction accuracy [pixels]

number of observed frames

number of frames to predict

Proposed

T_P	T_F	SP		LP		EM		LR		LSTM		SH		GMG	GM
2	1	7.3	(8.2)	**2.1**	**(4.2)**	**2.1**	**(4.2)**	**2.1**	**(4.2)**	6.1	(7.0)	2.8	*		
	10	35.0	(47.4)	*13.7*	*(19.6)*	*13.7*	*(19.6)*	*13.7*	*(19.6)*	27.3	(36.9)	69.9	*		
	20	60.3	(81.2)	*31.2*	*(40.5)*	*31.2*	*(40.5)*	*31.2*	*(40.5)*	47.4	(62.9)	194.6	*		
	30	80.5	(106.1)	50.7	(62.2)	50.7	(62.2)	50.7	(62.2)	64.8	(82.7)	292.7	*		
10	1	7.2	(8.2)	*2.8*	*(3.4)*	**2.7**	**(3.4)**	5.2	(5.8)	5.3	(6.5)	10.9	*		
	10	35.0	(46.8)	15.4	(18.7)	14.9	(18.3)	17.6	(20.2)	25.0	(35.5)	55.4	*		
	20	59.9	(79.4)	32.0	(38.0)	31.3	(37.5)	34.1	(39.2)	44.0	(60.5)	140.3	*		
	30	79.4	(101.8)	49.7	(57.6)	49.0	(57.0)	51.9	(58.8)	59.9	(78.2)	199.8	*		
20	1	7.2	(8.2)	3.6	(3.8)	*3.3*	*(3.7)*	12.1	(11.5)	5.6	(6.5)	11.1	*		
	10	34.9	(46.2)	18.4	(21.7)	17.1	(20.4)	26.1	(26.8)	27.1	(35.9)	63.4	*		
	20	59.3	(76.8)	35.7	(41.5)	33.8	(39.7)	42.8	(45.5)	49.6	(57.2)	145.4	*		
	30	78.8	(99.3)	52.9	(60.7)	50.8	(58.7)	59.8	(64.2)	68.6	(76.3)	200.5	*		
30	1	7.2	(8.2)	4.0	(4.3)	*3.5*	*(3.9)*	19.7	(18.5)	5.9	(6.4)	11.7	*		
	10	34.6	(45.3)	20.3	(23.7)	18.1	(21.5)	34.2	(33.4)	28.2	(34.8)	53.6	*		
	20	59.2	(76.0)	38.0	(44.3)	35.1	(41.1)	50.8	(51.7)	49.4	(59.1)	193.3	*		
	30	78.8	(99.1)	55.4	(65.0)	52.0	(60.8)	67.2	(71.1)	65.1	(81.4)	247.5	*		

* Indicates at least one more order of magnitude

Prediction accuracy [pixels]

number of observed frames

number of frames to predict

Proposed

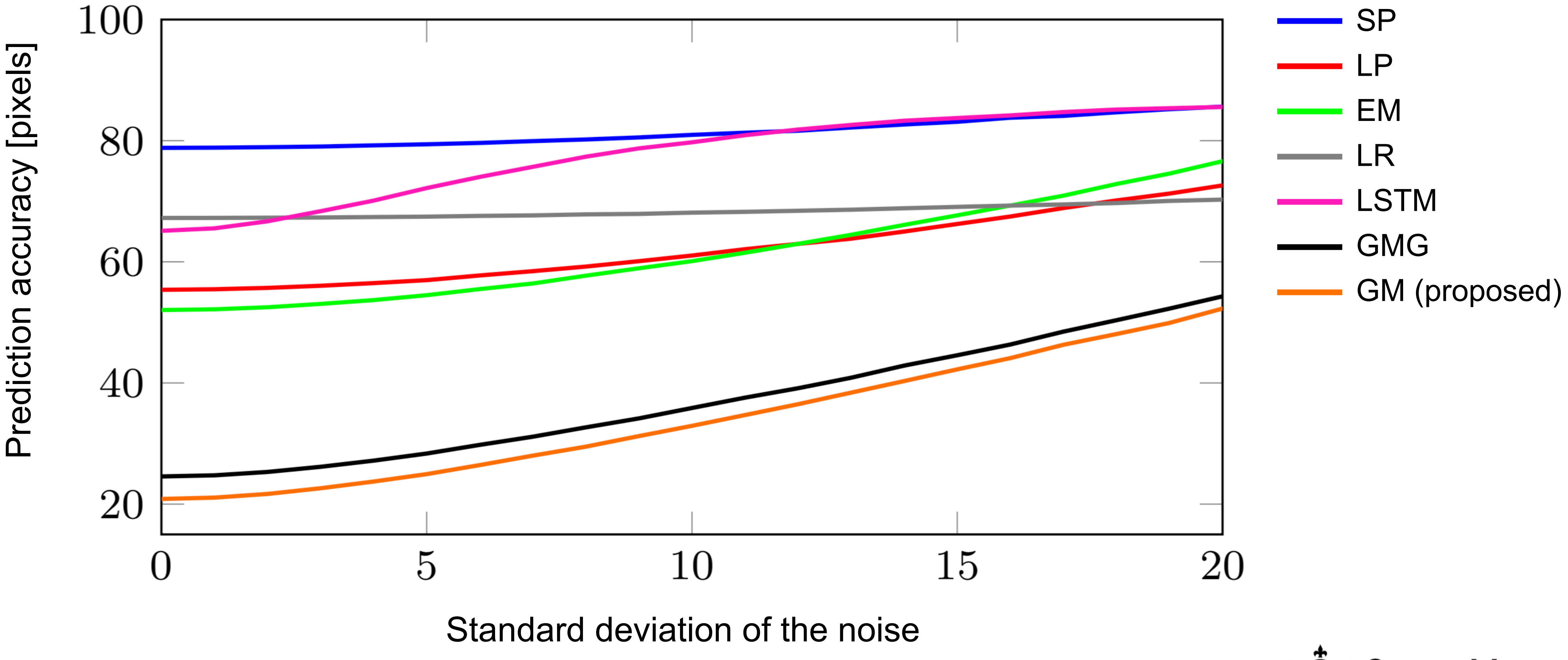
T_P	T_F	SP		LP		EM		LR		LSTM		SH		GMG		GM	
2	1	7.3	(8.2)	**2.1**	**(4.2)**	**2.1**	**(4.2)**	**2.1**	**(4.2)**	6.1	(7.0)	2.8	*	*2.2*	*(4.2)*	*2.2*	*(4.2)*
	10	35.0	(47.4)	*13.7*	*(19.6)*	*13.7*	*(19.6)*	*13.7*	*(19.6)*	27.3	(36.9)	69.9	*	16.0	(26.4)	**13.0**	**(16.8)**
	20	60.3	(81.2)	*31.2*	*(40.5)*	*31.2*	*(40.5)*	*31.2*	*(40.5)*	47.4	(62.9)	194.6	*	31.0	(42.6)	**25.1**	**(31.9)**
	30	80.5	(106.1)	50.7	(62.2)	50.7	(62.2)	50.7	(62.2)	64.8	(82.7)	292.7	*	*46.1*	*(61.3)*	**37.5**	**(46.0)**
10	1	7.2	(8.2)	*2.8*	*(3.4)*	**2.7**	**(3.4)**	5.2	(5.8)	5.3	(6.5)	[illegible]	[illegible]	3.5	(5.2)	2.9	(3.7)
	10	35.0	(46.8)	15.4	(18.7)	14.9	(18.3)	17.6	(20.2)	25.0	[illegible]	[illegible]	*	*11.5*	*(18.8)*	**9.5**	**(14.2)**
	20	59.9	(79.4)	32.0	(38.0)	31.3	(37.5)	34.1	(39.2)	44.0	[illegible]	[illegible]	*	*19.2*	*(30.4)*	**16.0**	**(24.2)**
	30	79.4	(101.8)	49.7	(57.6)	49.0	(57.0)	51.9	(58.8)	[illegible]	[illegible]	199.8	*	*26.7*	*(39.8)*	**22.3**	**(33.0)**
20	1	7.2	(8.2)	3.6	(3.8)	*3.3*	*(3.7)*	12.1	[illegible]	[illegible]	(6.5)	11.1	*	3.6	(5.6)	**3.0**	**(3.9)**
	10	34.9	(46.2)	18.4	(21.7)	17.1	(20.4)	[illegible]	[illegible]	27.1	(35.9)	63.4	*	*11.1*	*(18.6)*	**9.4**	**(15.1)**
	20	59.3	(76.8)	35.7	(41.5)	33.8	(39.7)	[illegible]	(45.5)	49.6	(57.2)	145.4	*	*17.8*	*(28.8)*	**15.2**	**(24.8)**
	30	78.8	(99.3)	52.9	(60.7)	50.8	(58.7)	59.8	(64.2)	68.6	(76.3)	200.5	*	*24.5*	*(37.5)*	**20.6**	**(32.3)**
30	1	7.2	(8.2)	4.0	(4.3)	*3.5*	*(3.9)*	19.7	(18.5)	5.9	(6.4)	11.7	*	3.6	(5.3)	**3.1**	**(4.2)**
	10	34.6	(45.3)	20.3	(23.7)	18.1	(21.5)	34.2	(33.4)	28.2	(34.8)	53.6	*	*11.3*	*(18.6)*	**9.7**	**(16.2)**
	20	59.2	(76.0)	38.0	(44.3)	35.1	(41.1)	50.8	(51.7)	49.4	(59.1)	193.3	*	*18.0*	*(29.1)*	**15.5**	**(25.9)**
	30	78.8	(99.1)	55.4	(65.0)	52.0	(60.8)	67.2	(71.1)	65.1	(81.4)	247.5	*	*24.6*	*(37.9)*	**20.8**	**(33.3)**

Up to x 2.75 times more accurate

* Indicates at least one more order of magnitude

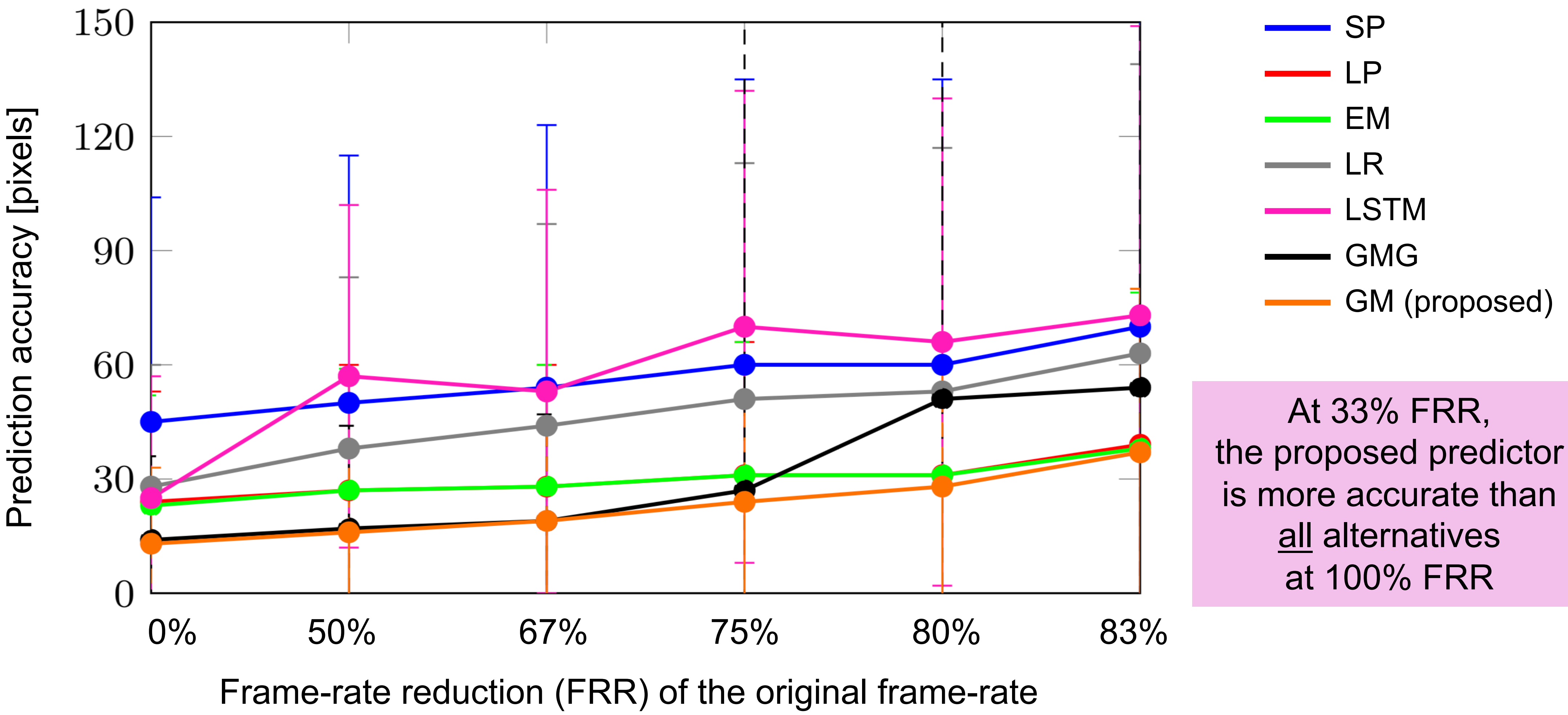
Robustness to noisy observations

- Gaussian noise introduced to observed past locations
- Prediction accuracy on all objects and for all frames of the test dataset



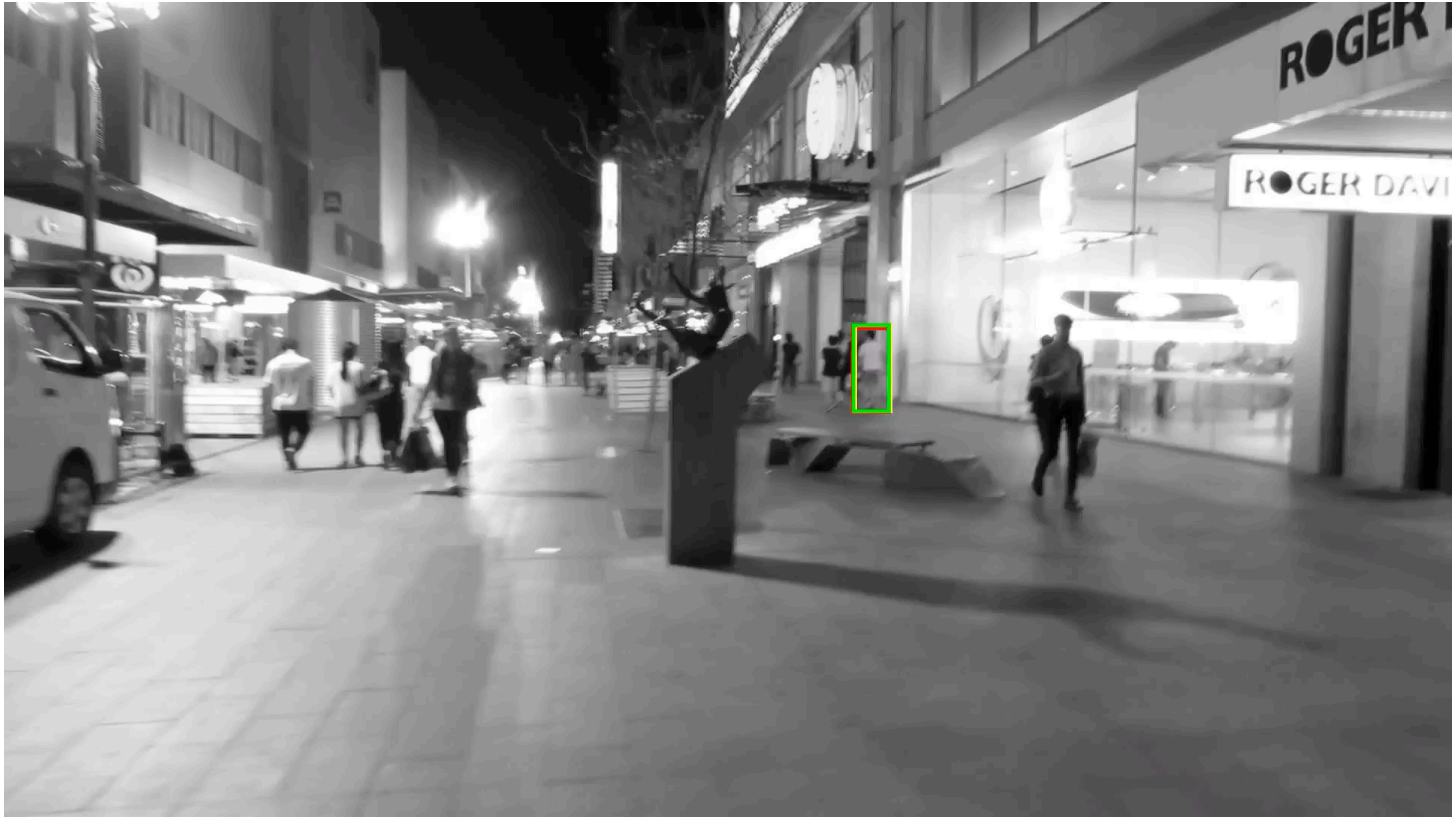
Robustness to frame-rate reduction

- Prediction accuracy on all objects and for all frames of the test dataset



At 33% FRR,
the proposed predictor
is more accurate than
all alternatives
at 100% FRR

Example



Manual annotation

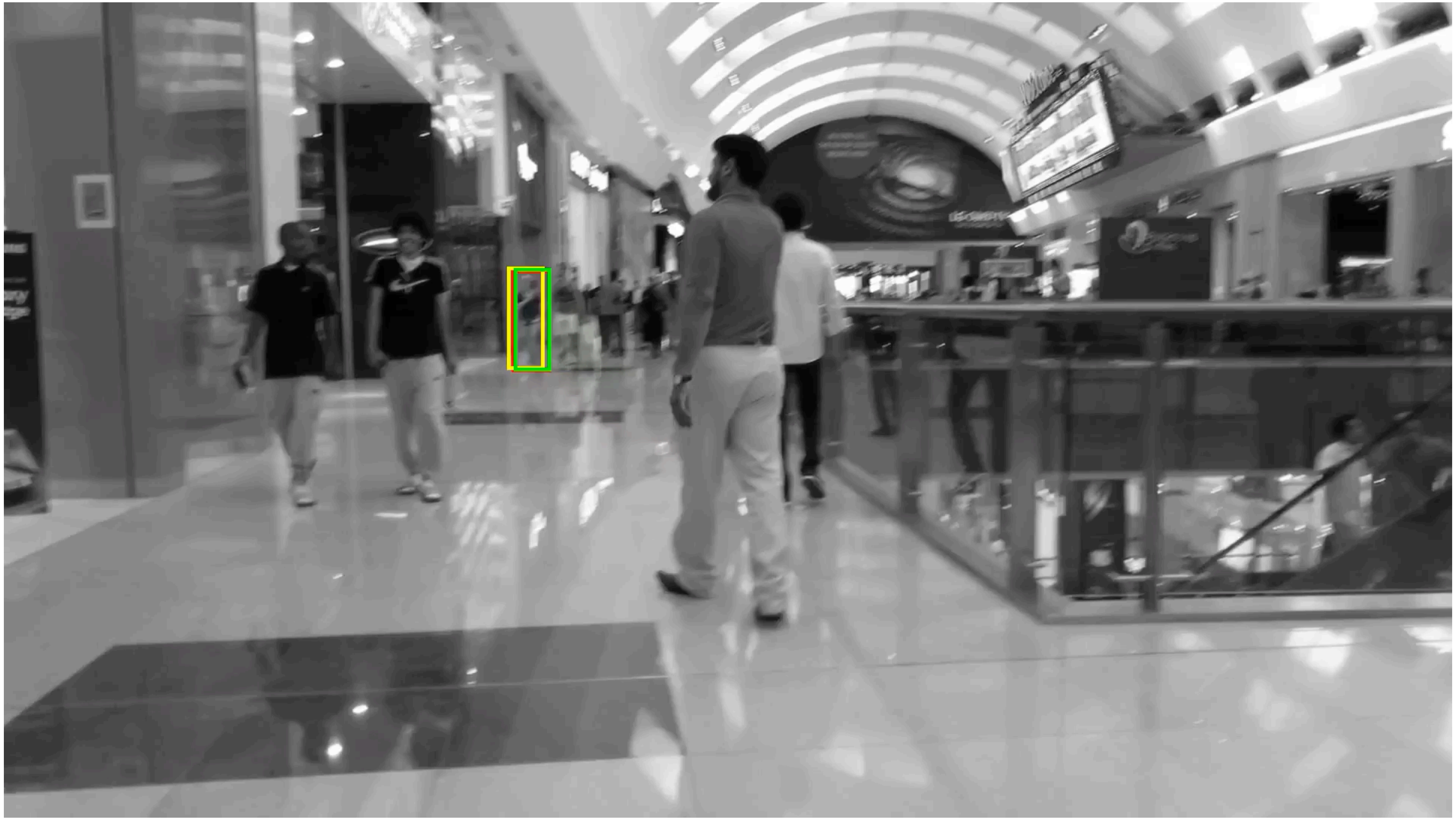
Linear prediction

Proposed

$T_P = 10$ past observed frames

$T_F = 60$ future predicted frames

Example



Manual annotation

Linear prediction

Proposed

$T_P = 10$ past observed frames

$T_F = 60$ future predicted frames

Example of error (occlusions)

Manual annotation

Linear prediction

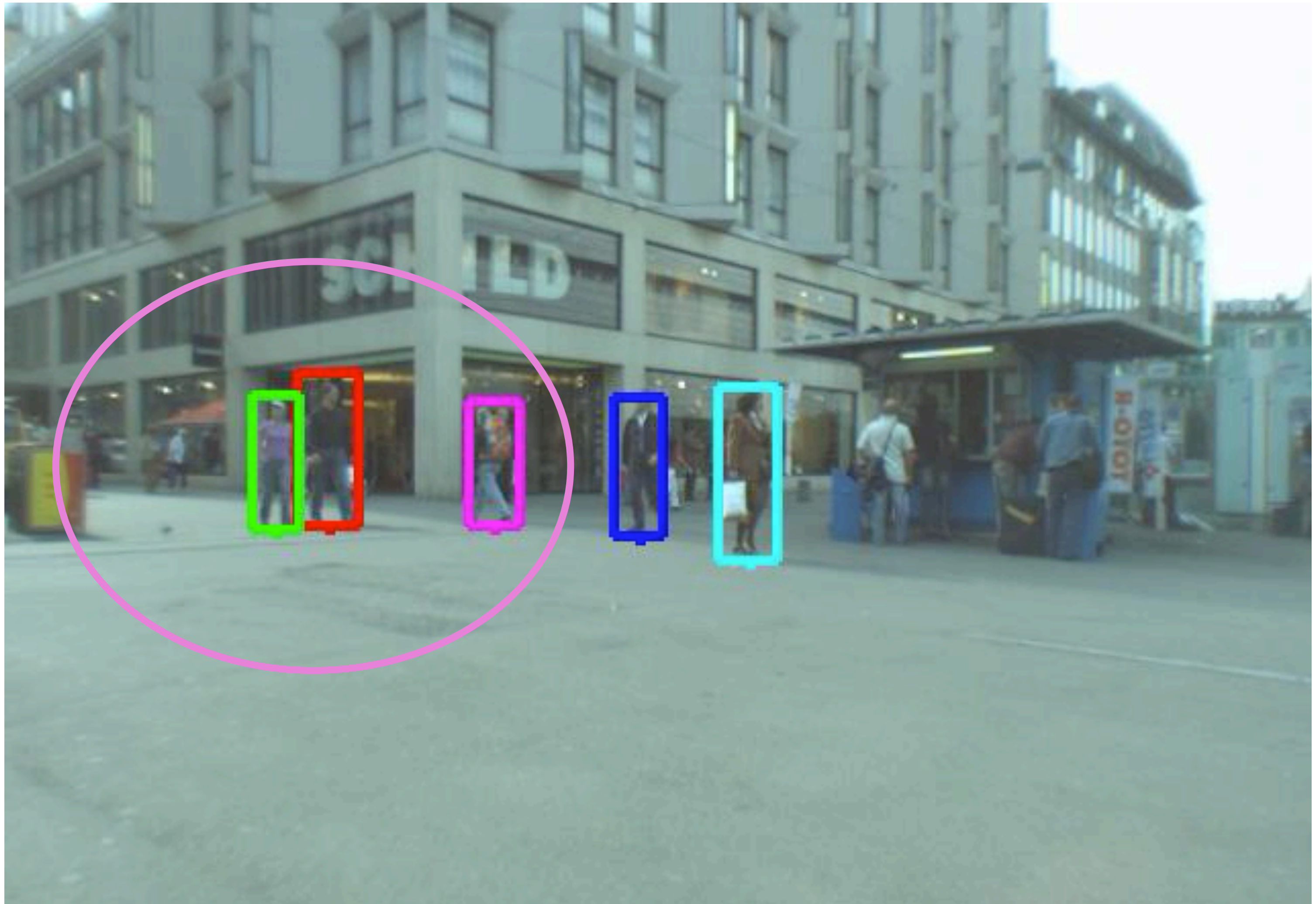
Proposed

$T_P = 10$ past observed frames

$T_F = 60$ future predicted frames

Tracking with linear motion prediction

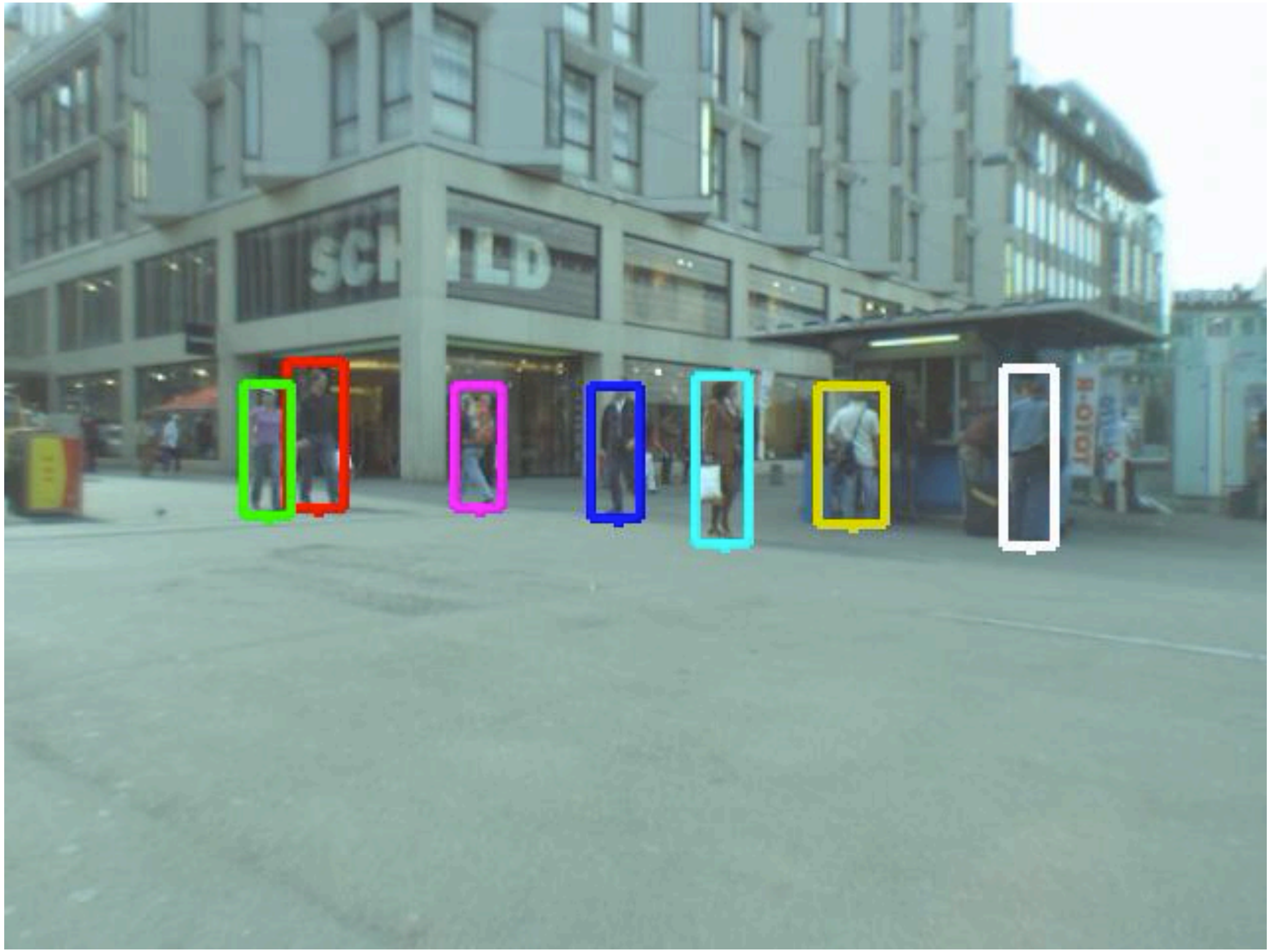
False-positive initialisations



R. Sanchez-Matilla et al., *Online multi-target tracking with strong and weak detections*. In Proc. ECCVw. 2016

Note
Tracking does not use appearance features

Sample of tracking with proposed motion prediction



Note
Tracking does not use appearance features

Conclusions

- Simple object motion predictor that
  - requires no camera calibration, scene nor object location assumptions
  - is 3x more accurate than LSTM and 2x more accurate than linear prediction
  - is as accurate as linear prediction when processing only 1/3 of the frames
  - is real-time
  - improves tracking results
- Future work
  - integration on a moving platform (robot) for navigation
  - energy consumption analysis