FaceHugger: The ALIEN Tracker Applied to Faces

DEMO

VISUAL OBJECT TRACKING

Tracking arbitrary objects in unconstrained environment by simply specifying a *single (one-shot) training example* at run-time.

CHALLENGES

Illumination, shadow, occlusion, viewpoint, clutter, translucent reflective material, camera sensor noise, motion blur, sensor quantization.

THE KEY IDEA

Our method has been inspired by studying the effects of Scale Invariant Feature Transform (SIFT) when applied to objects assumed to be flat even though they aren't. We argue that positive performance is intrinsically in the multiview local appearance representation.

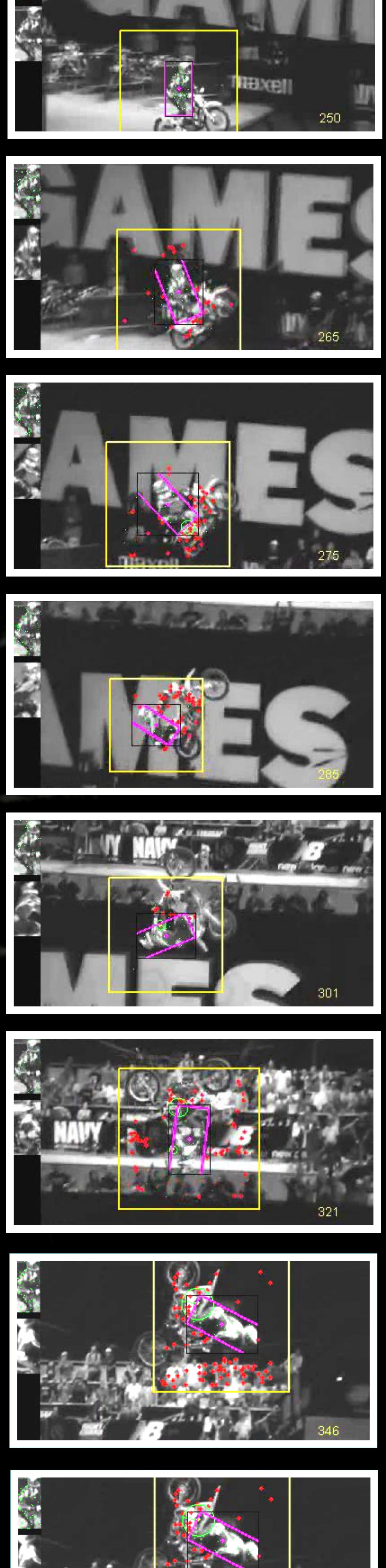
HOW AND WHY IT WORKS

Deviations from flatness induce nuisance factors that act on the feature representation in a manner for which no general local invariants can be computed. Hence deviations are (over)-represented through multiple instances of the same features.

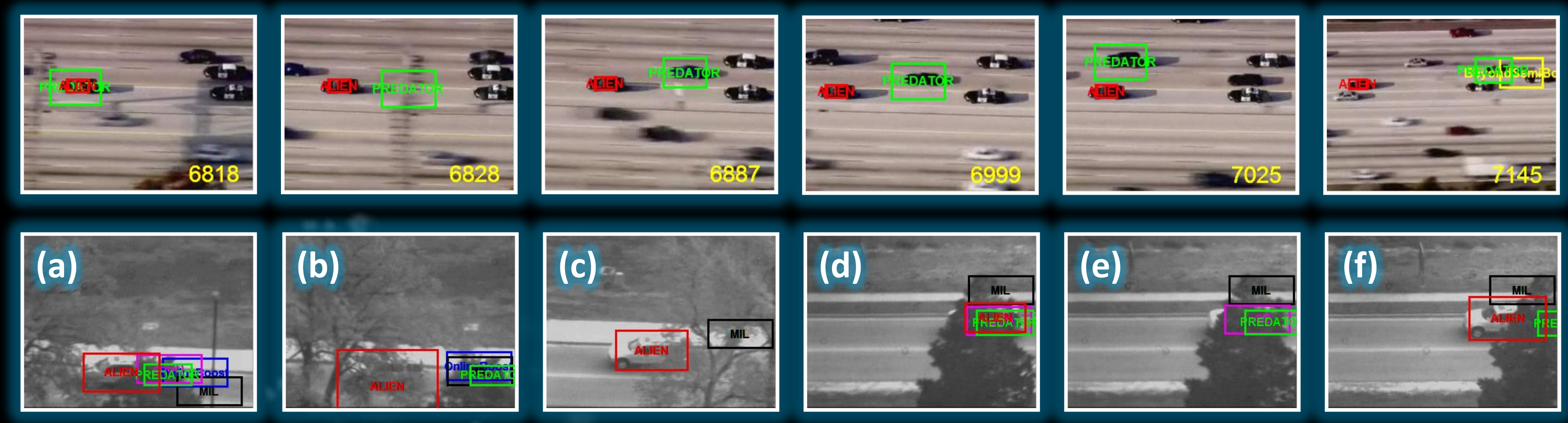
DOWNLOAD

http://www.micc.unifi.it/pernici/

Federico Pernici & Alberto Del Bimbo
MICC – Media Integration and Communication Center
University of Florence Italy

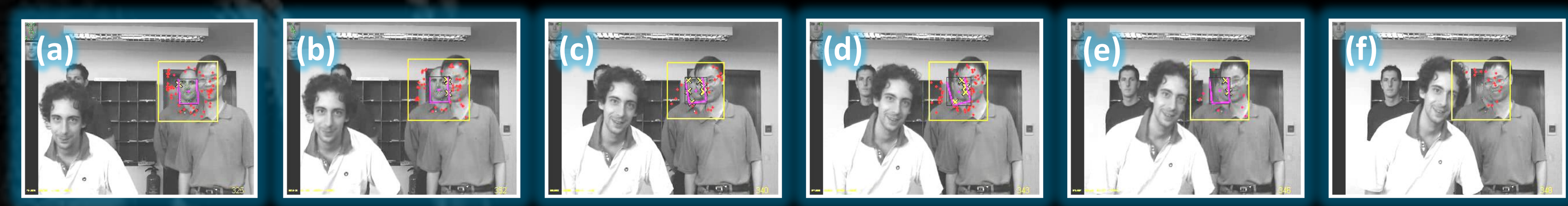


ALIEN vs PREDATOR



Sequence	Frames	OB [46]	SB [53]	BS [54]	MIL [16]	CoGD[27]	TLD [7]	ALIEN
David	761	0.01 / 0.01 / 0.01	0.27 / 0.27 / 0.27	0.16 / 0.12 / 0.13	0.06 / 0.06 / 0.06	0.99 / 0.99 / 0.99	**1.00 / 1.00 / 1.00**	*0.99 / 0.98 / 0.99*
Jumping	313	0.41 / 0.04 / 0.08	0.14 / 0.08 / 0.10	0.06 / 0.05 / 0.05	0.37 / 0.37 / 0.37	**1.00 / 0.99 / 1.00**	*0.99 / 0.99 / 0.99*	0.99 / 0.87 / 0.92
Pedestrian 1	140	0.36 / 0.09 / 0.14	0.20 / 0.14 / 0.16	0.10 / 0.04 / 0.05	0.42 / 0.42 / 0.42	0.99 / 0.99 / 0.99	**1.00 / 1.00 / 1.00**	**1.00 / 1.00 / 1.00**
Pedestrian 2	338	0.74 / 0.12 / 0.21	0.55 / 0.46 / 0.50	1.00 / 0.02 / 0.04	0.10 / 0.12 / 0.11	0.71 / 0.90 / 0.79	*0.89 / 0.92 / 0.91*	**0.93 / 0.92 / 0.93**
Pedestrian 3	184	1.00 / 0.33 / 0.49	0.41 / 0.33 / 0.36	0.81 / 0.40 / 0.54	0.49 / 0.58 / 0.53	0.84 / 0.99 / 0.91	**0.99 / 1.00 / 0.99**	*1.00 / 0.90 / 0.95*
Car	945	0.89 / 0.57 / 0.69	1.00 / 0.67 / 0.80	0.99 / 0.56 / 0.72	0.11 / 0.12 / 0.11	0.91 / 0.92 / 0.91	*0.92 / 0.97 / 0.94*	**0.95 / 1.00 / 0.98**
Motocross	2665	0.13 / 0.00 / 0.00	0.01 / 0.00 / 0.00	0.14 / 0.00 / 0.00	0.02 / 0.01 / 0.01	0.80 / 0.26 / 0.39	**0.67 / 0.58 / 0.62**	*0.49 / 0.58 / 0.54*
Volkswagen	8576	0.04 / 0.00 / 0.00	0.00 / 0.00 / 0.00	0.00 / 0.00 / 0.00	0.26 / 0.03 / 0.05	0.41 / 0.03 / 0.06	*0.54 / 0.64 / 0.59*	**0.99 / 0.70 / 0.82**
Carchase	9928	0.73 / 0.03 / 0.05	0.79 / 0.04 / 0.08	0.38 / 0.09 / 0.14	0.49 / 0.03 / 0.05	0.87 / 0.04 / 0.08	*0.50 / 0.40 / 0.45*	**0.73 / 0.68 / 0.70**
mean	-	0.40 / 0.10 / 0.15	0.30 / 0.18 / 0.20	0.33 / 0.11 / 0.15	0.21 / 0.15 / 0.15	0.68 / 0.55 / 0.55	*0.68 / 0.68 / 0.68*	**0.73 / 0.69 / 0.71**

FaceHugger



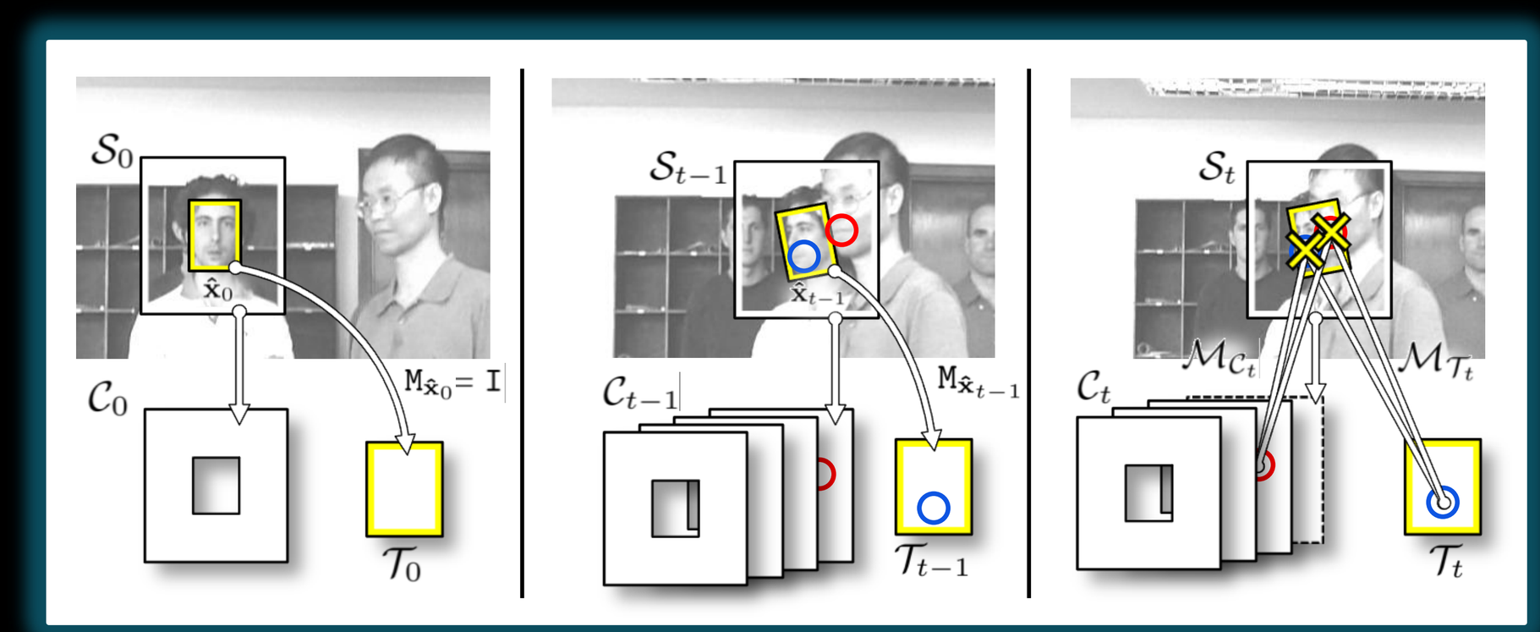
Object Representation

Weakly Aligned Multi-instance Local Features

For this motivation we call our method ALIEN, **A**ppearance **L**earning **I**n **E**vidential **N**uisance, since it is based on the physical observation that if the object is reasonably convex, known critical nuisance factors which **cannot** be neutralized, can be managed based on multiple instances of features selected and updated according to a weak global shape model.

Appearance Learning

Object/Context Appearance Update



- Additional information describing the arrangement object/context features is specified according to the transitive property: $$(\mathcal{T}_t \sim \mathcal{S}_t) \wedge (\mathcal{C}_t \sim \mathcal{S}_t) \Longrightarrow \mathcal{T}_t \sim \mathcal{C}_t$$
- Features are removed performing set-wise difference between the matching indexes: $\mathcal{F}_t = \mathcal{M}_{\mathcal{T}_t} \setminus \mathcal{M}_{\mathcal{C}_t}$

Occlusion Detection

To prevent an improper update, occlusion is detected before updating the template.

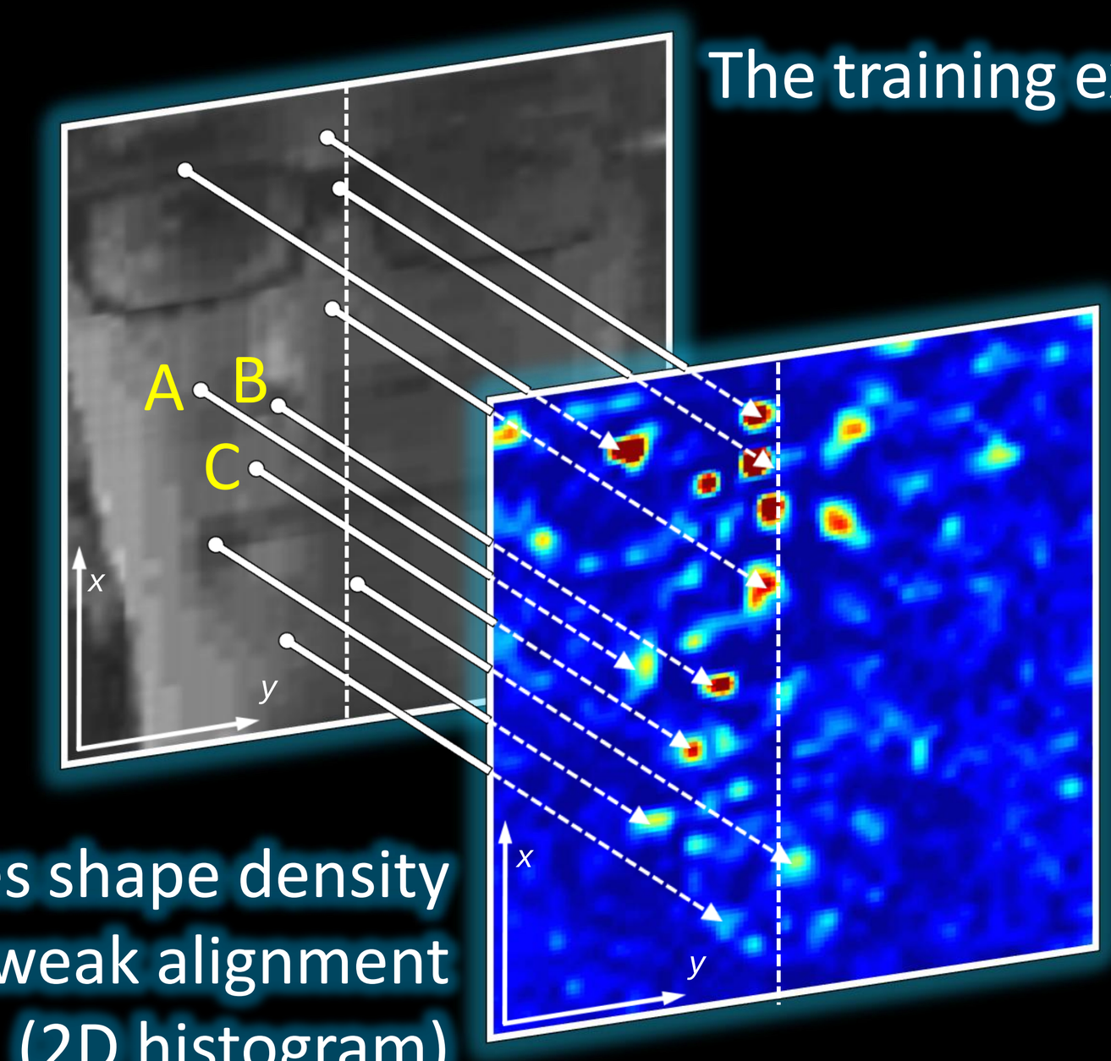
- The space time context is used to intercept possible occluders: the tracker is vulnerable to failures when the appearance of an occluders persistently added to the object template

$$\mathcal{O}_t = \{(\mathbf{p}, \mathbf{d}) \in \mathcal{M}_{\mathcal{C}_t} \mid \mathbf{p} \in \mathrm{OBB}(\hat{\mathbf{x}}_t)\} = \{\times\}$$

- The features in $\mathcal{O}_t$ may originate from:
  - object/context ambiguous features,
  - object/context boundary features,
  - features belonging to occluding objects.
- Hence occlusion is detected when:

$$|\mathcal{O}_t| \geq N_{\mathcal{O}}$$

The training example



Features shape density after weak alignment (2D histogram)