Visual Tracking by Sampling Tree-Structured Graphical Models

Seunghoon Hong Bohyung Han

Computer Vision Lab.
Dept. of Computer Science and Engineering
POSTECH

POSTECH

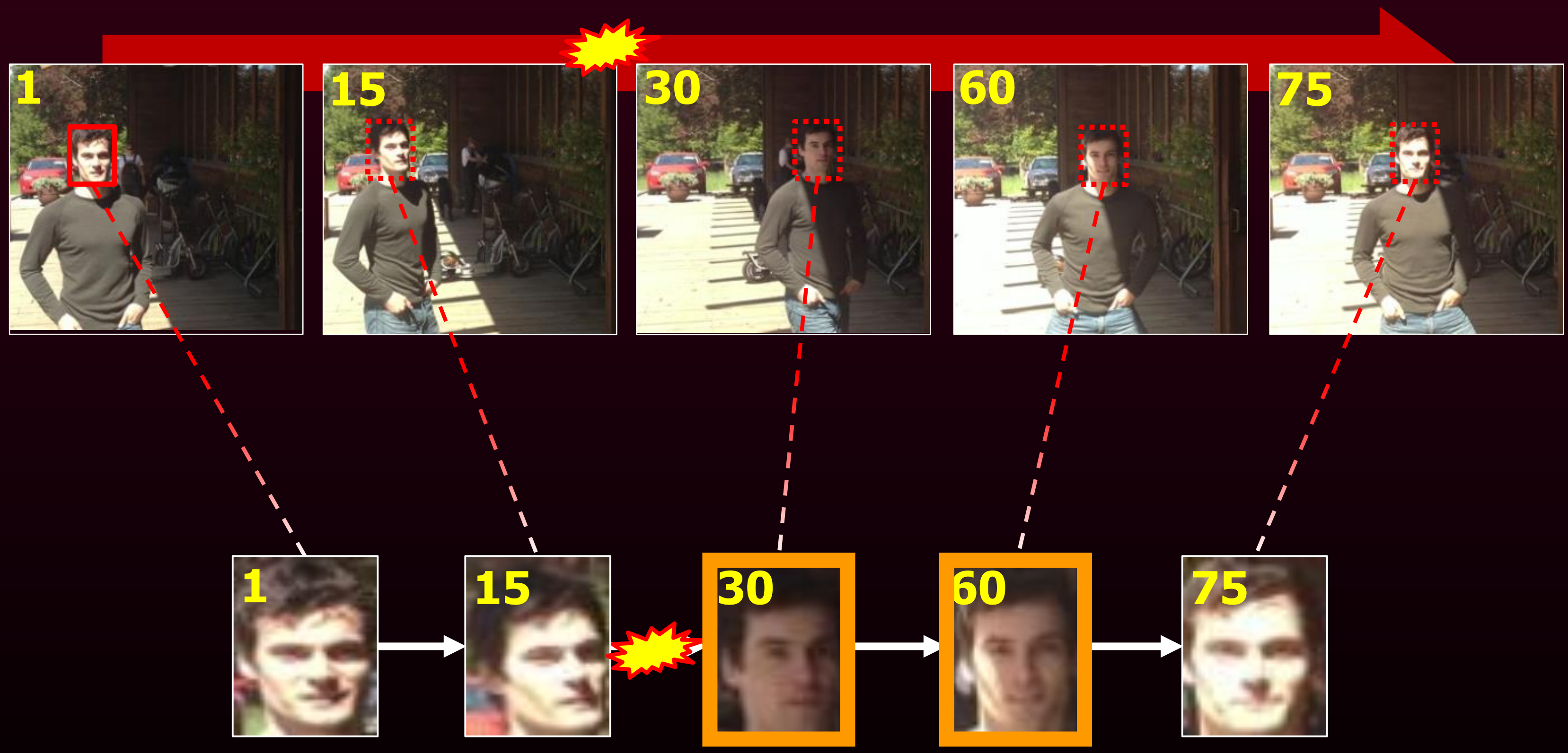
ECCV '14

Goal of Visual Tracking

- Robust estimation of accurate target states throughout an input video

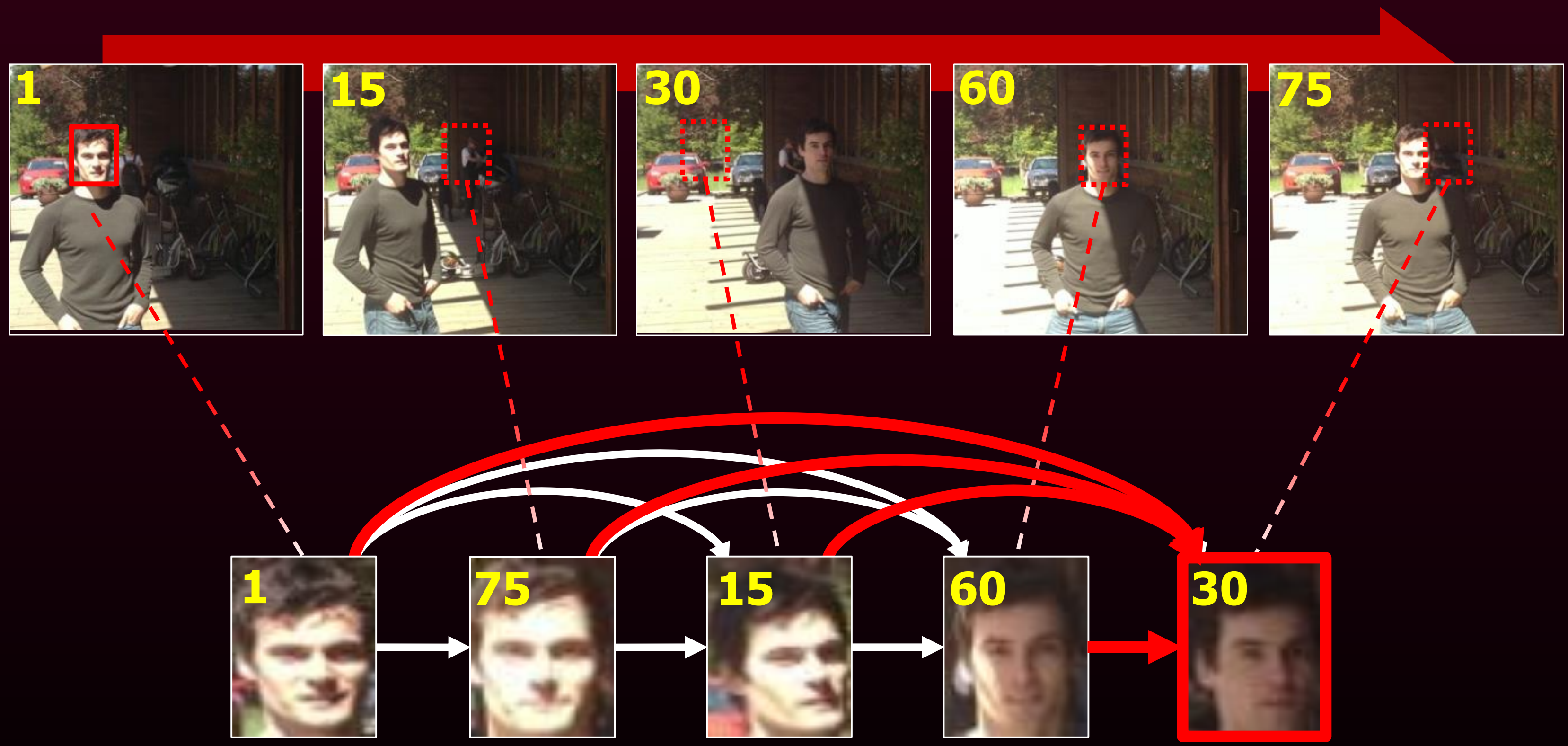
Conventional Tracking Approaches

- Sequential tracking based on chain model



Orderless Tracking[HongICCV2013]

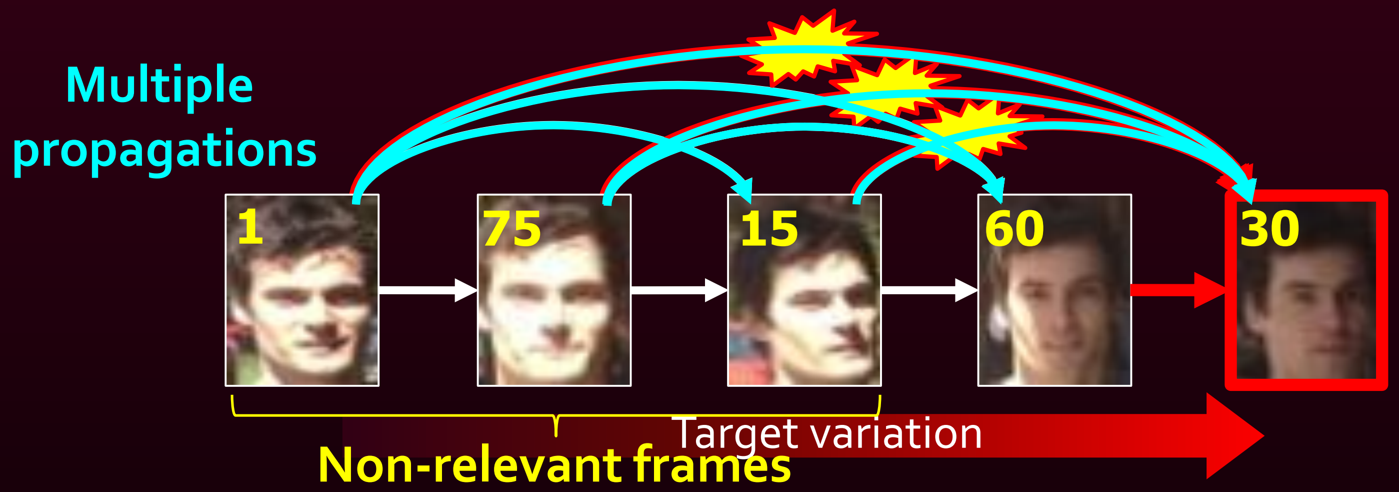
- Orderless tracking based on Bayesian model averaging



[HongICCV2013] S. Hong, S. Kwak and B. Han. Orderless Tracking through Model-Averaged Posterior Estimation. In ICCV, 2013

Orderless Tracking[HongICCV2013]

- Orderless tracking based on Bayesian model averaging



Advantages:

- Robust to error propagation

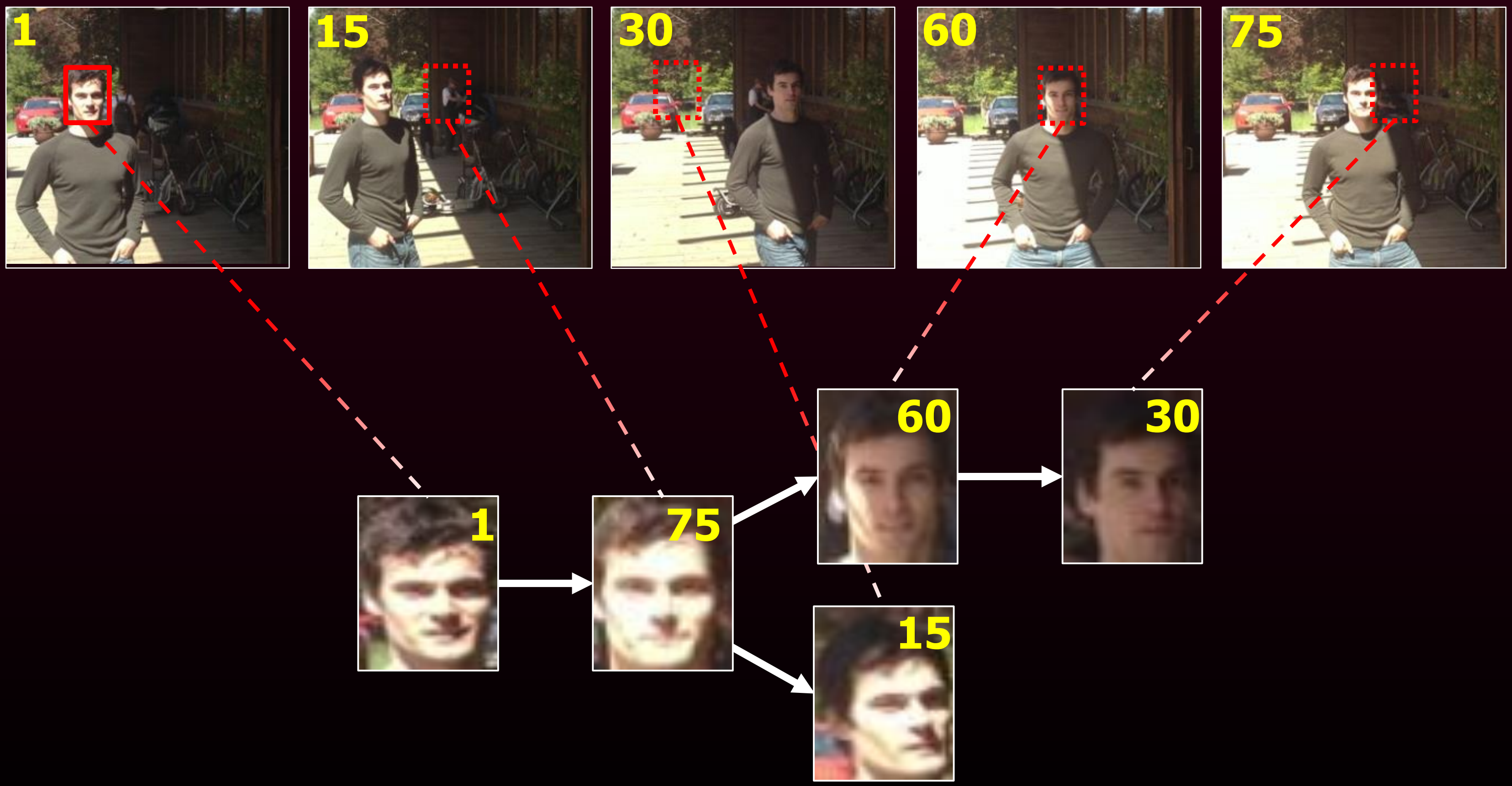
Limitation:

- Ineffective to handle multi-modal variation

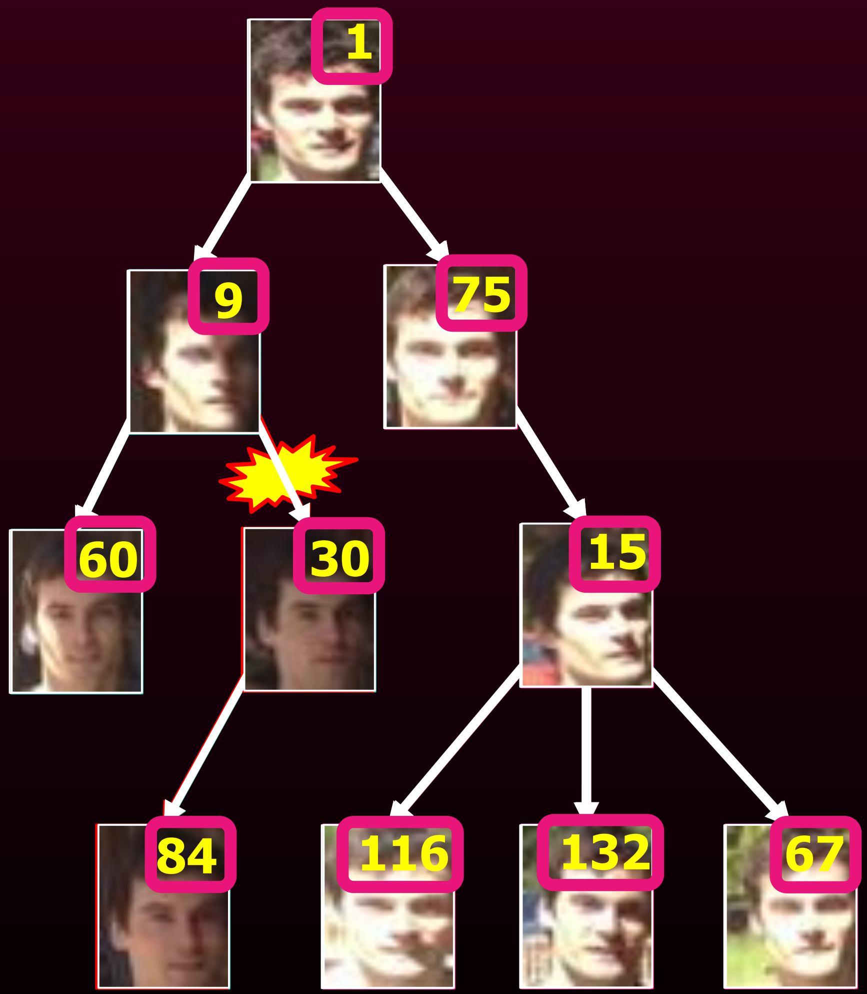
[HongICCV2013] S. Hong, S. Kwak and B. Han. Orderless Tracking through Model-Averaged Posterior Estimation. In ICCV, 2013

Our Approach

- Tracking on tree-structured graphical model



Tracking on Tree-Structure



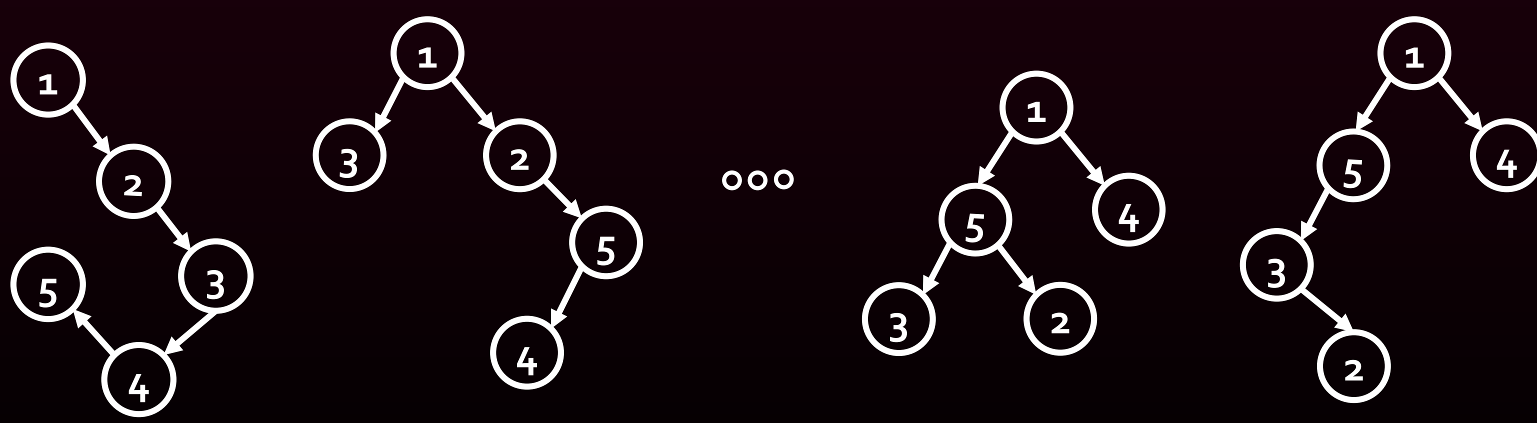
- Advantages

1. Multi-modality is handled by independent branch
2. Failures are isolated at local branch
3. Frames are ordered based on tracking difficulty

Challenges

- Tree learning and tracking are mutually dependent

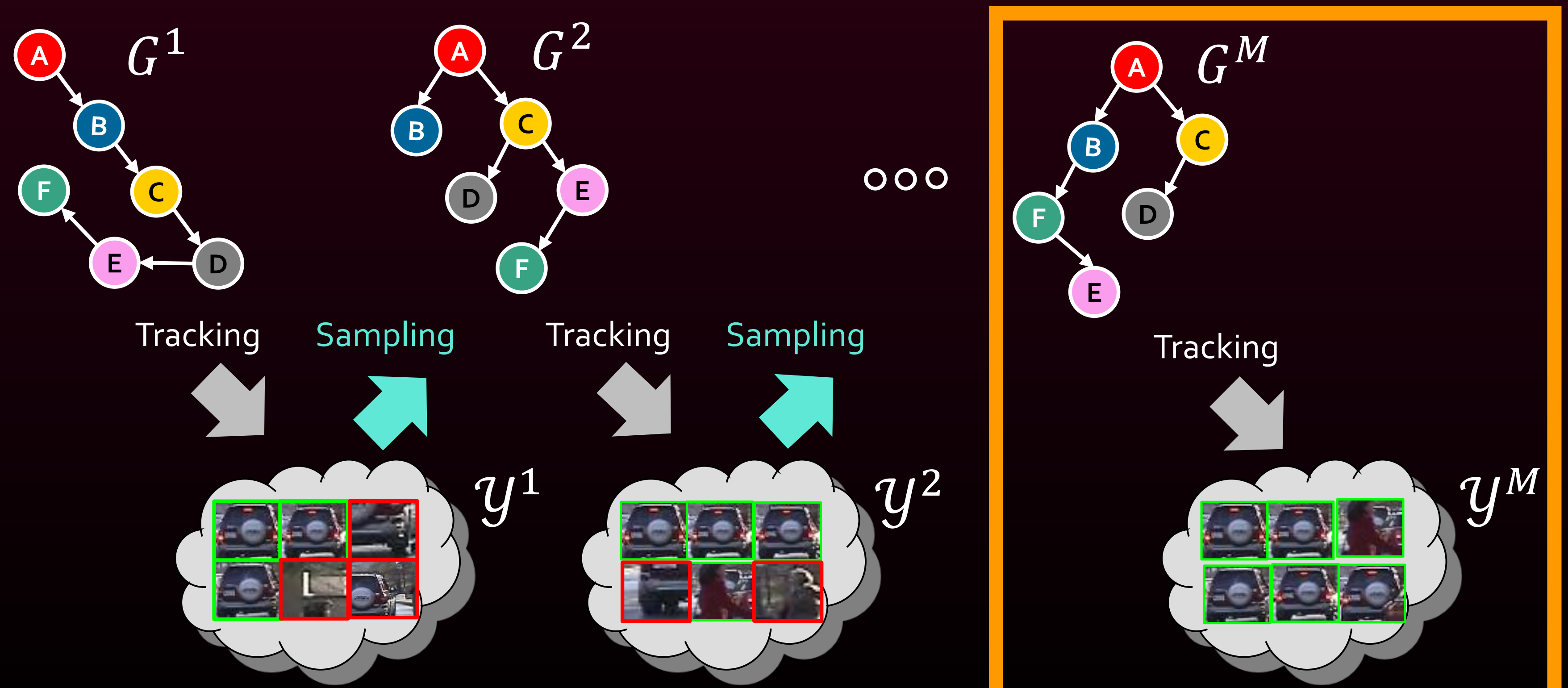
Which tree is good for tracking?



Our Approach

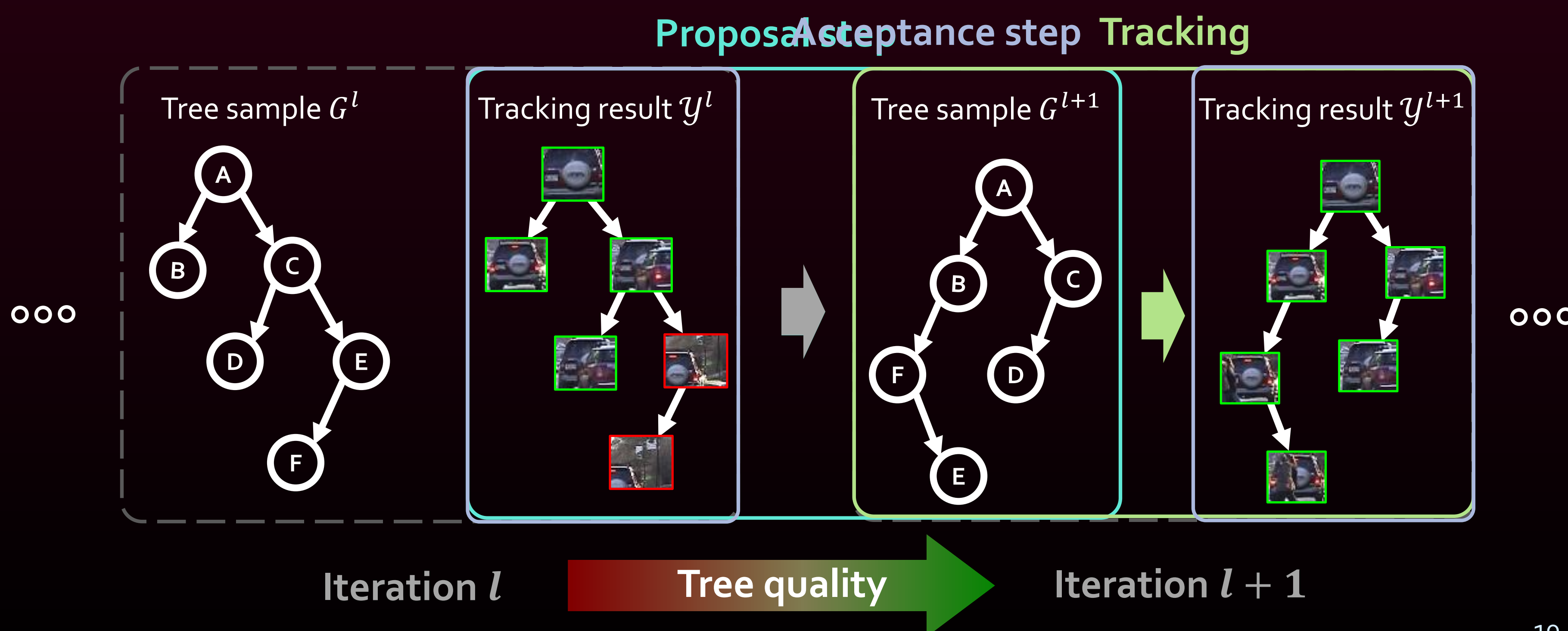
- Joint tree learning and tracking based on sampling

$$\hat{G} = \underset{G^l}{\operatorname{argmin}} - \log p(\mathcal{Y}^l | G^l), \qquad l = 1, \ldots, M$$



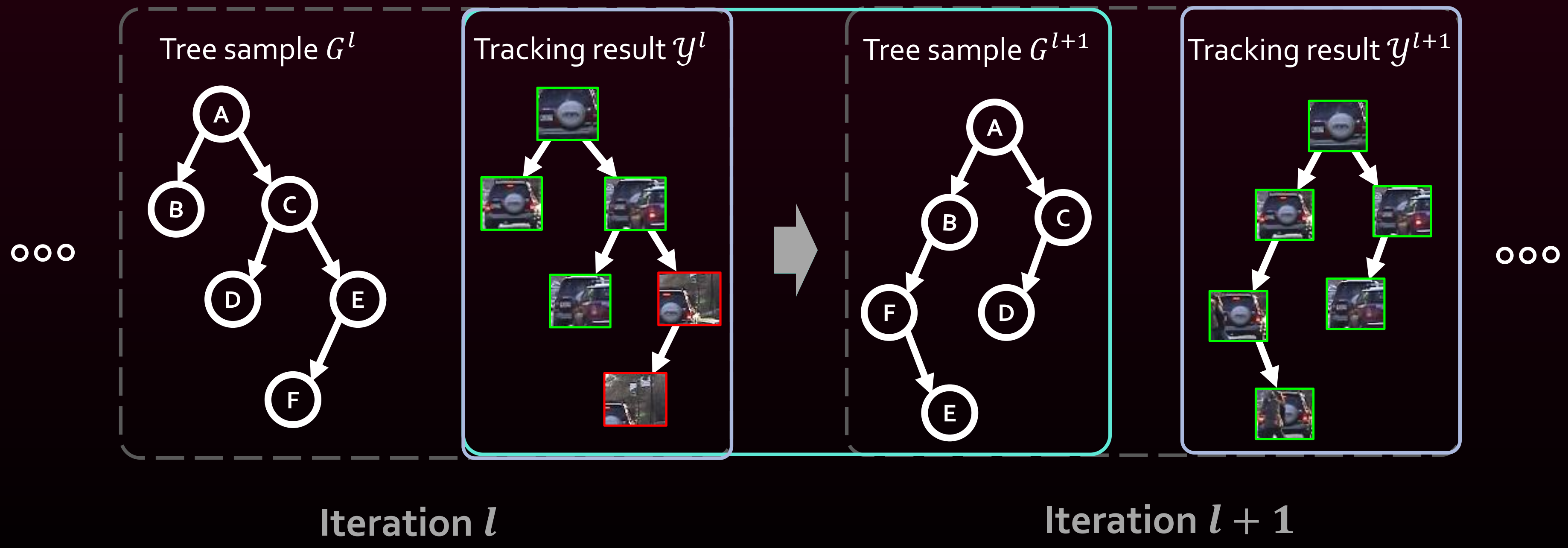
Sampling Tree Structure by MCMC

- Optimization by MCMC sampling
  - ***Propose*** a new sample by proposal distribution $Q(G^{l+1}; G^{l})$
  - ***Accept*** a new sample by acceptance ratio α



Challenges

- How to *propose a better tree*?
- How to *measure the quality of tree for tracking*?



Proposing A New Tree

- Single edge delete/add operation

$$Q\left(G^{l+1}; G^{l}\right) = P_{delete} * P_{add}$$

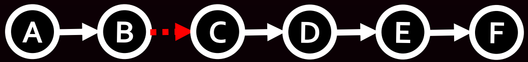
$$P_{delete}(i,j) = \frac{\exp\left(\boxed{d(i,j)}\right)}{\sum_{(a,b)\in\mathcal{E}} \exp\left(d(a,b)\right)}, \quad (i,j) \in \mathcal{E}$$

distance between target templates $d(i,j)$

$d(a,b)$ $d(b,c)$

y_a y_b y_c

$d(a,b) \ll d(b,c)$



(a) edge deletion

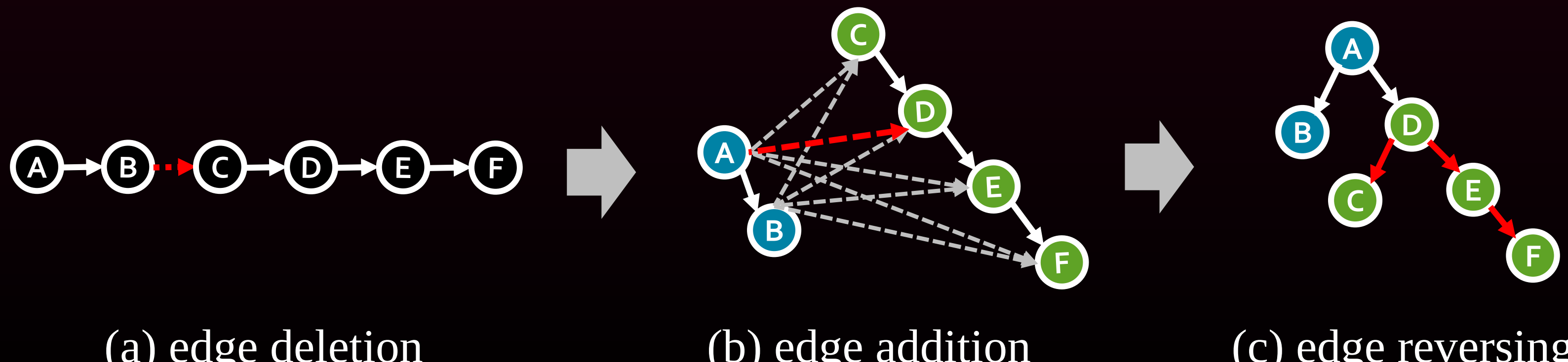
Proposing A New Tree

- Single edge delete/add operation

$$Q\left(G^{l+1}; G^{l}\right) = P_{delete} * P_{add}$$

$$P_{delete}(i,j) = \frac{\exp(d(i,j))}{\sum_{(a,b)\in\mathcal{E}} \exp\left(d(a,b)\right)}, \quad (i,j) \in \mathcal{E}$$

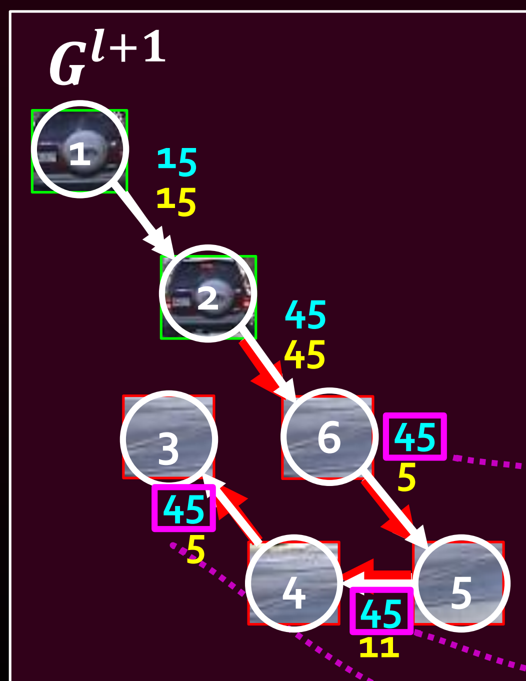
$$P_{add}(i,j) = \frac{\exp(-d(i,j))}{\sum_{(a,b)\in\bar{\mathcal{E}}} \exp\left(-d(a,b)\right)}, \quad (i,j) \in \bar{\mathcal{E}}$$



Validating A New Tree

- Accept a new tree structure
  - by measuring quality of the tree for tracking

$$-\log p(\mathcal{Y}^l|G^l) = \sum_i \boxed{c_i} = \sum_i \boxed{\max(d(i,p_i), c_{p_i})}$$



tracking cost
$d(i,j)$

Accumulated cost
$c_i = \max(d(i,p_i), c_{p_i})$

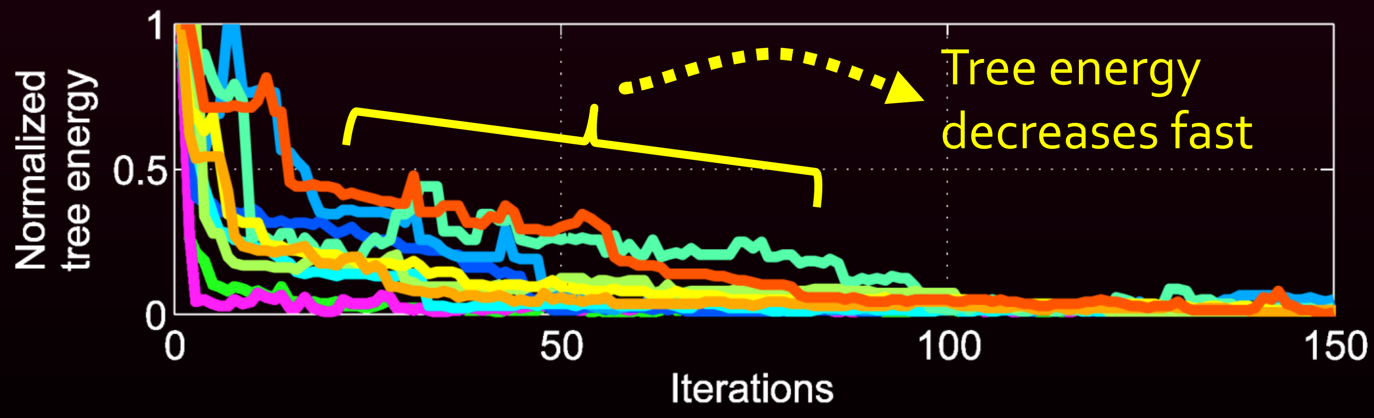
Error propagation
by accumulated cost

Validating A New Tree

- Accept a new tree structure
  - With an acceptance ratio α

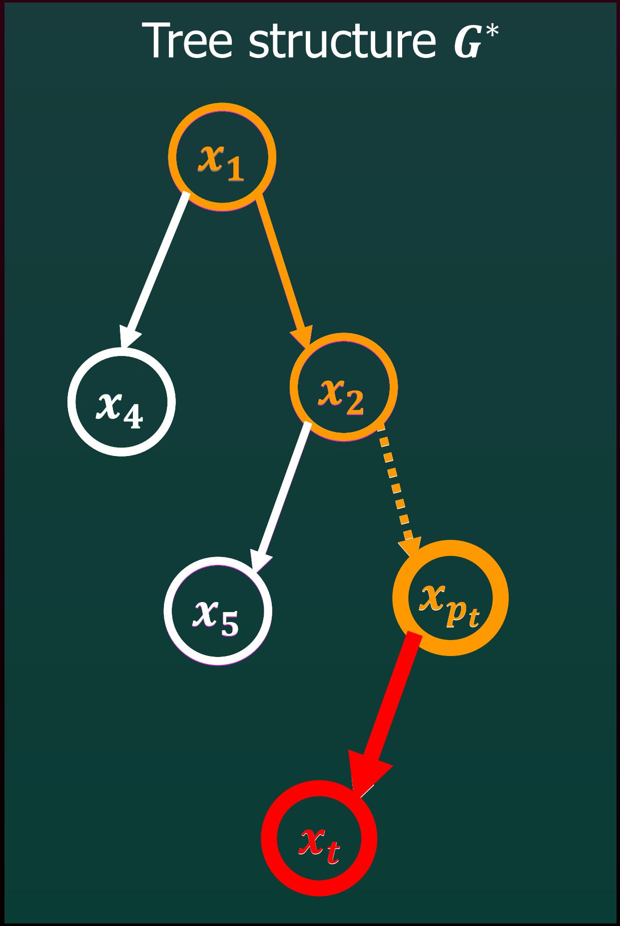
$$\alpha = \min\left[1, \quad \frac{\boxed{\left[-\log p(\mathcal{Y}^{\boldsymbol{l+1}}|G^{l+1})\right.}\left.\right]^{-1} Q(G^{l}; G^{l+1})}{\boxed{\left[\quad -\log p(\mathcal{Y}^{l}|G^{l})\right.}\left.\quad\right]^{-1} Q(G^{l+1}; G^{l})}\right]$$

  - Tree energy over MCMC iterations



Tracking on Tree Structure

- Density propagation in tree structure



Density propagation by sequential Bayesian filtering

$$p\left(x_t \middle| \{z_i\}_{i=1,\dots,t}, G\right)$$

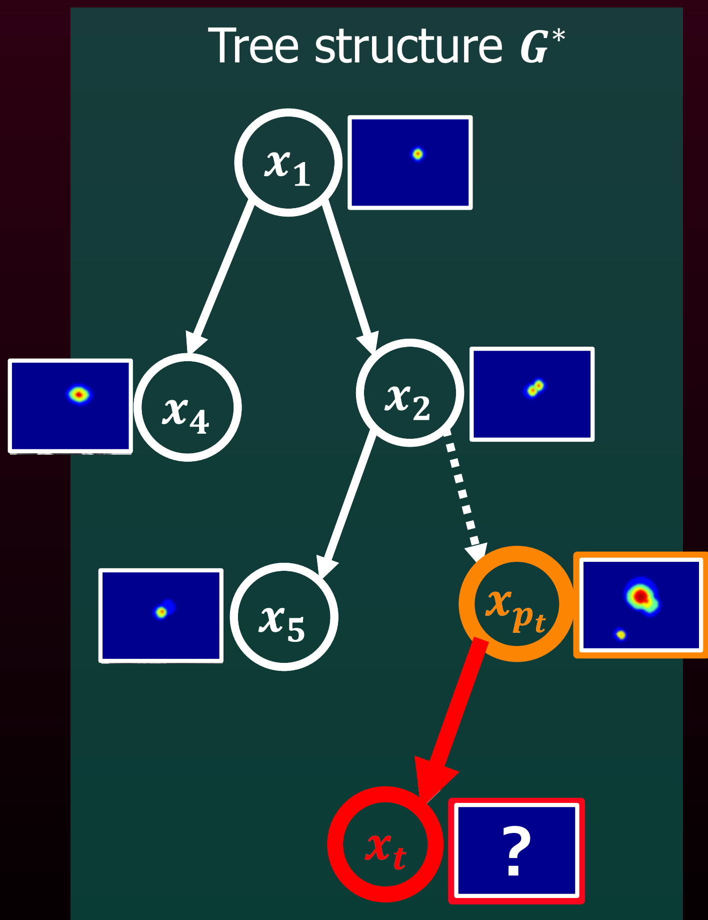
$$\propto \underbrace{p(z_t|x_t)}_{\text{Observation at current frame } t} \int_{x_{p_t}} \underbrace{p\left(x_t \middle| x_{p_t}\right) p\left(x_{p_t} \middle| \{z_i\}_{i=1,\dots,p_t}, G\right) dx_{p_t}}_{\text{Prediction from parent frame } p_t}$$

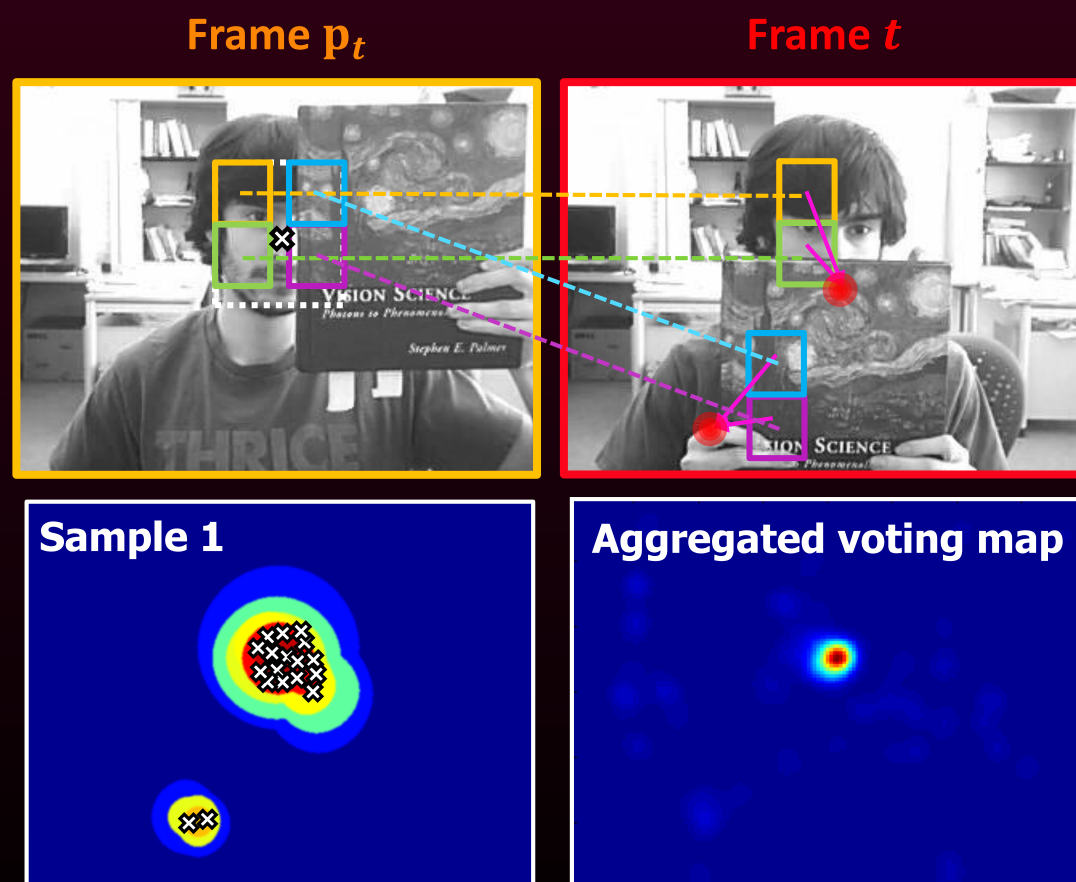
$$\approx \sum_{\mathrm{x}^{\mathrm{i}}_{p_\mathrm{t}} \in \mathbb{S}_{p_t}} \boxed{P(Z_t|x_t) P\left(x_t \middle| x^i_{p_t}\right)}$$

Patch matching and voting process[HongICCV2013]

$$\approx \sum_{x_{p_t}^{i} \in \mathbb{S}_{p_t}} \boxed{P(Z_t|x_t)P(x_t|x_{p_t}^{i})}$$

Patch matching and voting process[1],[2]

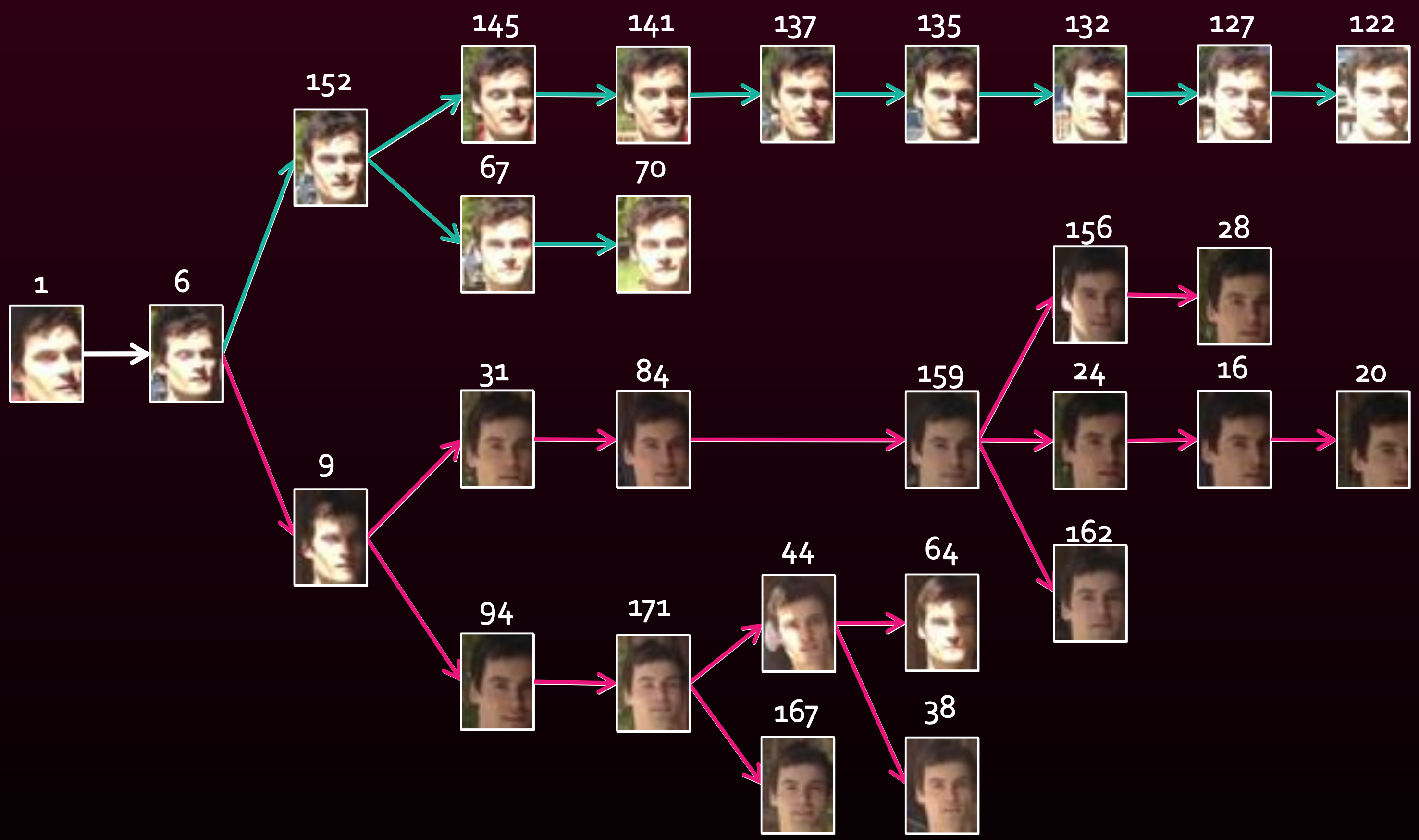




[1] S. Hong, S. Kwak and B. Han. Orderless Tracking through Model-Averaged Posterior Estimation. In ICCV, 2013
[2] S. Korman and S. Avidan. Coherency sensitive hashing. In ICCV, 2011

Identified Tree Structure

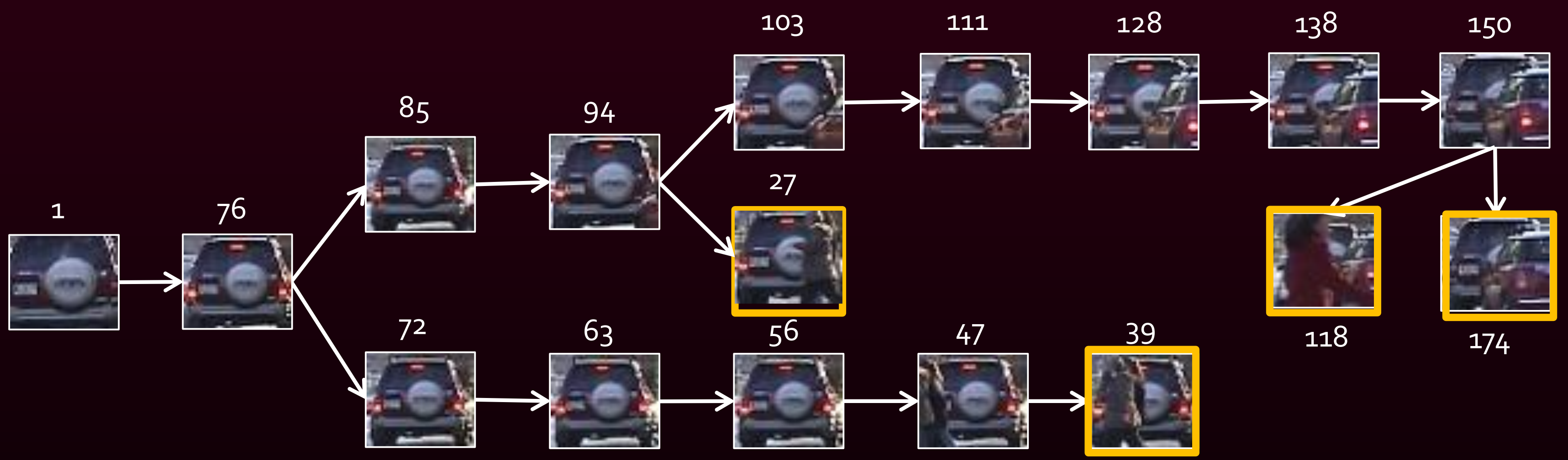
- Multi-modal appearance change



<*sunshade* sequence>

Identified Tree Structure

- Occlusion

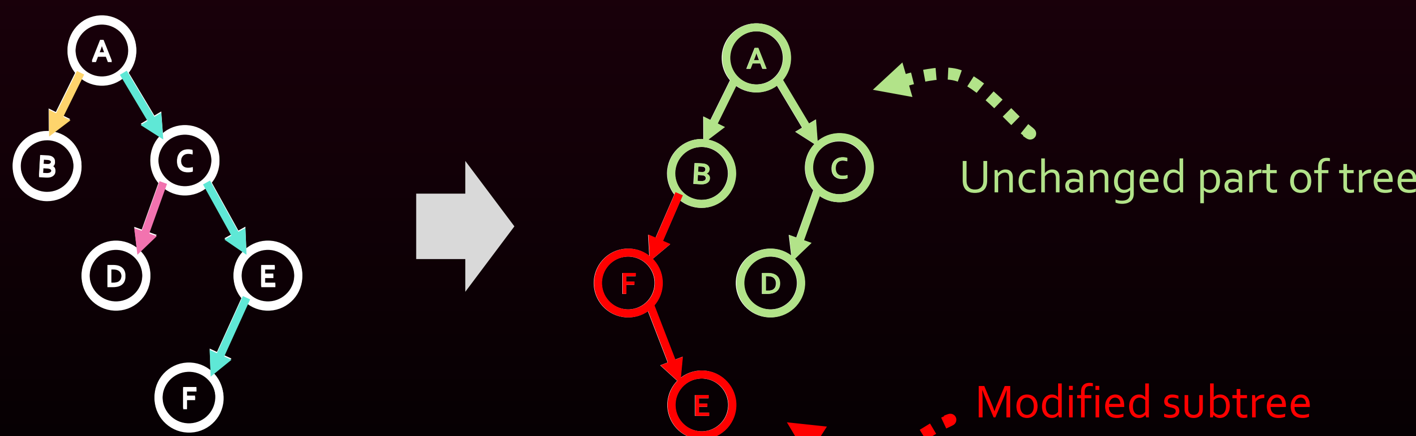


<*campus* sequence>

Computational Complexity

- Overall complexity : $O(MN)$
  - M : number of iterations
  - N : number of frames

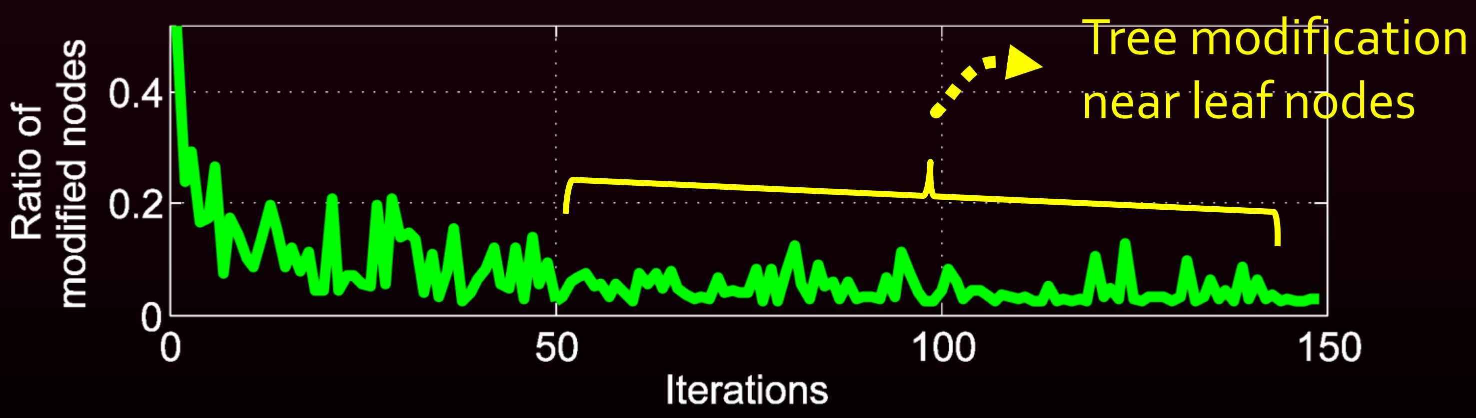
- Efficiency: posterior reusability



Computational Complexity

- Overall complexity : $O(MN)$
  - M : number of iterations
  - N : number of frames

- Efficiency: posterior reusability



Computational Complexity

- Overall complexity : $O(MN)$
  - M : number of iterations
  - N : number of frames

- Efficiency: posterior reusability

- We can further reduce a theoretical bound by a hierarchical approach : $O(kM + N - k)$

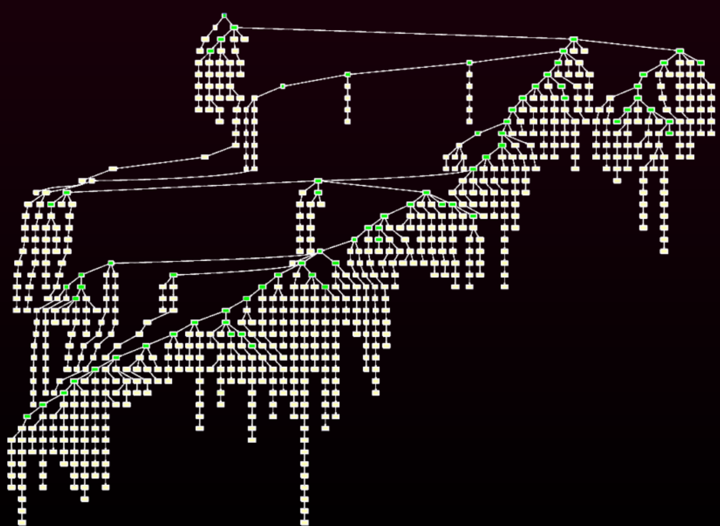
Hierarchical Approach

Original Video

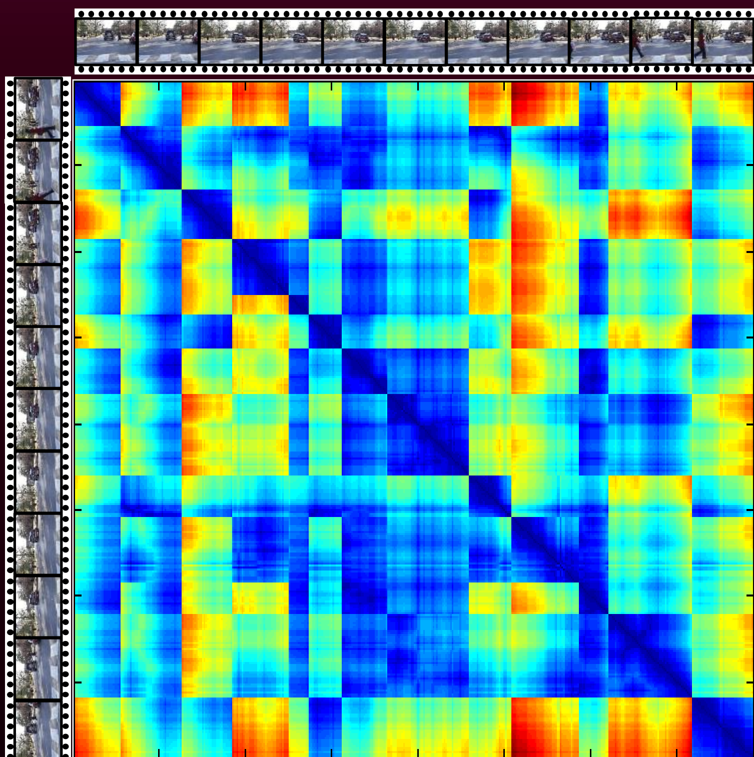
Key frame selection

Tree construction for key frames

Tree extension to entire video
and tracking on the tree



Key Frame Selection[HongICCV2013]

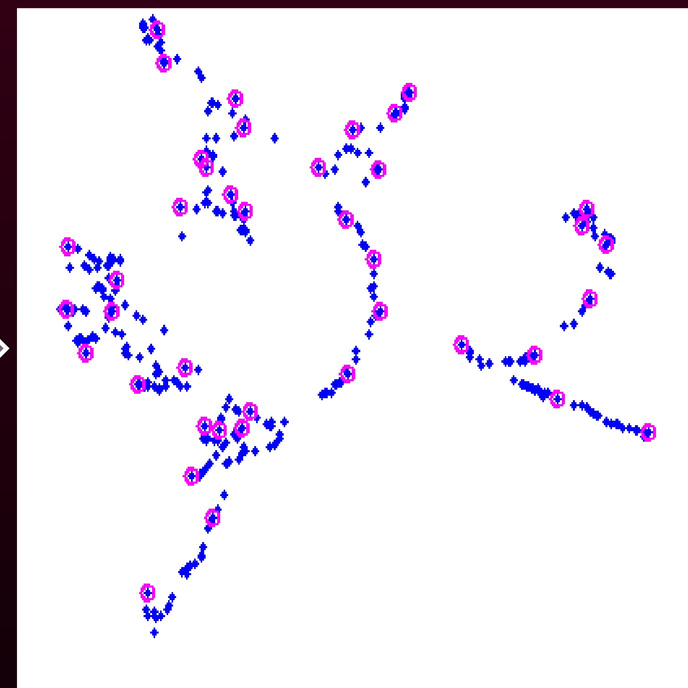


Dissimilarities between all pairs of frames

Geodesic distances from a sparse nearest neighbor graph

$$\frac{1}{n_1}\sum_{P\in I_1}\min_{Q\in I_2} d(P,Q)+\frac{1}{n_2}\sum_{Q\in I_2}\min_{P\in I_1} d(P,Q)$$

ISOMAP



Embedded frames

k most representative frames obtained by a *k-means clustering*

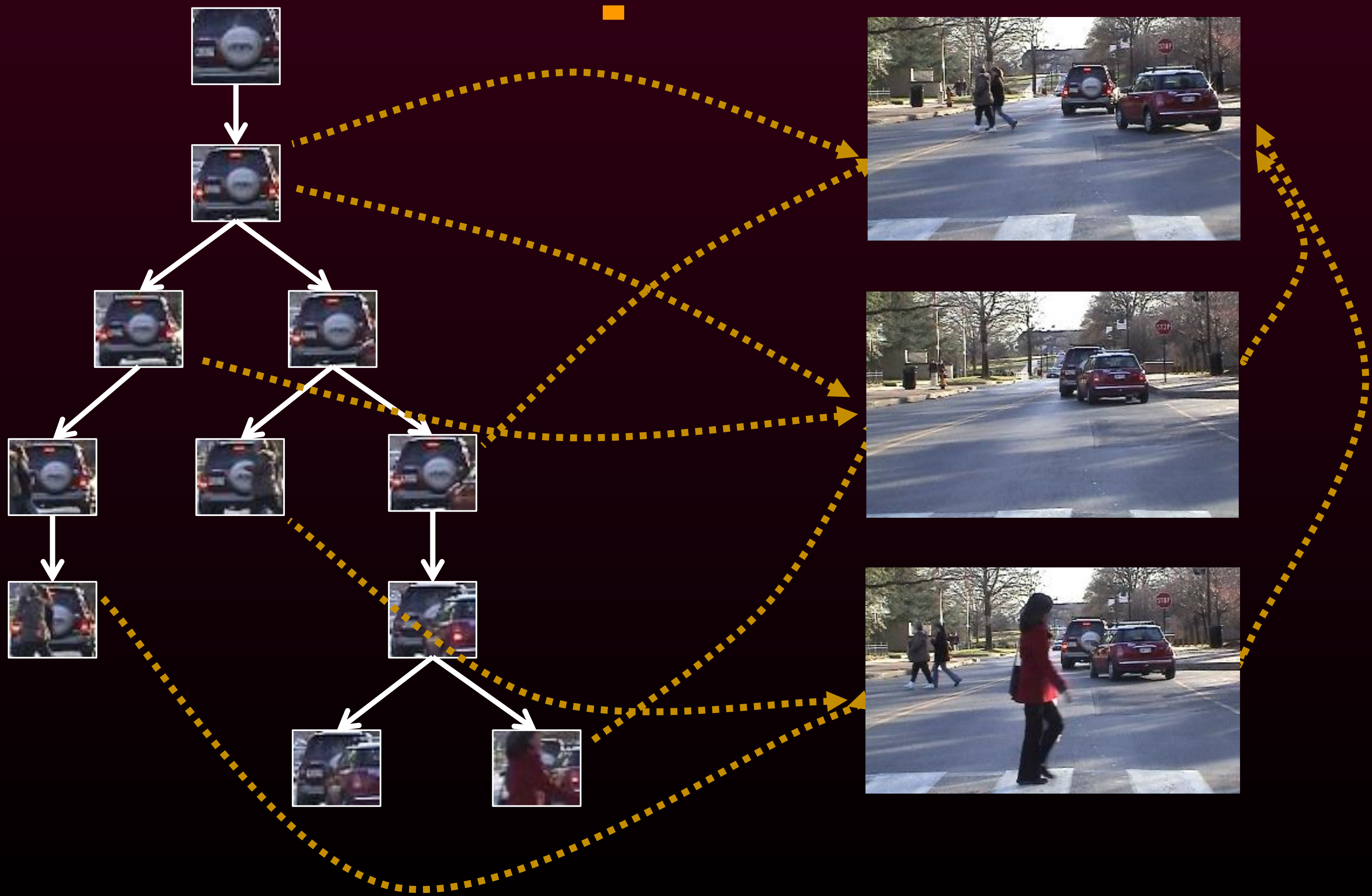
[HongICCV2013] S. Hong, S. Kwak and B. Han. Orderless Tracking through Model-Averaged Posterior Estimation. In ICCV, 2013

Tree Extension by Manifold Alignment

Key frames

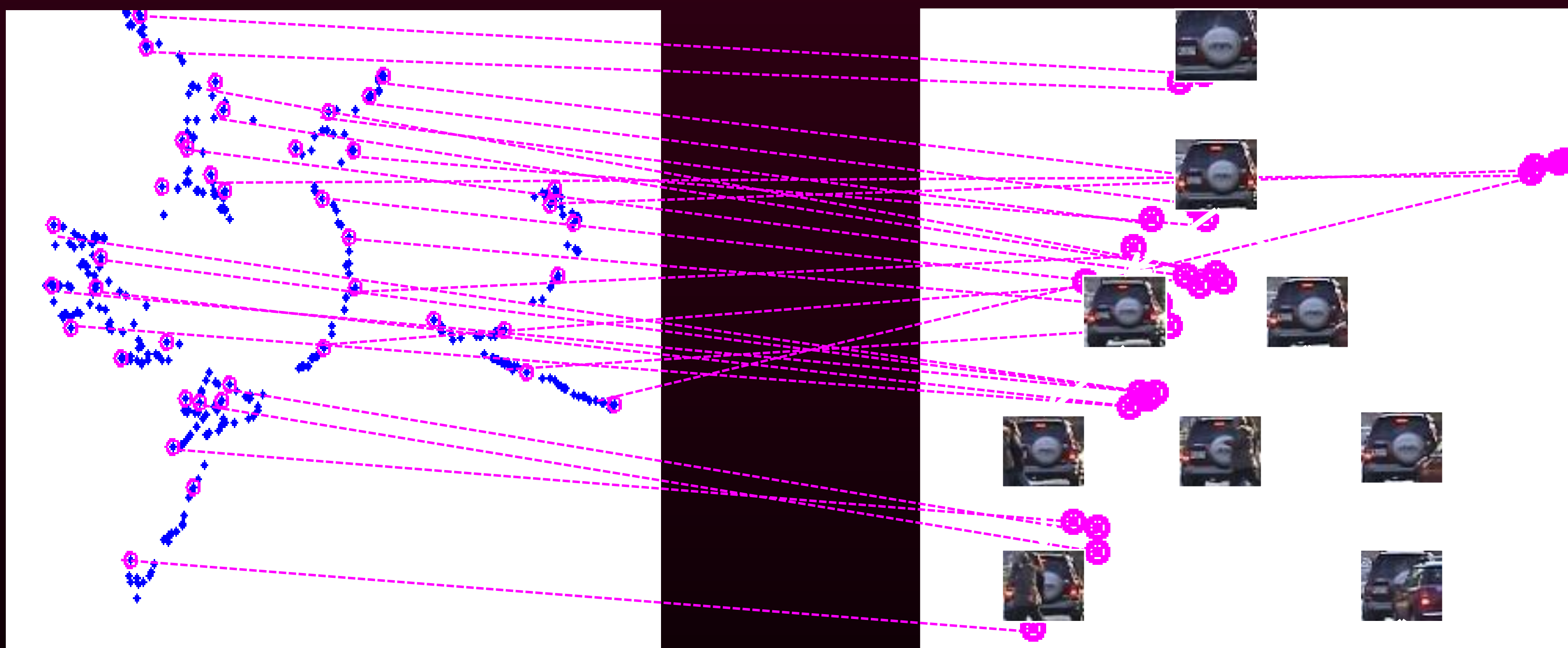
?

Non key frames



Tree Extension by Manifold Alignment

Entire frames **Key frames**



Embedding based on *scene* distance

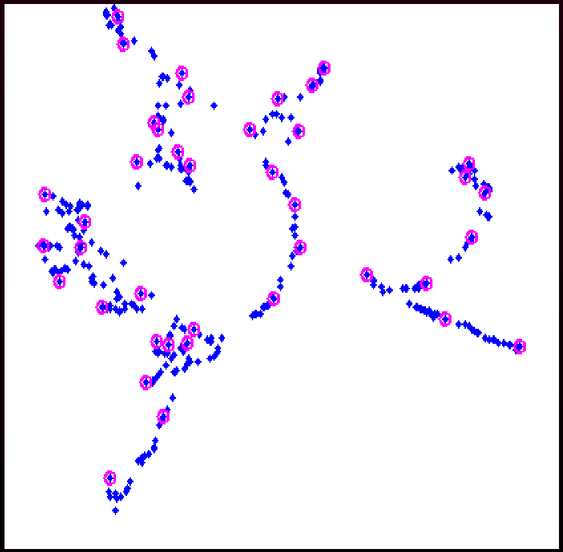
Embedding based on *target* distance

Tree Extension by Manifold Alignment

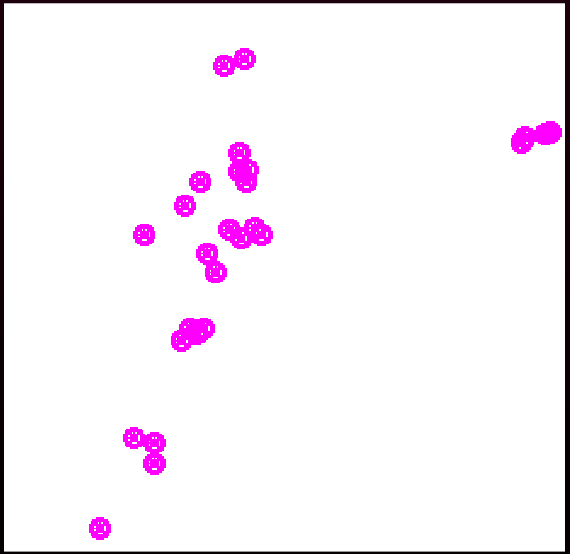
- Tree extension by semi-supervised manifold alignment[3]

$$\min_{s,t} \Phi(s,t) \equiv \mu \sum_{i \in \mathcal{K}} |s_i - t_i|^2 + s^T L^s s + t^T L^t t$$

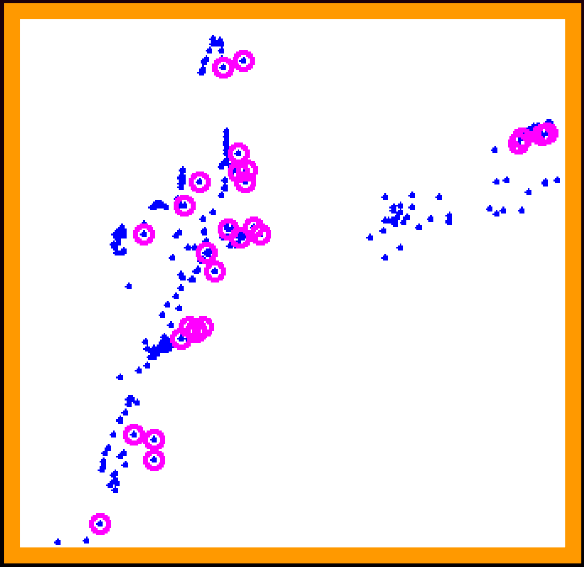
Scene based embedding s



Target based embedding t



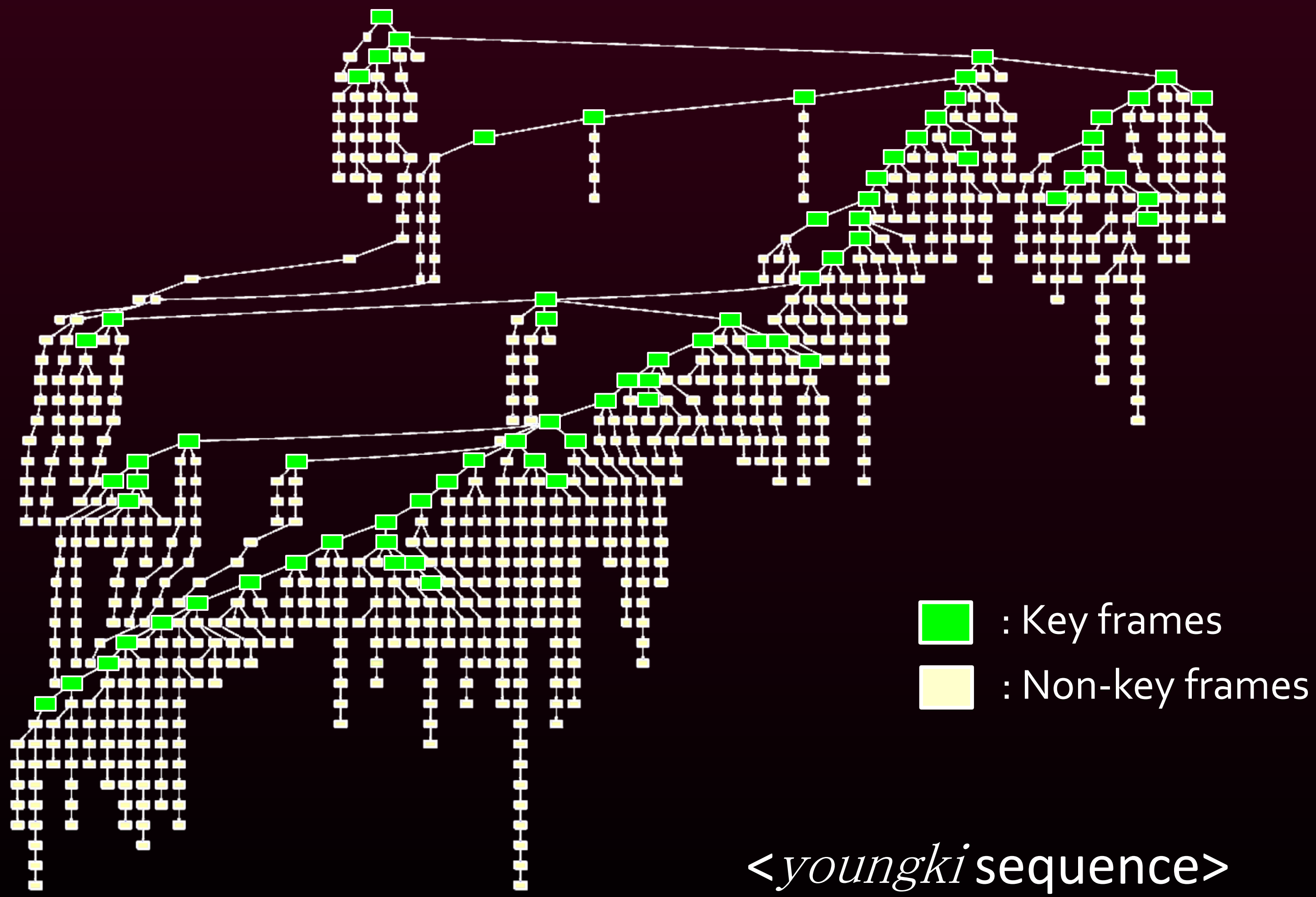
Joint embedding



[3] J. Ham, D. Lee, L. Saul, Semisupervised alignment of manifolds. In 10th International Workshop on Artificial Intelligence and Statistics, 2005

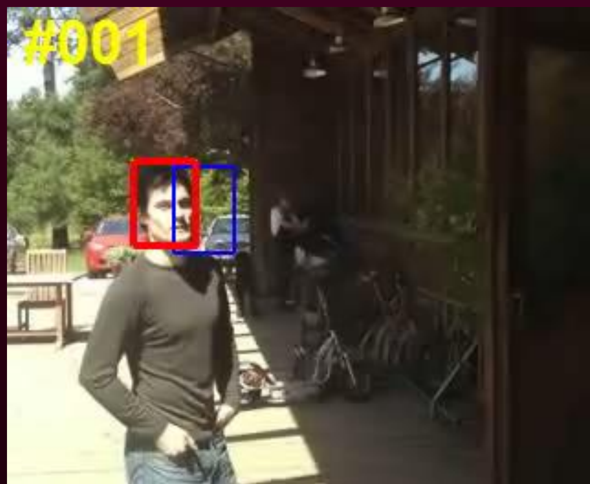
Tree Extension by Manifold Alignment

- Identified tree structures



<*youngki* sequence>

Qualitative Results



sunshade

bike

TUD

campus

OURS WLMC OTLE FRAG SCM L1APG

Qualitative Results

dance

skating

boxing

youngki

OURS WLMC OTLE FRAG SCM L1APG

Quantitative Results

Center location error

	FRAG	L1APG	CXT	ASLA	SCM	Struck	TLD	WLMC	OTLE	OMA	TST
campus	3.3	16.1	33.4	12.2	12.2	83.1	46.7	13.5	5.8	3.2	1.4
TUD	17.3	7.4	36.4	72.6	12.2	54.4	18.9	68.2	27.3	4.4	4.1
sunshade	35.8	42.8	30.6	37.2	44.9	3.9	19.9	61.1	9.1	88.1	5.3
bike	104.2	39.3	22.2	88.6	13.6	8.4	16.9	34.4	20.1	17.7	15.6
jumping	21.8	3.2	12.6	49.0	3.1	3.3	11.7	127.6	20.2	3.4	2.8
tennis	67.4	84.9	129.8	67.2	65.9	109.5	64.5	30.9	36.2	6.9	5.6
boxing	80.0	117.4	137.3	137.3	96.0	122.7	73.3	11.7	41.6	10.5	10.6
youngki	97.5	144.1	68.1	144.1	115.0	115.1	60.2	16.0	15.7	11.4	13.5
skating	35.4	143.9	41.5	45.2	49.4	23.8	35.3	14.7	18.3	8.0	6.1
dance2	132.4	167.2	176.8	176.8	208.0	107.1	105.0	39.7	118.8	15.1	18.6

Bounding box overlap ratio

	FRAG	L1APG	CXT	ASLA	SCM	Struck	TLD	WLMC	OTLE	OMA	TST
campus	0.77	0.52	0.56	0.63	0.62	0.24	0.50	0.52	0.72	0.78	0.86
TUD	0.59	0.85	0.51	0.30	0.67	0.30	0.67	0.38	0.49	0.82	0.80
sunshade	0.33	0.32	0.49	0.43	0.45	0.78	0.57	0.24	0.60	0.29	0.70
bike	0.08	0.18	0.39	0.16	0.46	0.54	0.45	0.39	0.27	0.40	0.56
jumping	0.31	0.77	0.40	0.20	0.76	0.75	0.56	0.07	0.26	0.74	0.79
tennis	0.11	0.29	0.08	0.12	0.11	0.28	0.10	0.43	0.33	0.63	0.74
boxing	0.22	0.13	0.01	0.03	0.13	0.04	0.21	0.65	0.38	0.70	0.71
youngki	0.19	0.02	0.38	0.12	0.13	0.09	0.24	0.62	0.54	0.62	0.56
skating	0.25	0.02	0.25	0.13	0.20	0.40	0.33	0.46	0.41	0.42	0.55
dance2	0.14	0.02	0.08	0.10	0.07	0.08	0.07	0.45	0.30	0.52	0.52

Summary

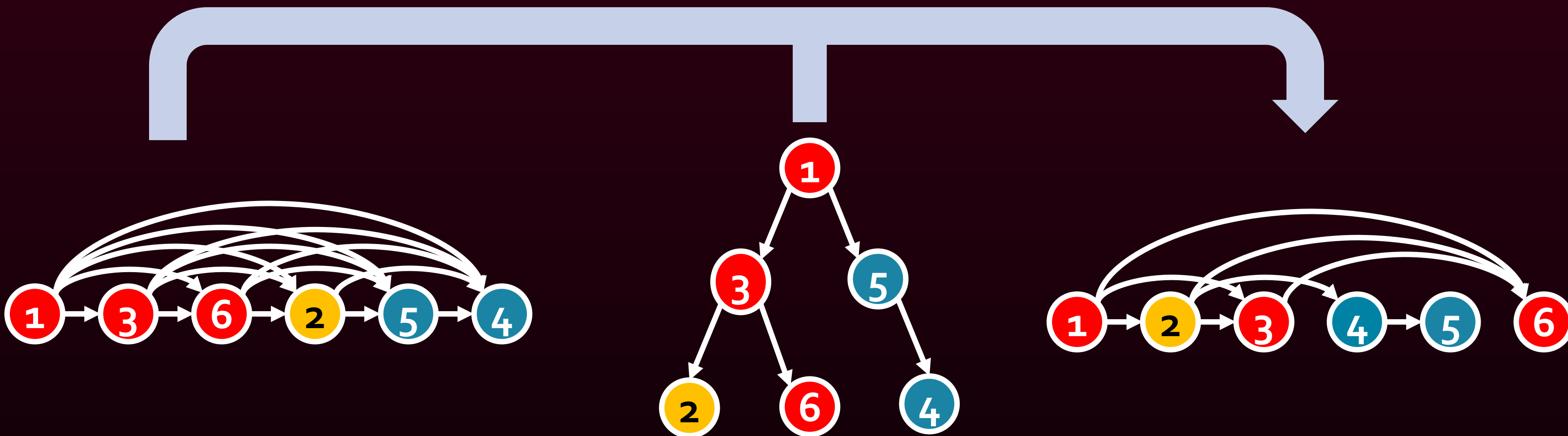
- Tree-structured graphical model for tracking
  - More general than chain model and blind model averaging

- Joint optimization of tree-learning and tracking
  - Based on MCMC sampling technique

- Hierarchical tracking
  - Based on semi-supervised manifold alignment

Advertisement

- We have a poster about online graph-based tracking algorithm.



(a) Bayesian model averaging (b) Tree-structure (c) Online-graph structure

Poster 3B-13, Wednesday