Generalized Background Subtraction Based on Hybrid Inference by Belief Propagation and Bayesian Filtering

Suha Kwak   Taegyu Lim   Woonhyun Nam   Bohyung Han   Joon Hee Han
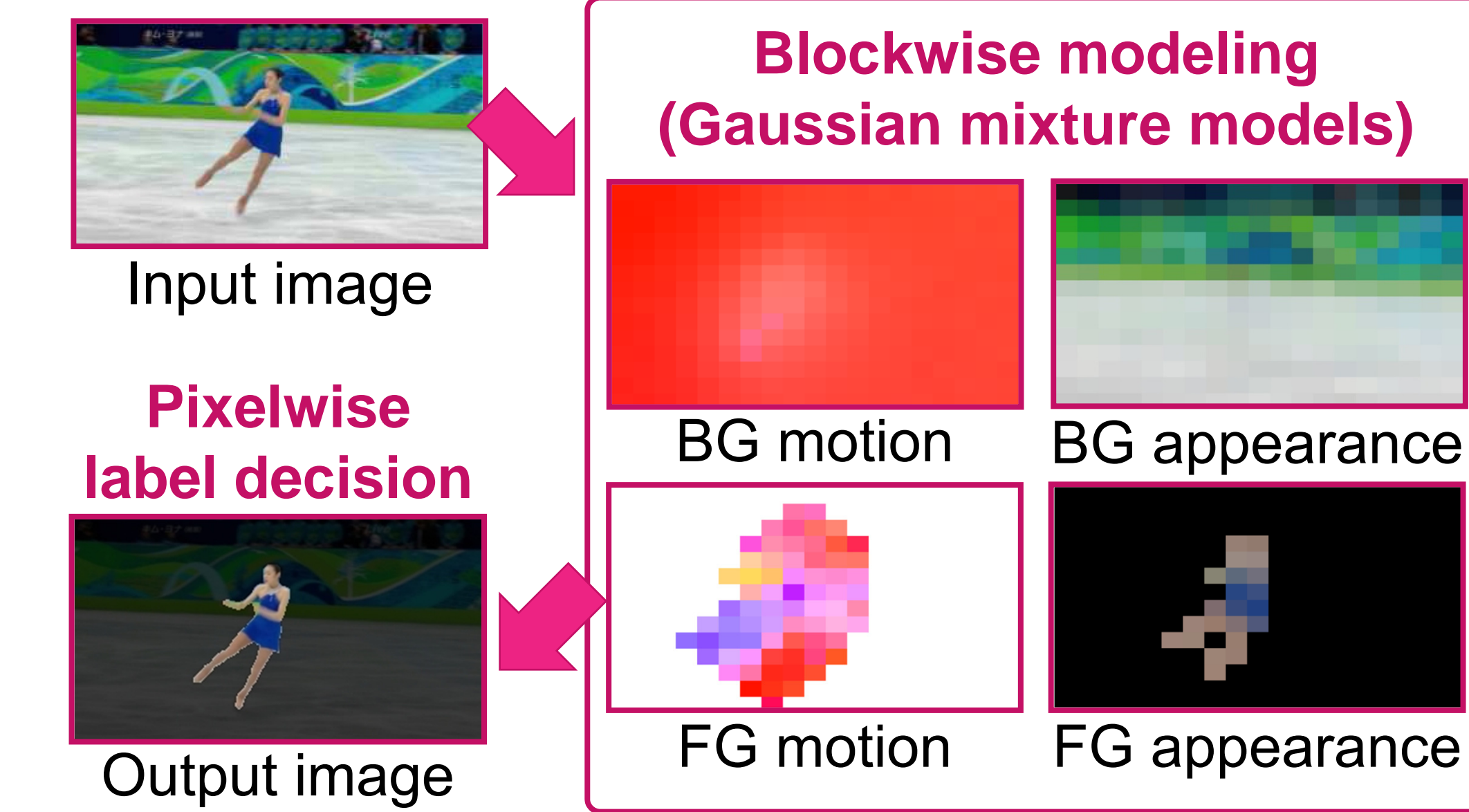
Department of Computer Science and Engineering, POSTECH, Korea

Problem

- **Generalized background subtraction:** background subtraction for videos captured by a moving camera
- **Our method**

Input image / Output image

Blockwise modeling (Gaussian mixture models): BG motion, BG appearance, FG motion, FG appearance
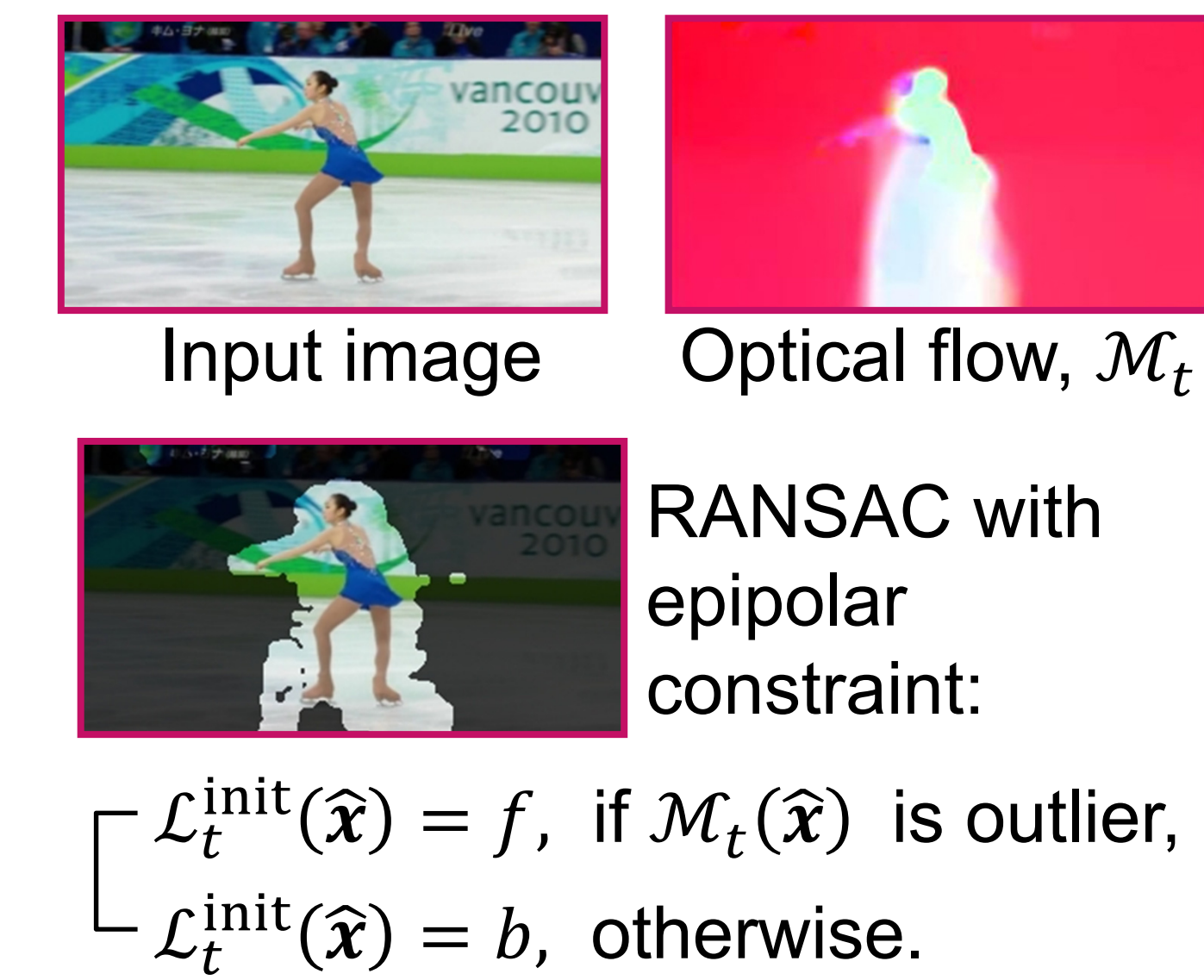
Pixelwise label decision

- Probabilistic inference for motion models, which is insensitive to noises and occlusions
  → Nonparametric belief propagation (NBP) [1]
- Reliable appearance estimation by prediction via motion and update with current observations
  → Sequential Bayesian filtering

Bayesian filtering for appearance estimation

Prediction via motion

Nonparametric belief propagation for motion estimation
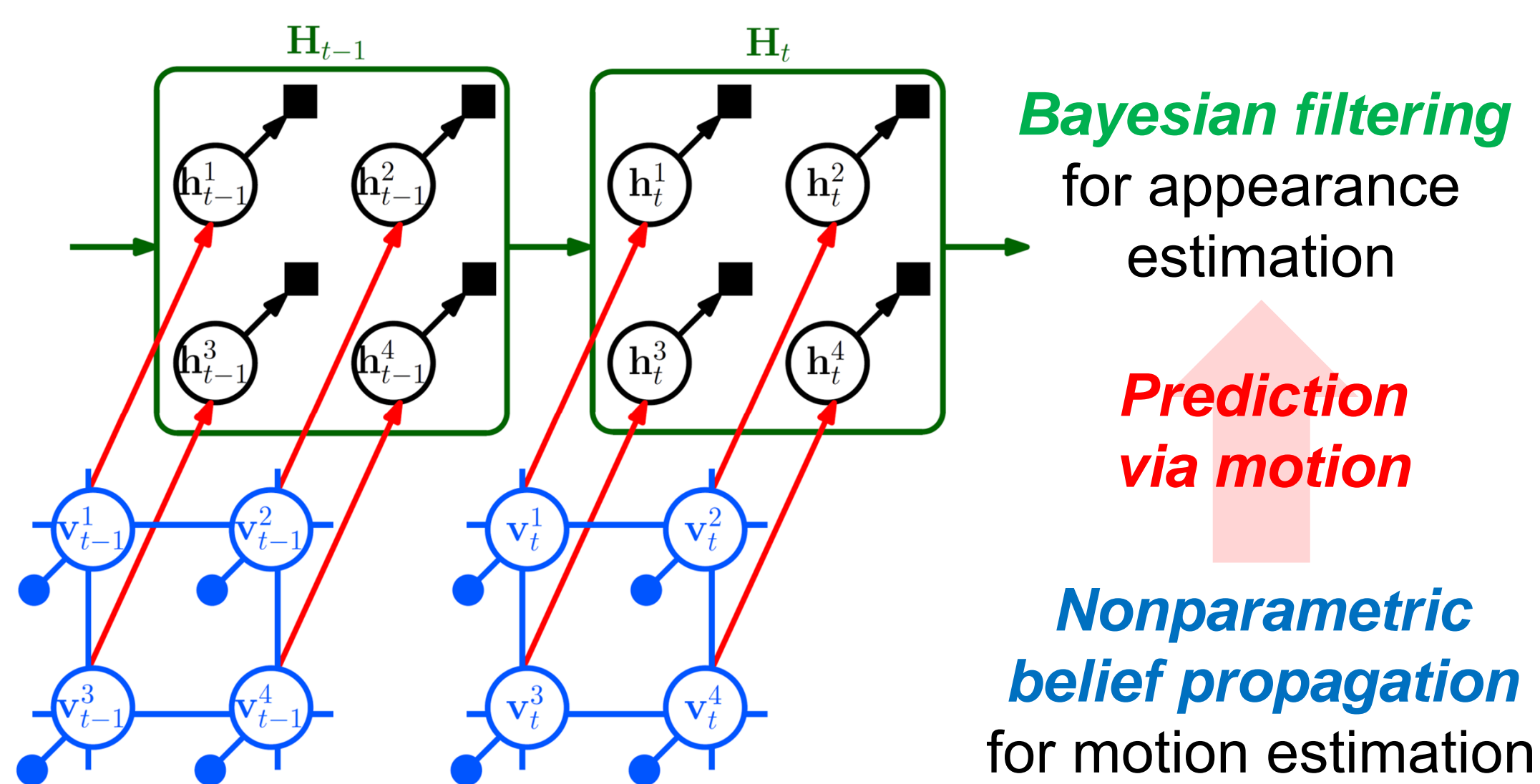
Initial Pixel-Labeling

- **Motion segmentation**
  - Optical flow + epipolar const.

Input image   Optical flow, $\mathcal{M}_t$

RANSAC with epipolar constraint:

$$\mathcal{L}_t^{\text{init}}(\hat{\boldsymbol{x}}) = f, \quad \text{if } \mathcal{M}_t(\hat{\boldsymbol{x}}) \text{ is outlier,}$$
$$\mathcal{L}_t^{\text{init}}(\hat{\boldsymbol{x}}) = b, \quad \text{otherwise.}$$

Motion Estimation by NBP

- **NBP in a four neighborhood MRF**
  - Joint posterior probability w.r.t. the motion models

$$p(V_t|\mathcal{M}_t) \propto \prod_{i\in\mathcal{V}} \Phi(\boldsymbol{v}_t^i, \mathcal{M}_t) \prod_{(i,j)\in\mathcal{E}} \Psi(\boldsymbol{v}_t^i, \boldsymbol{v}_t^j)$$

Gaussian KDE from flows in the ith block

$\mathcal{N}(\boldsymbol{v}_t^i - \boldsymbol{v}_t^j; \mathbf{0}, \Sigma_\Psi)$, spatial adjacency between motions

- NBP to estimate marginalized motion models [1]

noise removal / recovering from occlusion

Optical flow, $\mathcal{M}_t$ $p(\boldsymbol{v}_t^i|\mathcal{M}_t)$ for BG $p(\boldsymbol{v}_t^i|\mathcal{M}_t)$ for FG
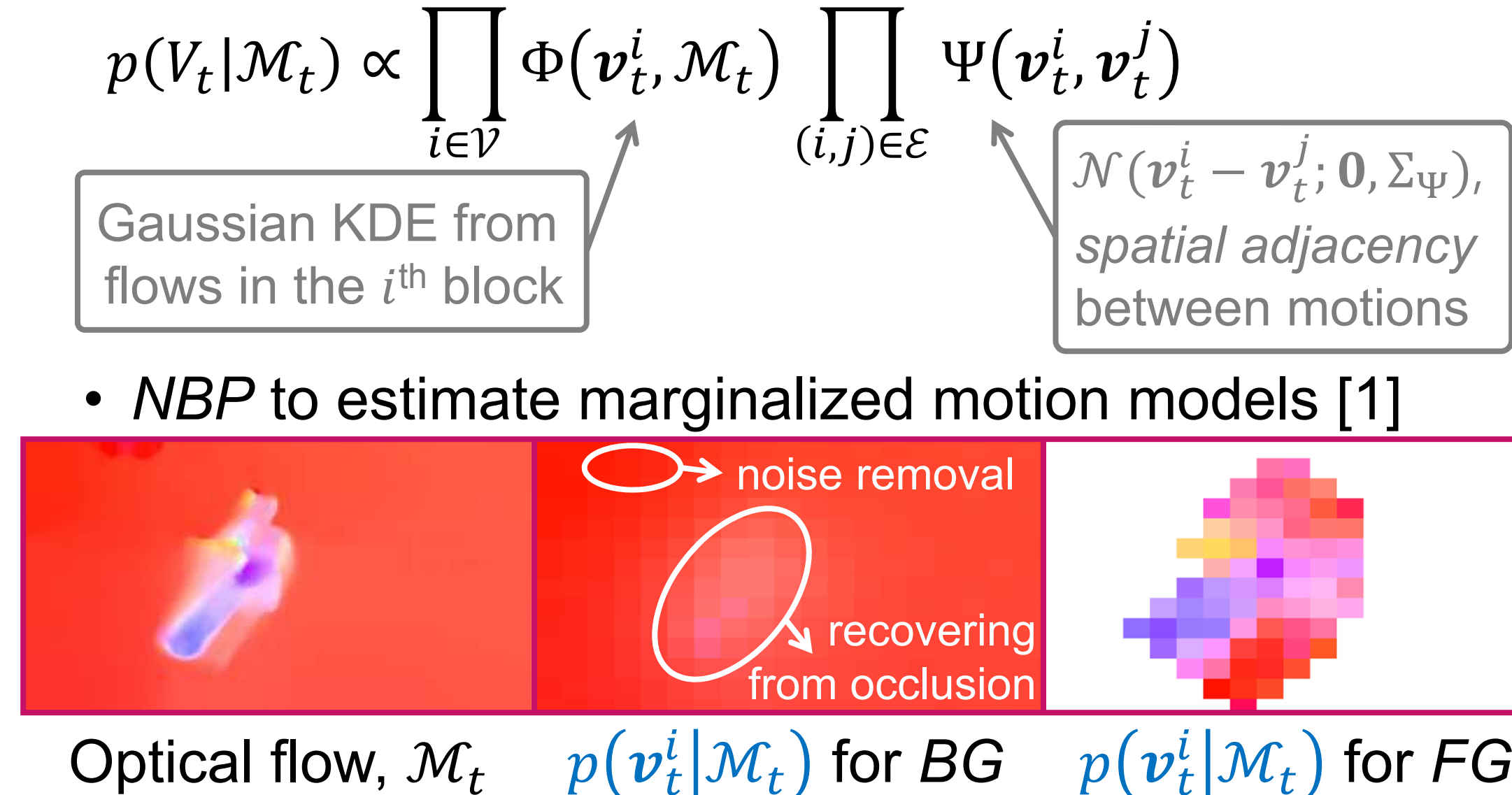
Pixelwise Label Estimation

- **Pixelwise belief propagation for binary labeling** [2]
  - Observation clique potentials

$$\varphi(\mathcal{L}_t(\boldsymbol{x})=f) = \frac{\zeta_t^i(\mathcal{M}_t(\boldsymbol{x}),f)\cdot\xi_t^i(\mathcal{I}_t(\boldsymbol{x}),f)}{\sum_{\ell\in\{b,f\}}\zeta_t^i(\mathcal{M}_t(\boldsymbol{x}),\ell)\cdot\xi_t^i(\mathcal{I}_t(\boldsymbol{x}),\ell)},$$
$$\varphi(\mathcal{L}_t(\boldsymbol{x})=b) = 1 - \varphi(\mathcal{L}_t(\boldsymbol{x})=f).$$

motion likelihood: $p(\boldsymbol{v}_{\ell,t}^i = \mathcal{M}_t(\boldsymbol{x})|\mathcal{M}_t)$

appearance likelihood: $p(\boldsymbol{h}_{\ell,t}^i = \mathcal{I}_t(\boldsymbol{x})|\mathcal{M}_{1:t}, \mathcal{I}_{1:t-1})$

missing leg / noisy head

Input   Motion   Appearance   Both of them
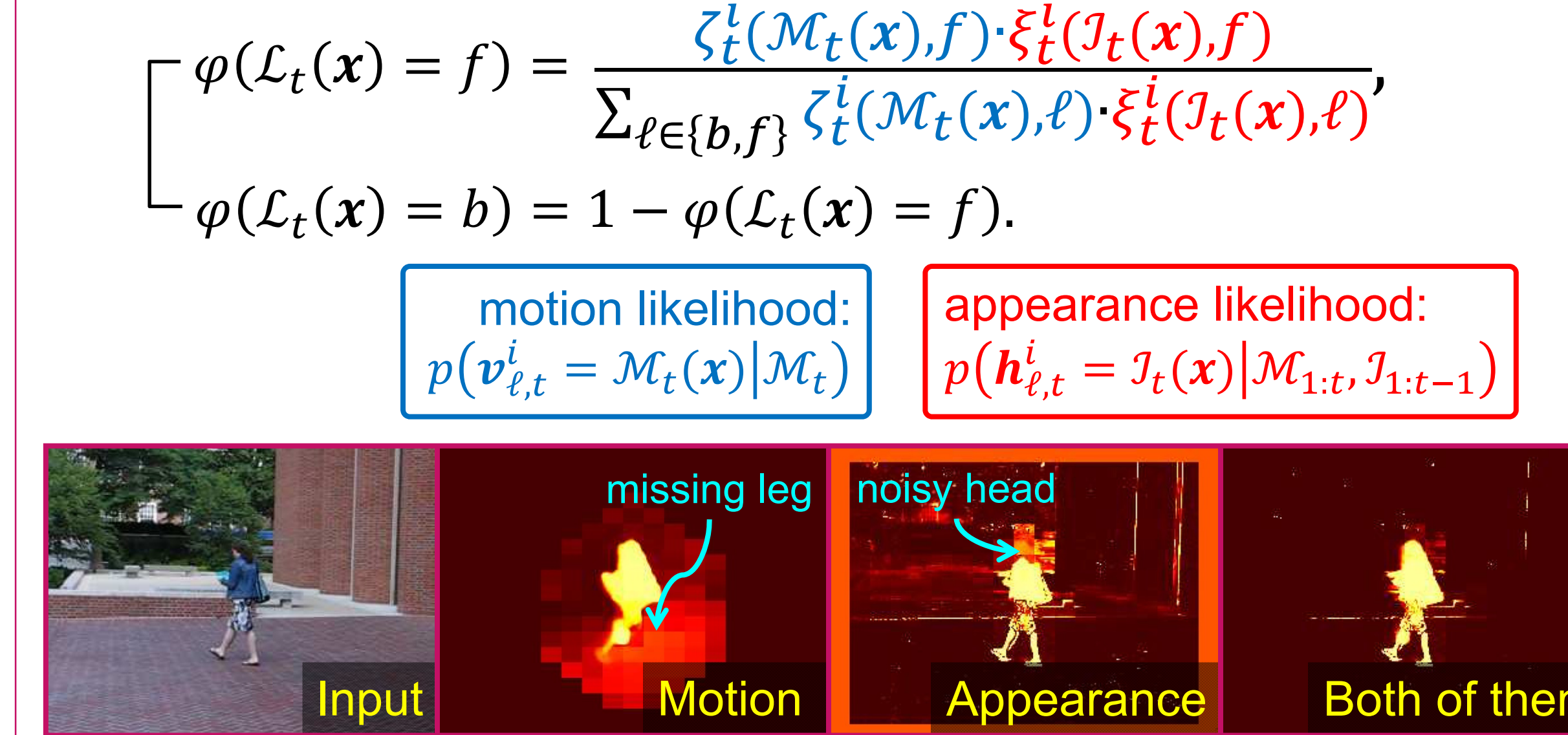
Appearance Prediction via Motion

- **Mixture of predictions**
  - Given a backward motion $\hat{\boldsymbol{v}}_t^i$, appearance prediction for the ith block at time t is given by

$$p(\boldsymbol{h}_t^i|\hat{\boldsymbol{v}}_t^i, \mathcal{M}_{1:t-1}, \mathcal{I}_{1:t-1}) = \sum_{k=1}^N \gamma_k^i(\hat{\boldsymbol{v}}_t^i)\cdot p_k(\boldsymbol{h}_t^i|\mathcal{M}_{1:t-1},\mathcal{I}_{1:t-1})\cdot$$

prediction of the kth block at time $t-1$:
$$\int_{\mathbb{R}^3} \mathcal{N}(\boldsymbol{h}_t^i - \boldsymbol{h}_{t-1}^k; \mathbf{0}, \Sigma_{tr})\cdot p(\boldsymbol{h}_{t-1}^k|\mathcal{M}_{1:t-1},\mathcal{I}_{1:t-1})d\boldsymbol{h}_{t-1}^k$$
→ smoothness of appearance variations over time

responsibility of the kth block at time $t-1$:
$$\gamma_k^i(\hat{\boldsymbol{v}}_t^i) \approx \int_{R(k)} \mathcal{N}(\boldsymbol{\tau}; \boldsymbol{c}^i + \hat{\boldsymbol{v}}_t^i, \Sigma_\gamma)d\boldsymbol{\tau}$$

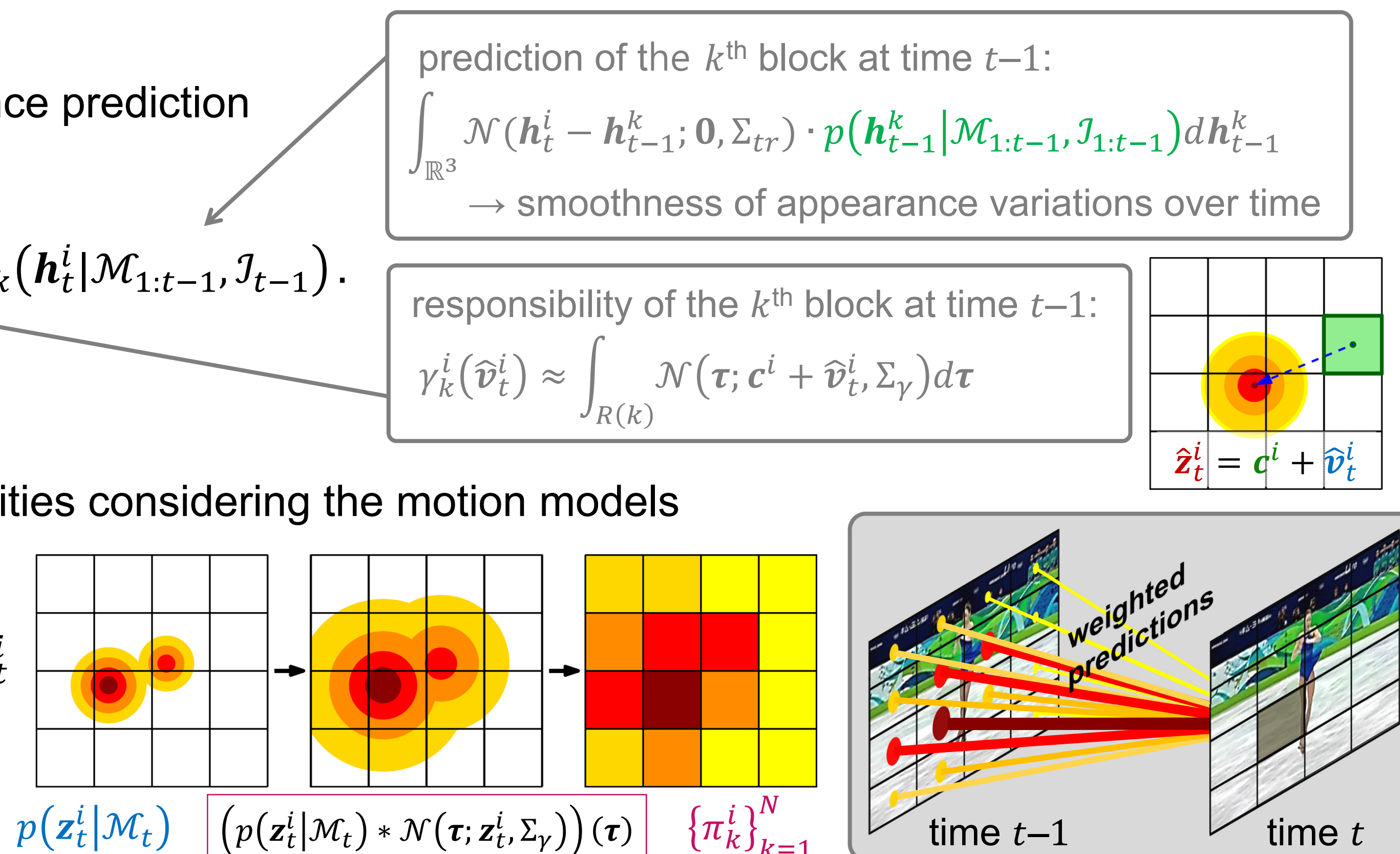$\boldsymbol{z}_t^i = \boldsymbol{c}^i + \hat{\boldsymbol{v}}_t^i$

- **Appearance prediction via motion**
  - Mixture weights = expected responsibilities considering the motion models

$$p(\boldsymbol{h}_t^i|\mathcal{M}_{1:t},\mathcal{I}_{1:t-1})$$
$$= \int_{\mathbb{R}^2} p(\boldsymbol{h}_t^i|\boldsymbol{v}_t^i,\mathcal{M}_{1:t-1},\mathcal{I}_{1:t-1})\cdot p(\boldsymbol{v}_t^i|\mathcal{M}_t)d\boldsymbol{v}_t^i$$
$$= \sum_{k=1}^N \pi_k^i\cdot p_k(\boldsymbol{h}_t^i|\mathcal{M}_{1:t-1},\mathcal{I}_{t-1})$$

$p(\boldsymbol{z}_t^i|\mathcal{M}_t)$ $\left(p(\boldsymbol{z}_t^i|\mathcal{M}_t)*\mathcal{N}(\boldsymbol{\tau};\boldsymbol{z}_t^i,\Sigma_\gamma)\right)(\boldsymbol{\tau})$ $\{\pi_k^i\}_{k=1}^N$

weighted predictions — time $t-1$, time t

Appearance Estimation by Bayesian Filtering

- **Update the prediction by observation**

$$p(\boldsymbol{h}_t^i|\mathcal{M}_{1:t},\mathcal{I}_{1:t}) \propto p(\mathcal{I}_t|\boldsymbol{h}_t^i)\cdot p(\boldsymbol{h}_t^i|\mathcal{M}_{1:t},\mathcal{I}_{1:t-1})$$

Gaussian KDE from colors in the ith block

13   85   170

Experiments

Legend: ■ Ours   ■ Ours without motion inference   ■ Motion segmentation   ■ [Sheikh et al., ICCV'09] with optical flow

Hopkins–car1   Hopkins–people2   Skating   Cycle

Hopkins–Car1: Precision 0.92, 0.41, 0.24, 0.24; Recall 0.84, 0.22, 0.36, 0.33; F-measure 0.88, 0.29, 0.29, 0.28

Hopkins–People2: Precision 0.85, 0.91, 0.62, 0.76; Recall 0.89, 0.92, 0.94, 0.99; F-measure 0.87, 0.92, 0.75, 0.86

Skating: Precision 0.95, 0.79, 0.25, 0.52; Recall 0.92, 0.87, 0.94, 0.97; F-measure 0.93, 0.83, 0.40, 0.68

Cycle: Precision 0.85, 0.91, 0.15, 0.49; Recall 0.85, 0.29, 0.50, 0.64; F-measure 0.85, 0.44, 0.24, 0.56