Temporal Depth Video Enhancement Based On Intrinsic Static Structure

Lu Sheng, King Ngi Ngan & Songnan Li
The Chinese University of Hong Kong

Outline

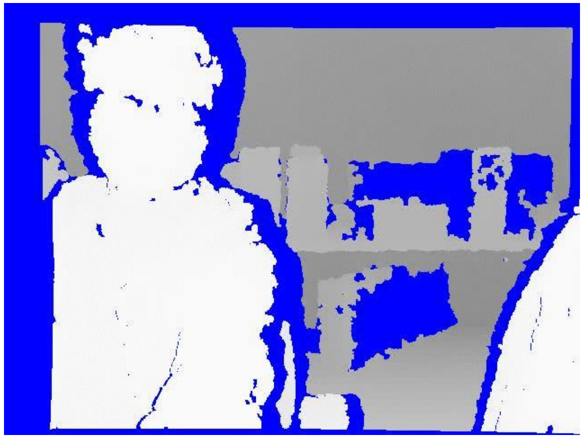
- Introduction & Related Work
- Intrinsic Static Structure
- Temporal Depth Video Enhancement
- Experimental Results
- Conclusion & Future Work

Introduction

- In a raw depth video capturing a natural scene
  - Various complex and even unpredictable dynamic contents
  - As well as spatial and temporal artifacts

- Artifacts in the depth video
  - Both in Spatial and temporal domain
  - Errors are caused by
    - Low resolution
    - Noise & outliers
    - Occlusions, material absorption & reflection
    - Distortion around object boundaries, and etc.

Raw Kinect video



Color-coded Raw TOF video

Introduction

- After the spatial enhancement
  - Reduce artifacts in spatial domain
  - But not enough to deal with ***temporal flickering***
    - No temporal consistency
    - Aggravate flickering artifacts

Raw Kinect video

Spatial JBF [1]

[1] J. Kopf et al, *Joint bilateral upsampling*, in ACM ToG., 2007.

Introduction

- After the spatial enhancement
  - Reduce artifacts in spatial domain
  - But not enough to deal with ***temporal flickering***
    - No temporal consistency
    - Aggravate flickering artifacts
- After a conventional spatiotemporal enhancement
  - Still contain temporal flickering
  - Distort the necessary depth variation along dynamic objects
- How to eliminate the temporal flickering while do not distort the necessary depth variation along the dynamic objects?

Spatial JBF

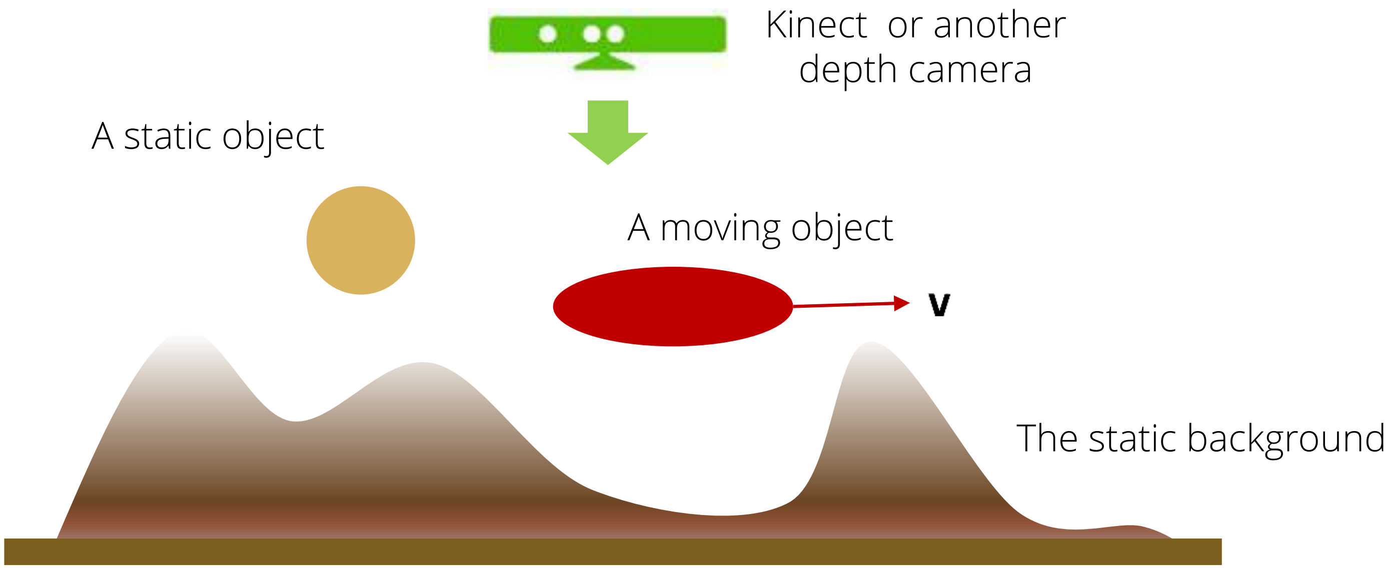
Coherent spatiotemporal JBF [2]

[2] N. Richardt, et al, Coherent spatiotemporal filtering, upsampling and rendering of RGBZ videos, Computer Graphics Forum, 2012

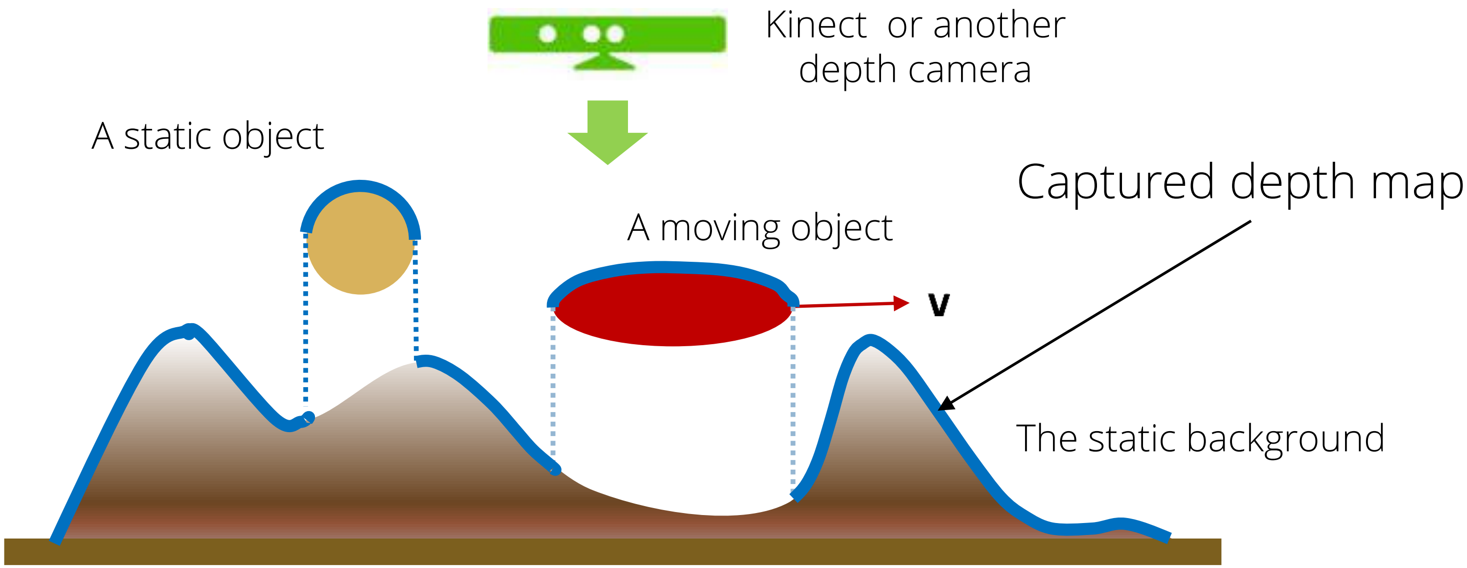
Related Work

- Spatial Enhancement
  - Global optimization
    - J. Diebel *et al.*, NIPS 2006
    - J. Yang *et al.*, ECCV 2012
    - J. Park *et al.*, ICCV 2011
    - and etc.
  - High dimensional Filtering
    - J. Kopf *et al.*, TOG 2007
    - J. Dolson *et al.*, CVPR 2010
    - B. Huhle *et al.*, CVIU 2010
    - and etc.
  - Patch Match and etc.
    - J. Lu *et al.*, CVPR 2013
    - L. Sheng *et al.*, ICIP 2013, and etc.
- Temporal Enhancement
  - Median filter
    - S. Matyunin *et al.*, WSCG 2011
  - $\mathrm{TV} - \ell_1$ norm
    - S. H. Chan *et al.*, TIP 2011
  - Texture consistency
    - Dongbo Min *et al.*, TIP 2012
    - Deliang Fu *et al.*, PCS 2010
  - Both texture and depth
    - N. A. Dodgson *et al.*, CGF 2012
  - Model the scene
    - J. Shen *et al.*, TIP 2013
    - J. Shen *et al.*, CVPR 2013

Intrinsic Static Structure

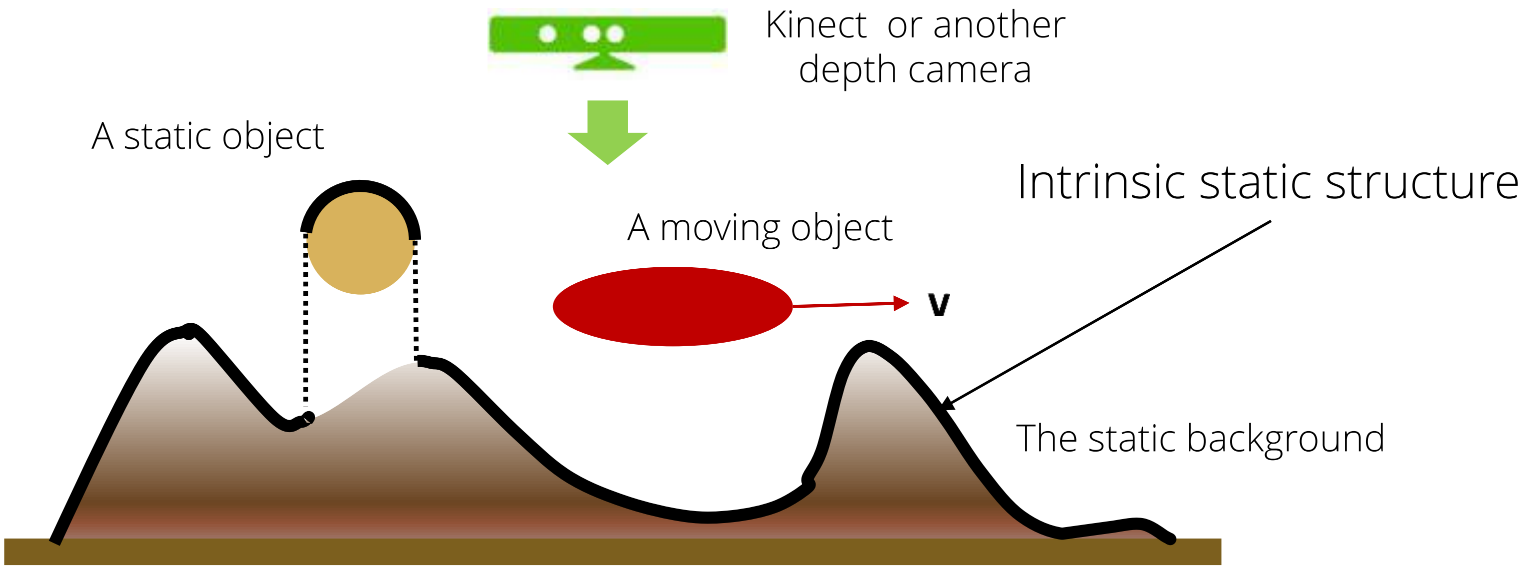


Intrinsic Static Structure



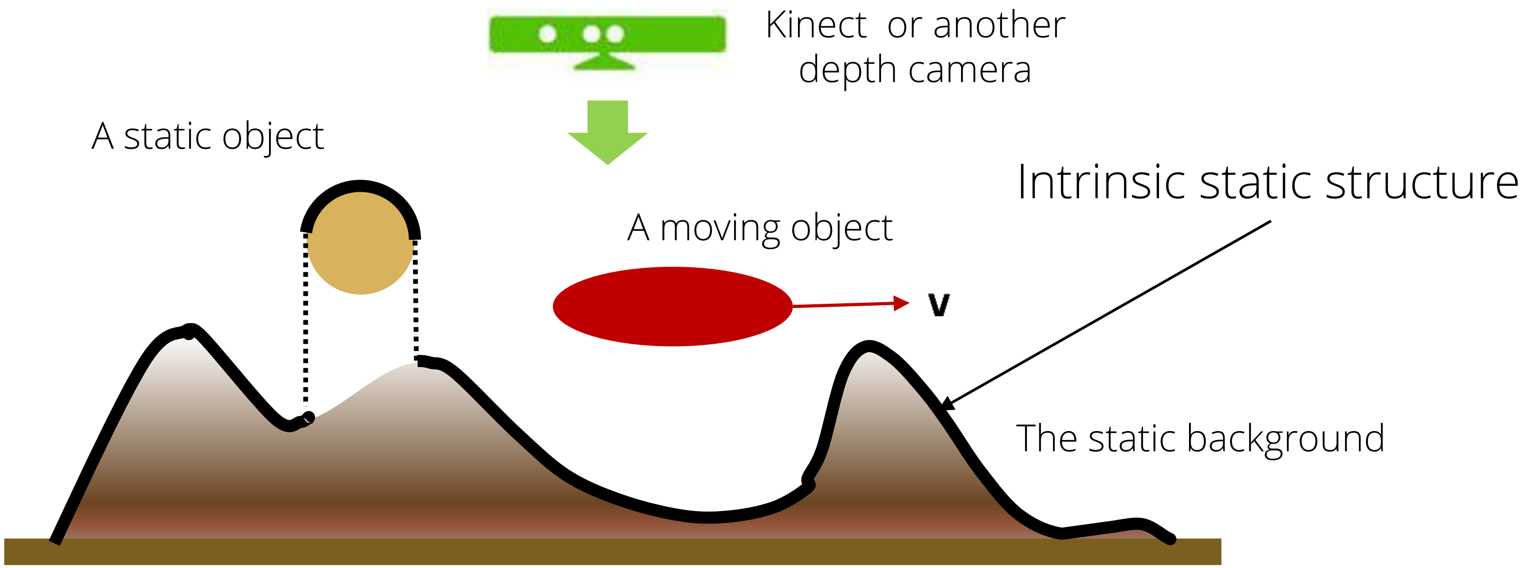
Intrinsic Static Structure

- Definition
  - A manifold lies on or behind the input depth map
  - Contains static structure of the captured scene



Intrinsic Static Structure

- Observation
  - Moving objects stay in its front
  - Static regions or visible background area are fused into it



Intrinsic Static Structure

- How to estimate the intrinsic static structure under an on-line fashion?

- Temporal enhancement of the depth video?

- A probabilistic generative model
- An online variational probabilistic update scheme with
  - A robust exclusion of the dynamic foreground objects
  - the inclusion of the once occluded background

- Robust static region detection by the intrinsic static structure
- Enhance the static region of input depth frame
  - Dynamic objects own limited temporal consistency, we just omit it
  - Fusing the static region with the estimated intrinsic static structure

Intrinsic Static Structure

- A Probabilistic Generative Model

  If incoming depth d_x^t belongs to

  - State-I: the intrinsic static structure

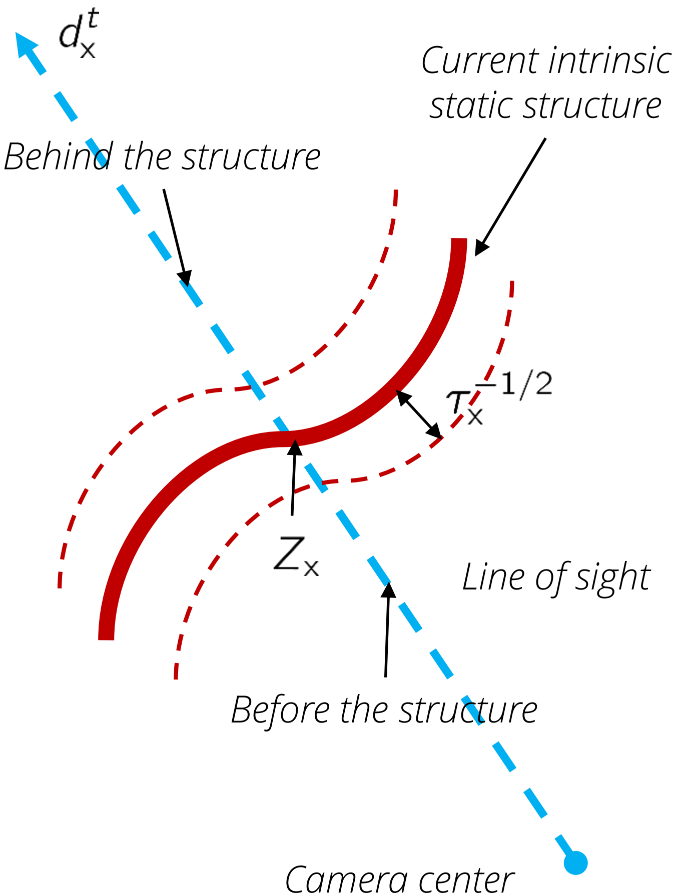
$$\mathcal{N}(d_x^t | Z_x, \tau_x^{-1})$$

  - State-II: outliers in the front or moving objects

$$\mathcal{U}_f(d_x^t | Z_x) = U_f \left[d_x^t < Z_x) \right]$$

  - State-III: outliers rearward or revealed background

$$\mathcal{U}_b(d_x^t | Z_x) = U_b \left[d_x^t > Z_x \right]$$



$[\cdot]$ is an indicate function that is equal to 1, when input argument is true and 0 vice visa

Intrinsic Static Structure

- A Probabilistic Generative Model

  If incoming depth d_{x}^t belongs to

  - State-I: the intrinsic static structure

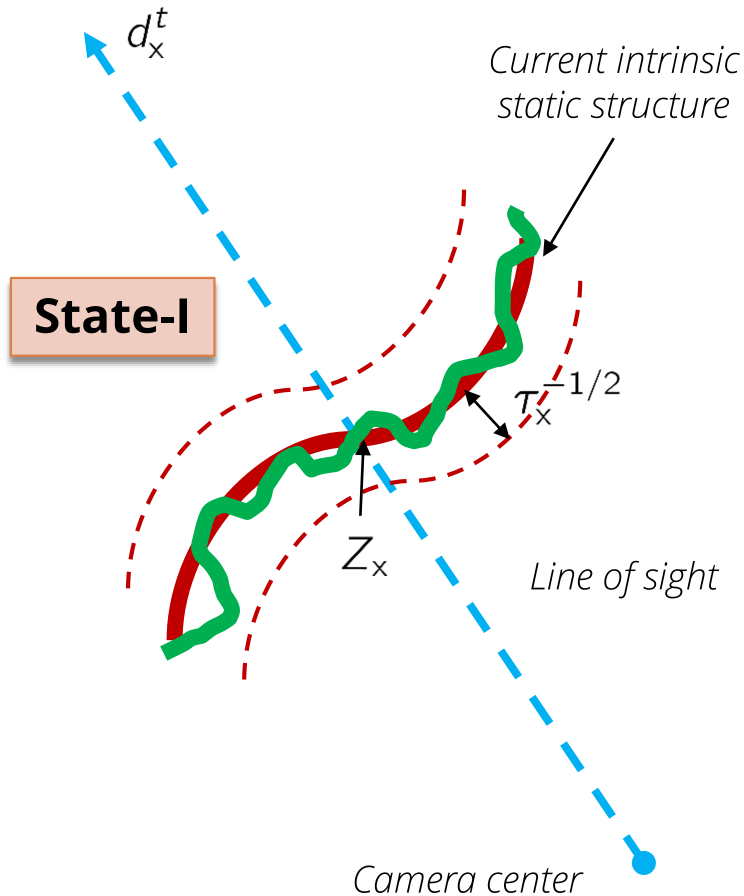
$$\mathcal{N}(d_{\mathrm{x}}^t | Z_{\mathrm{x}}, \tau_{\mathrm{x}}^{-1})$$

  - State-II: outliers in the front or moving objects

$$\mathcal{U}_f(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_f \left[d_{\mathrm{x}}^t < Z_{\mathrm{x}} \right]$$

  - State-III: outliers rearward or revealed background

$$\mathcal{U}_b(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_b \left[d_{\mathrm{x}}^t > Z_{\mathrm{x}} \right]$$



$[\cdot]$ is an indicate function that is equal to 1, when input argument is true and 0 vice visa

Intrinsic Static Structure

- A Probabilistic Generative Model

  If incoming depth d_{x}^t belongs to

  - State-I: the intrinsic static structure

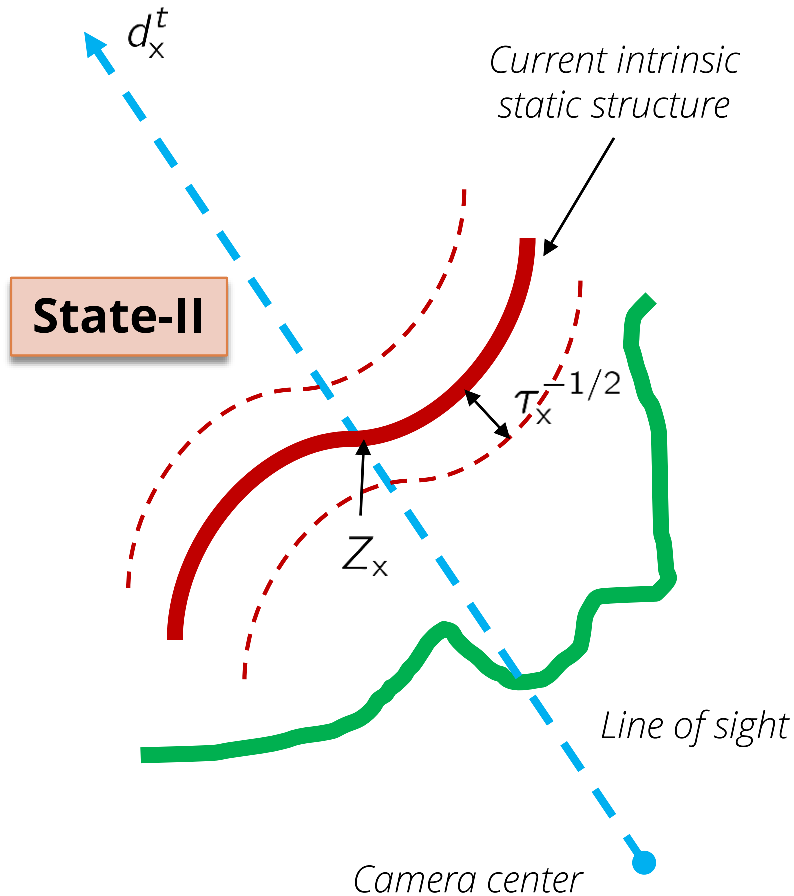
$$\mathcal{N}(d_{\mathrm{x}}^t | Z_{\mathrm{x}}, \tau_{\mathrm{x}}^{-1})$$

  - State-II: outliers in the front or moving objects

$$\mathcal{U}_f(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_f \left[d_{\mathrm{x}}^t < Z_{\mathrm{x}} \right]$$

  - State-III: outliers rearward or revealed background

$$\mathcal{U}_b(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_b \left[d_{\mathrm{x}}^t > Z_{\mathrm{x}} \right]$$



$[\cdot]$ is an indicate function that is equal to 1, when input argument is true and 0 vice visa

Intrinsic Static Structure

- A Probabilistic Generative Model

  If incoming depth d_{x}^t belongs to

  - State-I: the intrinsic static structure

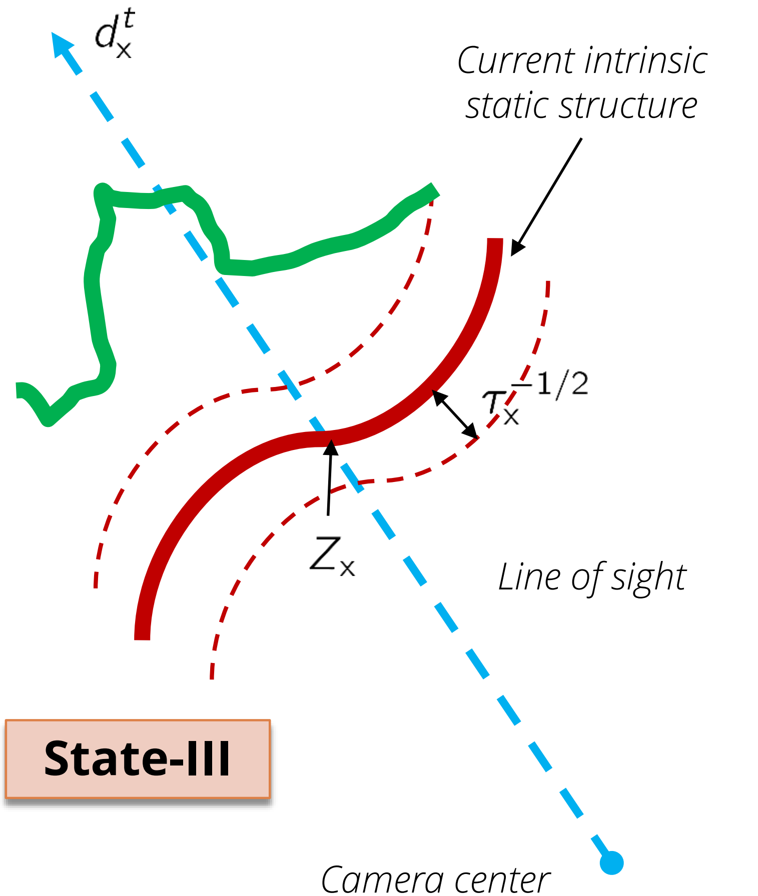
    $$\mathcal{N}(d_{\mathrm{x}}^t | Z_{\mathrm{x}}, \tau_{\mathrm{x}}^{-1})$$

  - State-II: outliers in the front or moving objects

    $$\mathcal{U}_f(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_f \left[d_{\mathrm{x}}^t < Z_{\mathrm{x}}) \right]$$

  - State-III: outliers rearward or revealed background

    $$\mathcal{U}_b(d_{\mathrm{x}}^t | Z_{\mathrm{x}}) = U_b \left[d_{\mathrm{x}}^t > Z_{\mathrm{x}} \right]$$



$[\cdot]$ is an indicate function that is equal to 1, when input argument is true and 0 vice visa

Intrinsic Static Structure

- A Probabilistic Generative Model
  - The likelihood of d_x^t w.r.t. the given intrinsic static structure

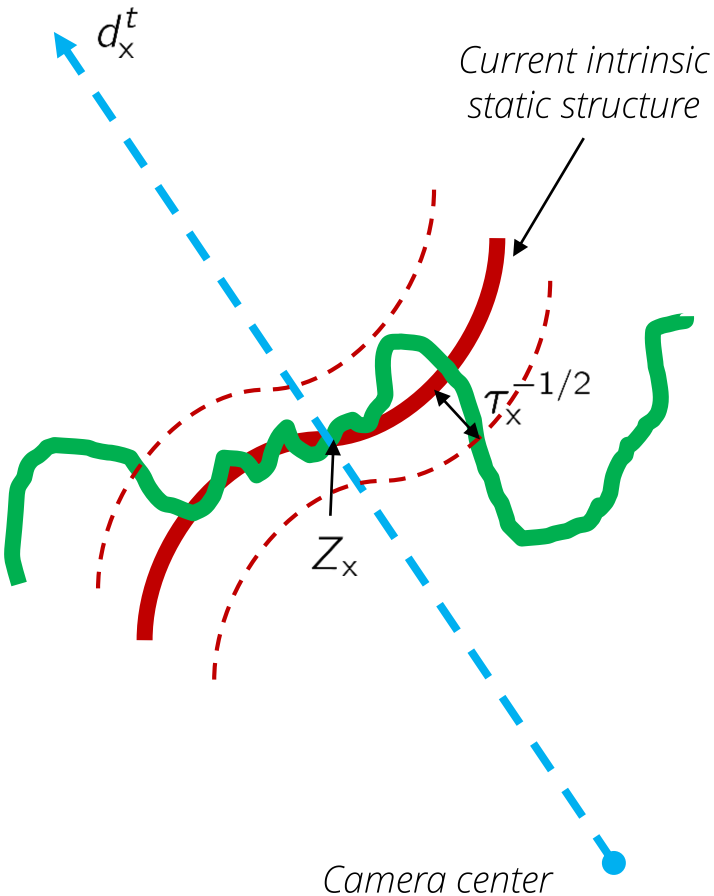
$$p(d_x^t|Z_x, \boldsymbol{\omega}_x) = \omega_x^1 \mathcal{N}(d_x^t|Z_x, \tau_x^{-1}) + \omega_x^2 \mathcal{U}_f(d_x^t|Z_x) + \omega_x^3 \mathcal{U}_b(d_x^t|Z_x)$$

  - Gaussian prior over Z_x

$$p(Z_x) = \mathcal{N}\left(Z_x|\mu_x, \lambda_x^{-1}\right)$$

  - Dirichlet prior over the frequency of each state

$$p(\boldsymbol{\omega}_x) = \mathrm{Dir}\left(\boldsymbol{\omega}_x|\alpha_{1,x}, \alpha_{2,x}, \alpha_{3,x}\right)$$



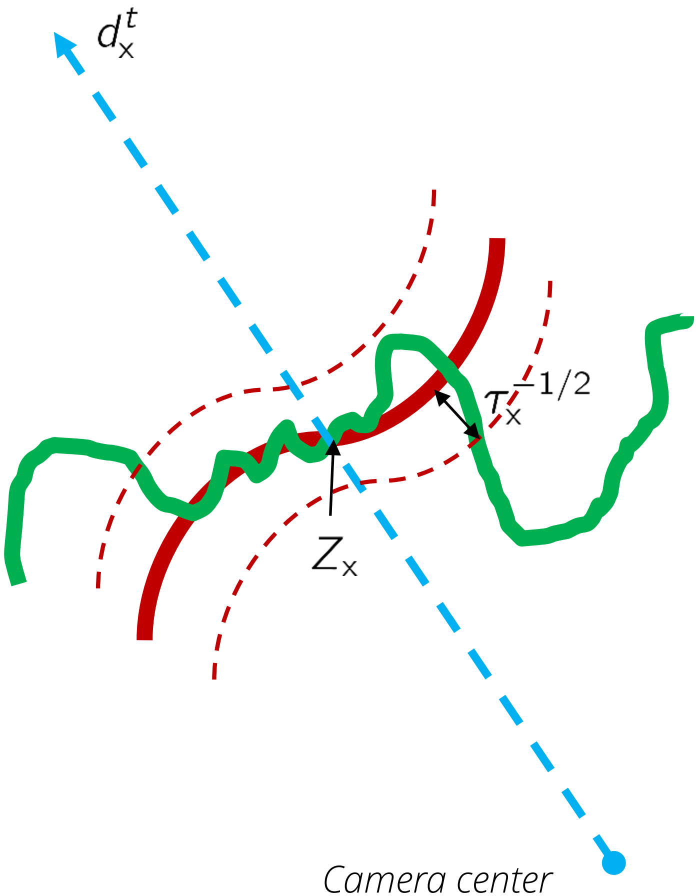
Intrinsic Static Structure

- A Probabilistic Generative Model
  - The posterior

$$p(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}} | \mathcal{D}_{\mathrm{x}}^{t}) = \frac{p(d_{\mathrm{x}}^{t} | Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}}) p(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}} | \mathcal{D}_{\mathrm{x}}^{t-1})}{p(d_{\mathrm{x}}^{t} | \mathcal{D}_{\mathrm{x}}^{t-1})}$$

$\mathcal{D}_{\mathrm{x}}^{t-1}$ is the set of previous depth samples

$\mathcal{D}_{\mathrm{x}}^{t} = \{\mathcal{D}_{\mathrm{x}}^{t-1}, d_{\mathrm{x}}^{t}\}$ is the set of current samples



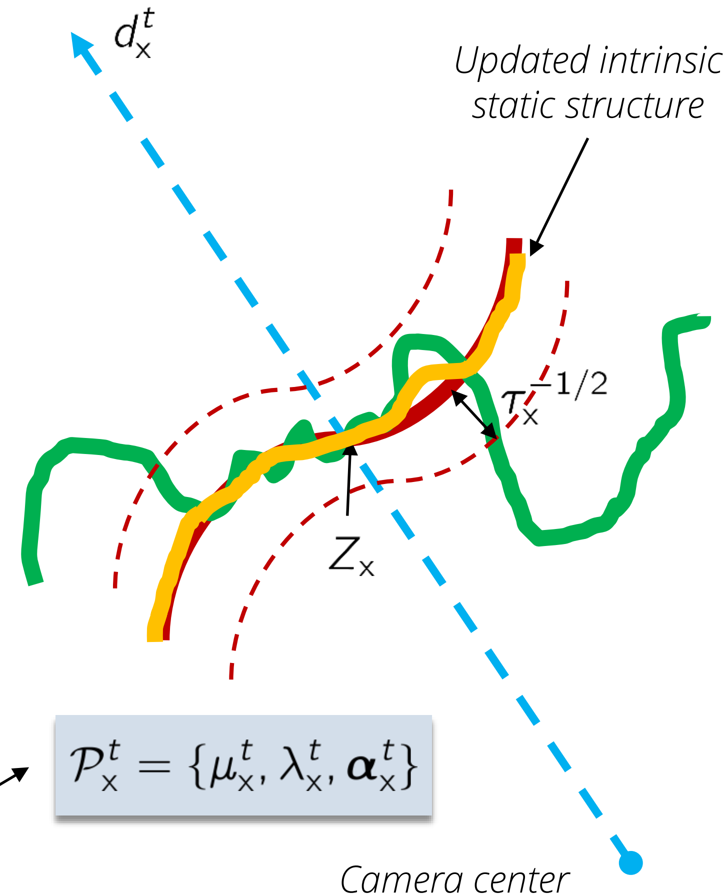
On-line Update Scheme

- A Probabilistic Generative Model
  - The posterior

$$p(Z_x, \boldsymbol{\omega}_x | \mathcal{D}_x^t) = \frac{p(d_x^t | Z_x, \boldsymbol{\omega}_x) p(Z_x, \boldsymbol{\omega}_x | \mathcal{D}_x^{t-1})}{p(d_x^t | \mathcal{D}_x^{t-1})}$$

$\mathcal{D}_x^{t-1}$ is the set of previous depth samples

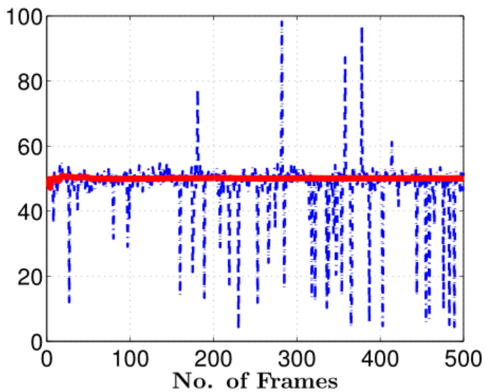
$\mathcal{D}_x^t = \{\mathcal{D}_x^{t-1}, d_x^t\}$ is the set of current samples



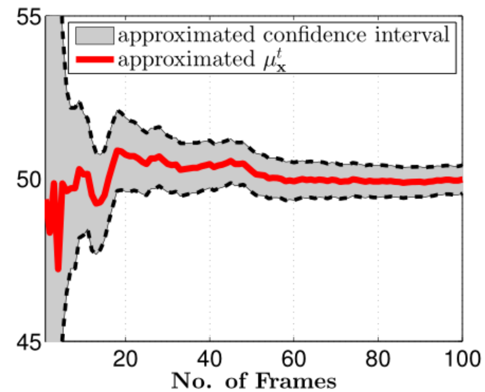
If the input frame only contains the static scene and outliers, the updated intrinsic static structure will be governed by the posterior, and we have

- The reliability of the model is $\mathbb{E}_{p(\boldsymbol{\omega}_x | \mathcal{D}_x^t)} [\omega_x^1]$
- Its probable depth is $\mathbb{E}_{p(Z_x | \mathcal{D}_x^t)} [Z_x]$
- Variational approximation for efficiency

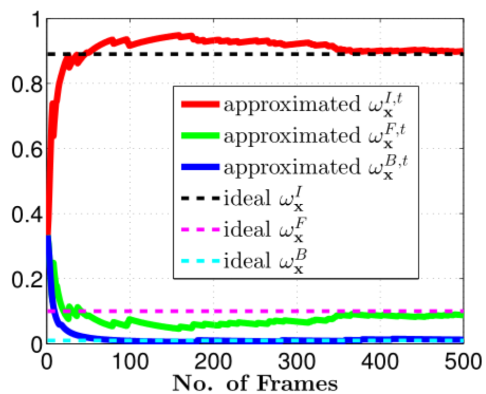
On-line Update Scheme



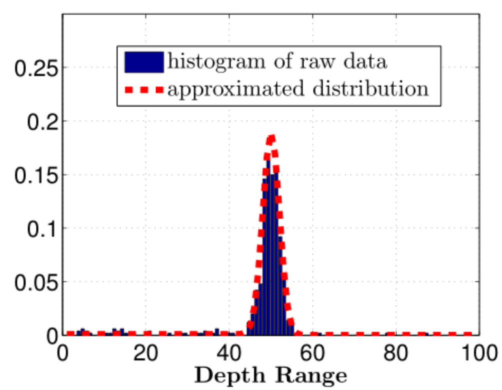
(a) $Z_{\mathbf{x}}^{t}$ *v.s.* $\mathcal{D}_{\mathbf{x}}^{t}$



(b) Confidence interval *w.r.t.* $Z_{\mathbf{x}}^{t}$



(c) Evolution of each states' portions



(d) $q^{T}(d_{\mathbf{x}}|\mathcal{P}_{\mathbf{x}}^{\mathcal{D},T})$ *v.s.* data histogram

- Variational approximation for a 1D depth signal

a) the probable depth values of the intrinsic static structure versus the raw depth signals

b) The confidence interval of the intrinsic static structure

c) The portions of three states

d) The approximated date distribution $q^{T}(d_{x}|\mathcal{P}_{x}^{T})$ versus the histogram of the raw data

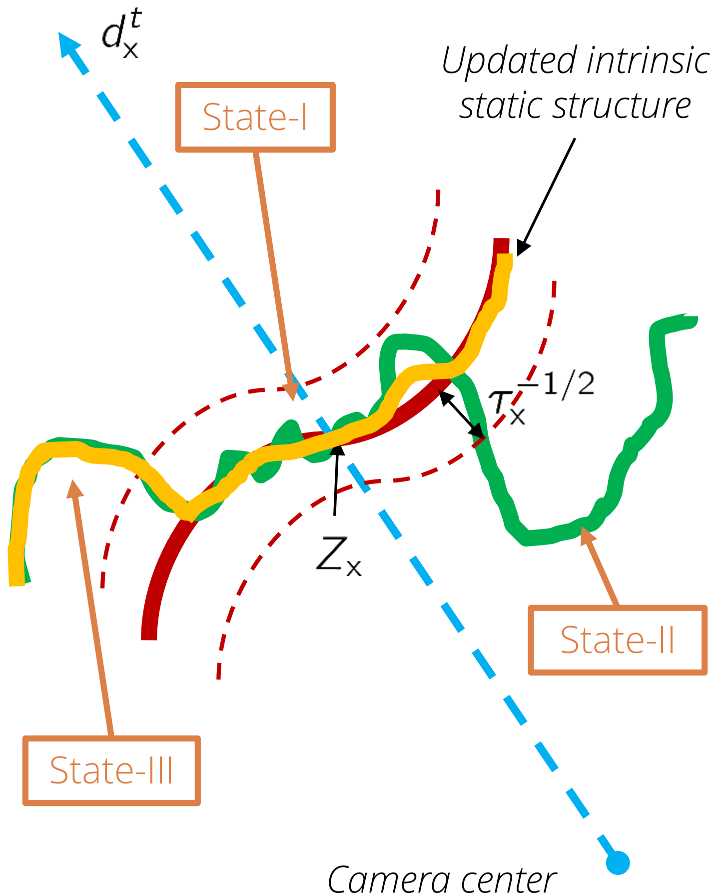
> The nature of the input data samples is still well captured by the variational approximation

On-line Update Scheme

If the input frame contains the dynamic content, the posterior w.r.t. each state indicates the criterion to update the intrinsic static structure

State-I: fuse the new samples into the current model
State-II: leave the current model unchanged
State-III: have a risk that current model is incorrect, might require re-initialization

- Results after variational approximation are valid for **State-I** and **State-II**
- Only need *an additional criterion* for **State-III**



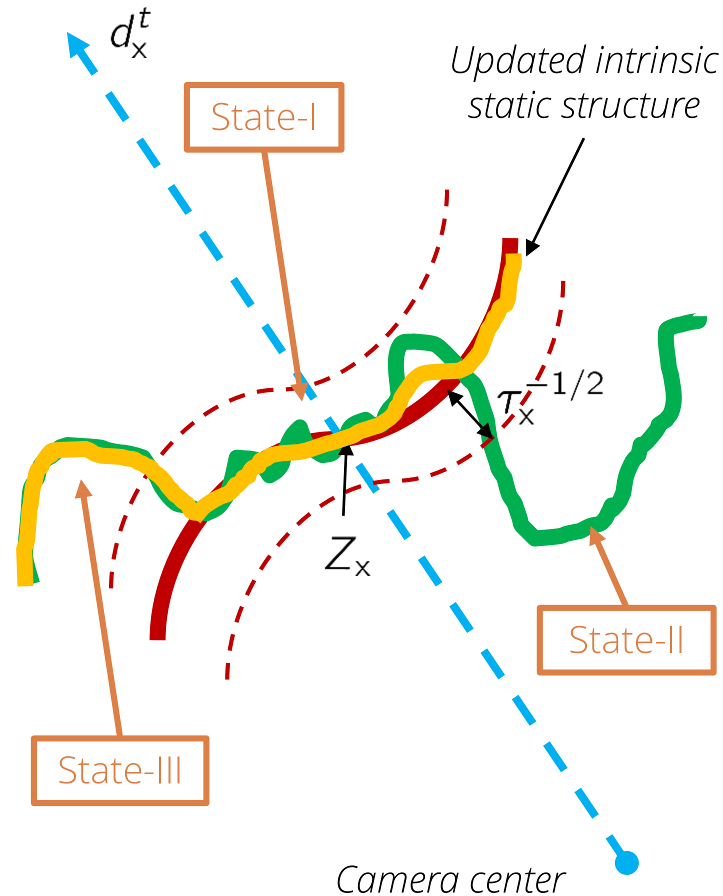
On-line Update Scheme

- Re-initialization to avoid incorrect estimation when
  - Ratio $\rho_x = \alpha^t_{3,x} / \alpha^t_{1,x} > T_\rho$

    It means the percentage of **State-III** is too large to provide enough confidence that current estimation is true
  - Increment $\Delta_F \alpha_{3,x} = \alpha^t_{3,x} - \alpha^{t-F}_{3,x} > T_F$

    It means there are a successive frames that the states of input depth samples are **State-III**
  - Or texture consistency is violated at the locations of **State-III**

    If RGB-D video is applicable. One step further, we can exploit the conditional random field to robustly find the re-initialized region



On-line Update Scheme

Algorithm 1: Intrinsic Static Structure Update Scheme

input : Input depth sequence $\{d_{\mathrm{x}}^{t} | t = 1, 2, \ldots, N\}$;
Initial parameter set $\mathcal{P}_{init} = \{\mu_{\mathrm{x}}^{0}, \lambda_{\mathrm{x}}^{0}, \boldsymbol{\alpha}_{\mathrm{x}}^{0}\}$;
output: Current parameter set $\mathcal{P}_{\mathrm{x}}^{t} = \{\mu_{\mathrm{x}}^{t}, \lambda_{\mathrm{x}}^{t}, \boldsymbol{\alpha}_{\mathrm{x}}^{t}\}$;

$\mu_{\mathrm{x}}^{0} \leftarrow d_{\mathrm{x}}^{1}$, $\mathcal{P}_{\mathrm{x}}^{1} \leftarrow \mathcal{P}_{init}$;

- ➢ Parameter initialization

for $t \leftarrow 2$ to N do
  if $d_{\mathrm{x}}^{t} > 0$ then

- ➢ When input depth sample is valid

    estimate parameter set $\mathcal{P}_{\mathrm{x}}^{t}$ based on $\mathcal{P}_{\mathrm{x}}^{t-1}$ by variational approximaion;
    $\rho_{\mathrm{x}} \leftarrow \alpha_{3,\mathrm{x}}^{t} / \alpha_{1,\mathrm{x}}^{t}$, $\Delta_{F}\alpha_{3,\mathrm{x}} \leftarrow \alpha_{3,\mathrm{x}}^{t} - \alpha_{3,\mathrm{x}}^{t-F}$;
    if $\rho_{\mathrm{x}} > T_{\rho}$ *or* $\Delta_{F}\alpha_{3,\mathrm{x}} > T_{F}$ then

- ➢ No texture information is applicable

      $\mu_{\mathrm{x}}^{0} \leftarrow d_{\mathrm{x}}^{t}$, $\mathcal{P}_{\mathrm{x}}^{t} \leftarrow \mathcal{P}_{init}$;
  else
    $\mathcal{P}_{\mathrm{x}}^{t} \leftarrow \mathcal{P}_{\mathrm{x}}^{t-1}$

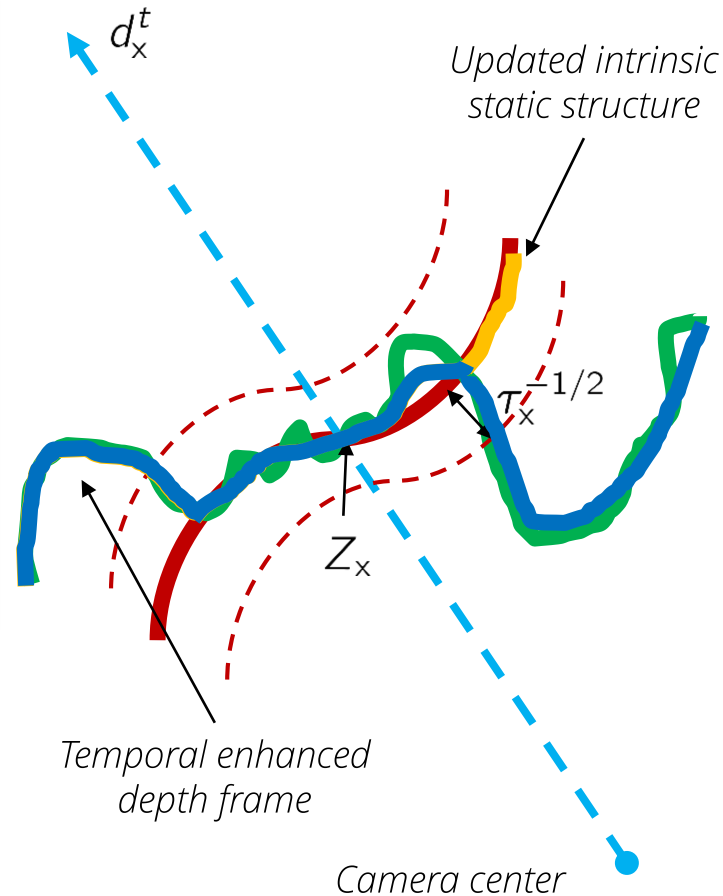
- ➢ When input depth sample is valid, just copy the previous model

Temporal Depth Video Enhancement

- Blend input depth frame/spatially enhanced depth frame with the estimated intrinsic static structure
  - The intrinsic static structure $d^t_{iss,\mathrm{x}} = \mu^t_{\mathrm{x}}$
  - The posterior of **state-I**

    $$\gamma_{iss}(d^t_{\mathrm{x}}) = \alpha^{t-1}_{1,\mathrm{x}} \mathcal{N}(d^t_{\mathrm{x}} | \mu^{t-1}_{\mathrm{x}}, \tilde{\tau}^{t-1}_{\mathrm{x}})^{-1}) / q^t(d^t_{\mathrm{x}})$$
  - The temporally enhanced depth value

    $$d^t_{S,\mathrm{x}} = (1 - \gamma_{iss}(d^t_{\mathrm{x}})) d^t_{\mathrm{x}} + \gamma_{iss}(d^t_{\mathrm{x}}) d^t_{iss,\mathrm{x}}$$

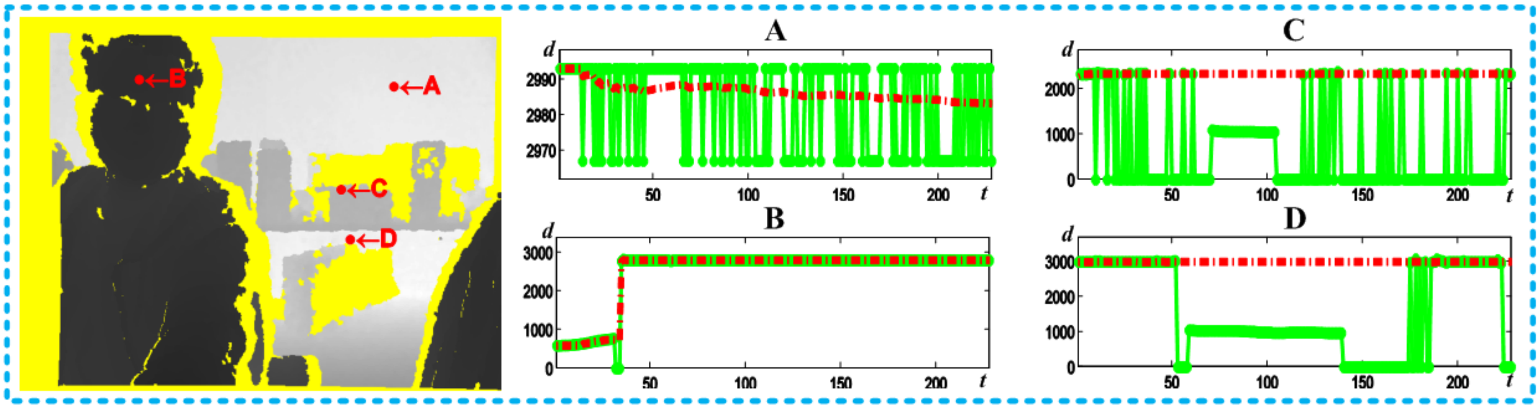


For RGB-D video, Both depth and texture information are helpful to fill depth holes

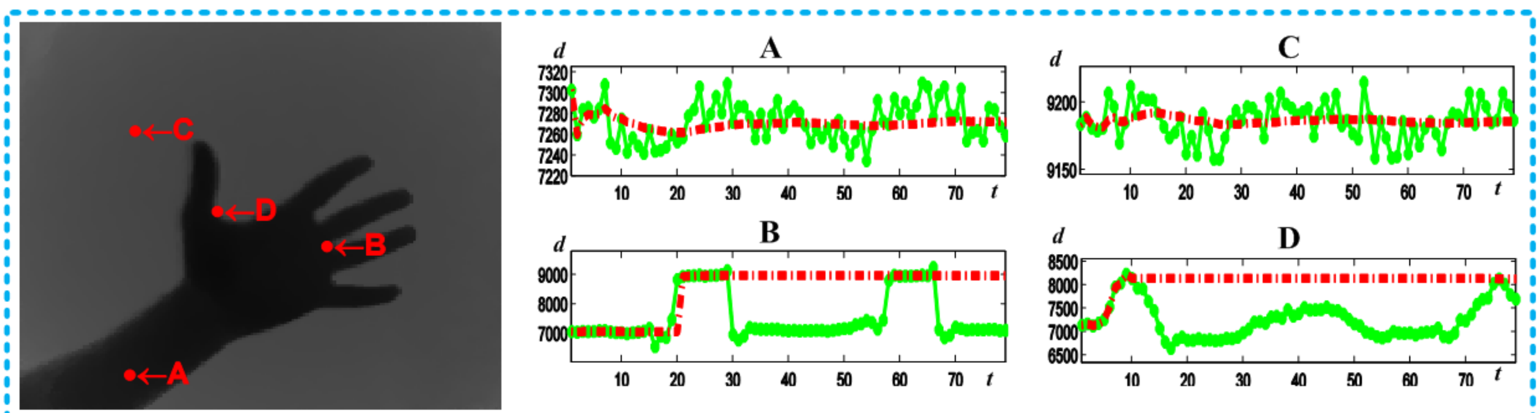
- Inference the chance that the hole region belongs to the intrinsic static structure

Experimental Results

- Intrinsic static structure estimation – 1D case



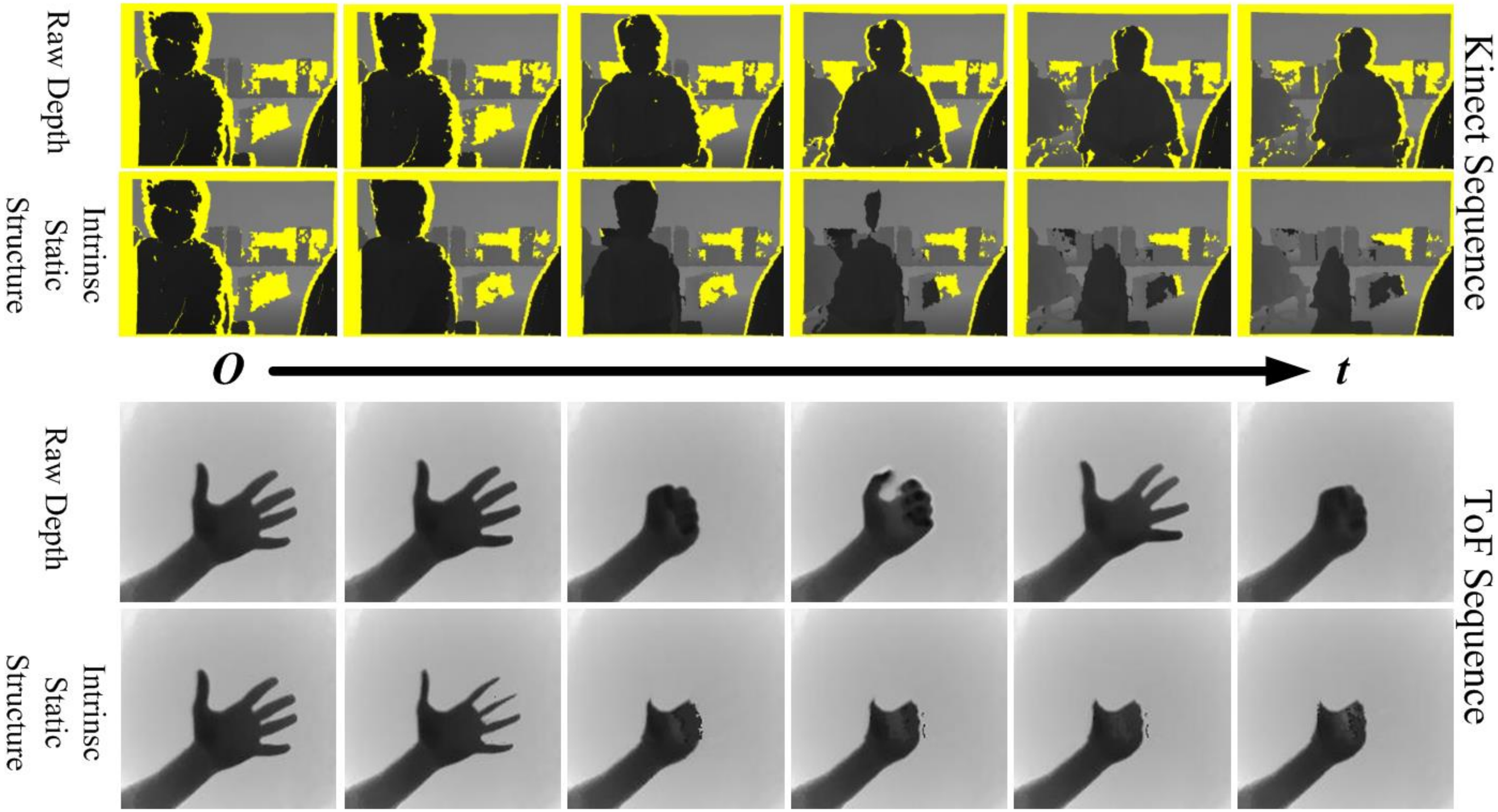
Kinect Sequence



ToF Sequence

Experimental Results

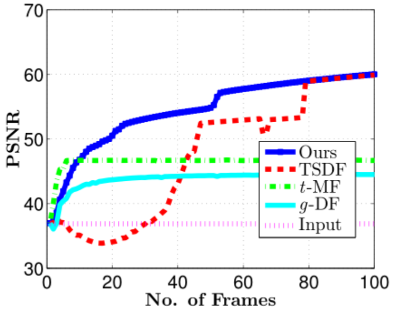
- Intrinsic static structure estimation – 2D case



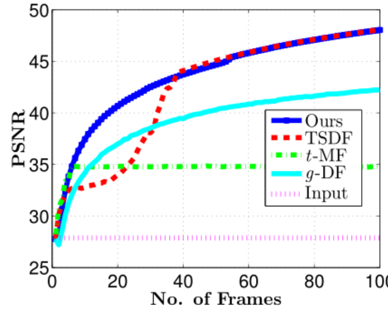
(b) Example: depth videos

Experimental Results

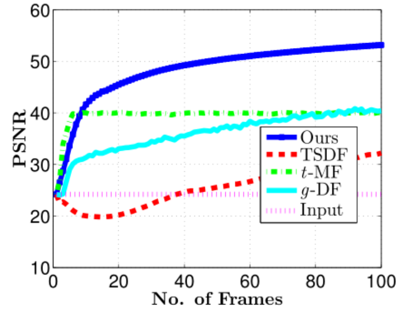
- Temporal enhancement on synthesis static scenes



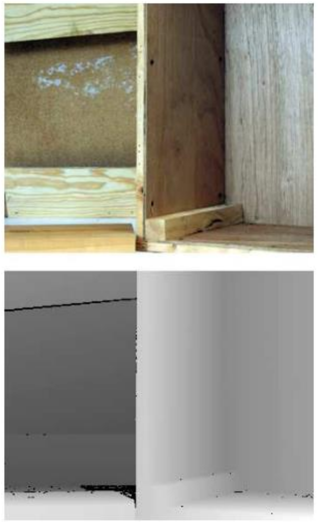
(a) $\omega_d = 10^{-3}, \sigma_d = 2.5$

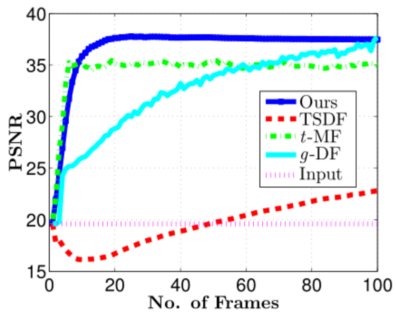


(b) $\omega_d = 10^{-3}, \sigma_d = 10$

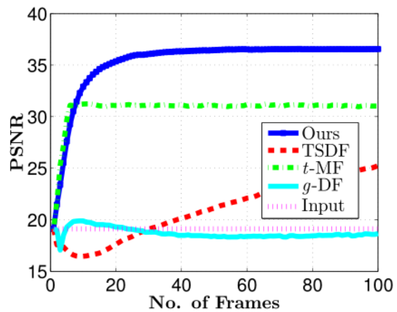


(c) $\omega_d = 5 \times 10^{-2}, \sigma_d = 5$





(d) $\omega_d = 10^{-1}, \sigma_d = 2.5$



(e) $\omega_d = 10^{-1}, \sigma_d = 10$

Input disparity map is contaminated by

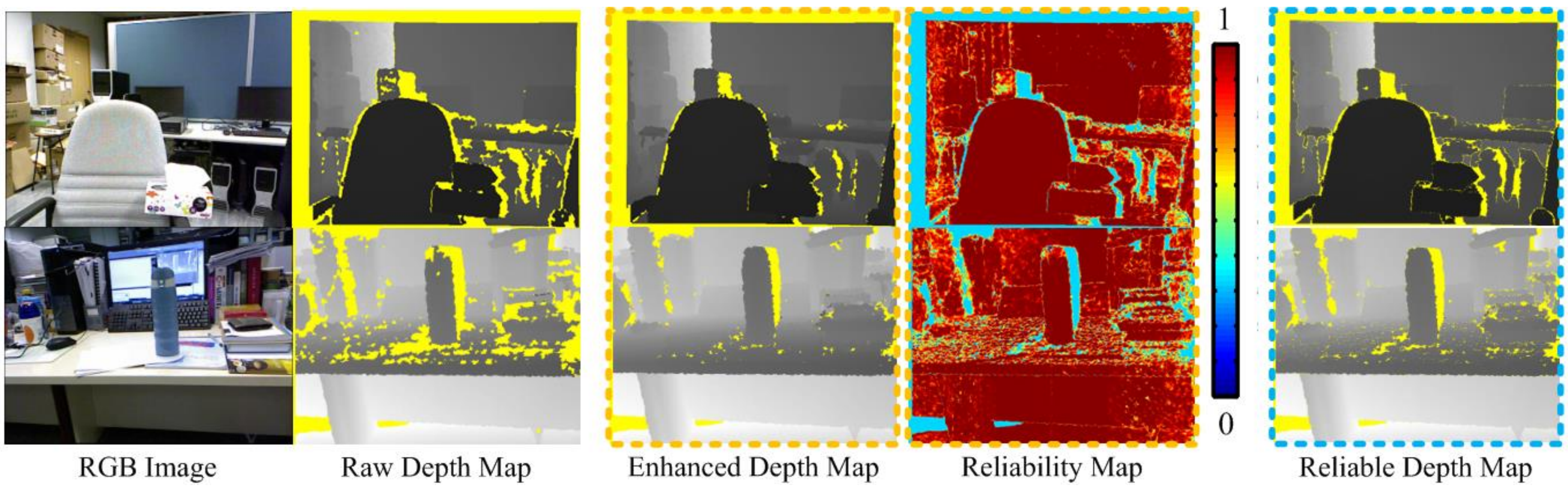
$$p(d_n|Z) = (1 - \omega_d)\mathcal{N}(d_n|Z, \sigma_d^2) + \omega_d \mathcal{U}(d_n)$$

Algorithms	t-MF (w=5/10)	g-DF	TSDF	Ours
Running time (s)	0.0188 / 0.0309	1.9186	0.6847	0.0223

Per-frame Running time comparison
(MALTAB Platform)

Experimental Results

- Temporal enhancement on static scenes

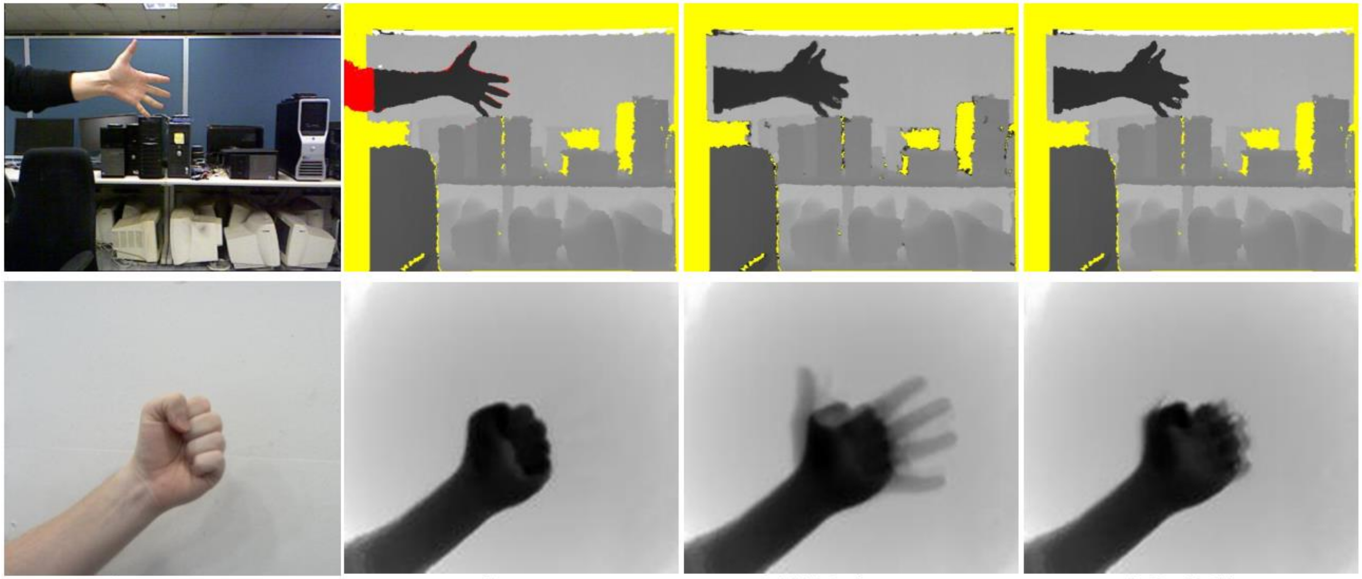


The reliability quantity $r_{\mathrm{x}} = \alpha^t_{1,\mathrm{x}} / \sum_{i=1}^{3} \alpha^t_{i,\mathrm{x}}$

- The reliability map indicates that most of the flat or smooth surfaces of the captured scene are of high reliability to be static
- Around edges or highly slanted surfaces, the reliability map owns lower values

Experimental Results

- Temporal enhancement on dynamic scenes



RGB Image | Ours | *TC*-Enhance | *t*-MedFilter

  - Effectively and accurately find the moving objects according to the **state classification**, with the help of conditional random field and the registered color frame
  - Less motion delay will occur

Experimental Results

Spatiotemporal depth video enhancement

(c) Coherent spatiotemporal JBF

(d) Spatiotemporal WMF

(e) Ours

Conclusion & Future Work

- Temporal flickering problem of conventional depth enhancement approaches
- Introduction of the intrinsic static structures and its application to temporal enhancement of a depth video
- Discuss the its potential to handle the static and dynamic scenes

- Proper way to further combine spatial and temporal enhancement based on the intrinsic static structure into a whole framework
- Research on a more general formulation of the static structure
- How to deal with depth video captured by moving cameras, and etc.

Thank you!

Any Questions?

On-line Update Scheme

- Variational Parameter Estimation $\mathcal{P}_{\mathrm{x}}^t = \{\mu_{\mathrm{x}}^t, \lambda_{\mathrm{x}}^t, \boldsymbol{\alpha}_{\mathrm{x}}^t\}$
  - Factorize the posterior into independent Gaussian and Dirichlet distributions

$$q^t(Z_{\mathrm{x}}) = \mathcal{N}(Z_{\mathrm{x}} | \mu_{\mathrm{x}}^t, (\lambda_{\mathrm{x}}^t)^{-1})$$

$$q^t(\boldsymbol{\omega}_{\mathrm{x}}) = \mathrm{Dir}(\boldsymbol{\omega}_{\mathrm{x}} | \alpha_{1,\mathrm{x}}^t, \alpha_{2,\mathrm{x}}^t, \alpha_{3,\mathrm{x}}^t)$$

$$q^t(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}}) = q^t(Z_{\mathrm{x}}) q^t(\boldsymbol{\omega}_{\mathrm{x}}) \sim p(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}} | \mathcal{D}_{\mathrm{x}}^t)$$

    - The reliability of the model $\mathbb{E}_{p(\boldsymbol{\omega}_{\mathrm{x}}|\mathcal{D}_{\mathrm{x}}^t)}[\boldsymbol{\omega}_{\mathrm{x}}] \approx \mathbb{E}_{q^t(\boldsymbol{\omega}_{\mathrm{x}})}[\boldsymbol{\omega}_{\mathrm{x}}]$
    - The probable depth is $\mathbb{E}_{p(\boldsymbol{\omega}_{\mathrm{x}}|\mathcal{D}_{\mathrm{x}}^t)}[\boldsymbol{\omega}_{\mathrm{x}}] \approx \mu_{\mathrm{x}}^t$

  - The posterior can be approximated by

$$\boxed{q^t(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}})} \sim \frac{p(d_{\mathrm{x}}^t | Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}}) \boxed{q^{t-1}(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}})}}{q^t(d_{\mathrm{x}}^t)} = Q(Z_{\mathrm{x}}, \boldsymbol{\omega}_{\mathrm{x}} | d_{\mathrm{x}}^t)$$

Recursive estimation is possible!

On-line Update Scheme

- Variational Parameter Estimation $\mathcal{P}_x^t = \{\mu_x^t, \lambda_x^t, \boldsymbol{\alpha}_x^t\}$
  - Factorize the posterior into independent Gaussian and Dirichlet distributions

$$q^t(Z_x, \boldsymbol{\omega}_x) = q^t(Z_x) q^t(\boldsymbol{\omega}_x) \sim p(Z_x, \boldsymbol{\omega}_x | \mathcal{D}_x^t)$$

$$q^t(Z_x) = \mathcal{N}(Z_x | \mu_x^t, (\lambda_x^t)^{-1})$$

$$q^t(\boldsymbol{\omega}_x) = \mathrm{Dir}(\boldsymbol{\omega}_x | \alpha_{1,x}^t, \alpha_{2,x}^t, \alpha_{3,x}^t)$$

  - The posterior can be approximated by

$$q^t(Z_x, \boldsymbol{\omega}_x) \sim \frac{p(d_x^t | Z_x, \boldsymbol{\omega}_x) q^{t-1}(Z_x, \boldsymbol{\omega}_x)}{q^t(d_x^t)} = Q(Z_x, \boldsymbol{\omega}_x | d_x^t)$$

  - Moment matching to estimate the hyperparameters

Closed-form solutions!

$$q^t(Z_x) : \begin{cases} \mu_x^t = \mathbb{E}_{Q(Z_x|d_x^t)}[Z_x] \\ \lambda_x^t = \dfrac{1}{\mathbb{E}_{Q(Z_x|d_x^t)}[Z_x^2] - \mathbb{E}_{Q(Z_x|d_x^t)}[Z_x]^2} \end{cases}$$

$$q^t(\boldsymbol{\omega}_x) : \begin{cases} \alpha_{i,x}^t = \mathbb{E}_{Q(\omega_x^i|d_x^t)}[\omega_i] \cdot \alpha_{o,x}^t, \quad i = 1, 2, 3 \\ \alpha_{o,x}^t = \dfrac{\sum_{i=1}^3 \mathbb{E}_{Q(\omega_x^i|d_x^t)}[\omega_x^i] - \mathbb{E}_{Q(\omega_x^i|d_x^t)}[(\omega_x^i)^2]}{\sum_{i=1}^3 \mathbb{E}_{Q(\omega_x^i|d_x^t)}[(\omega_x^i)^2] - \mathbb{E}_{Q(\omega_x^i|d_x^t)}[\omega_x^i]^2} \end{cases}$$