A Generative Model for Online Depth Fusion

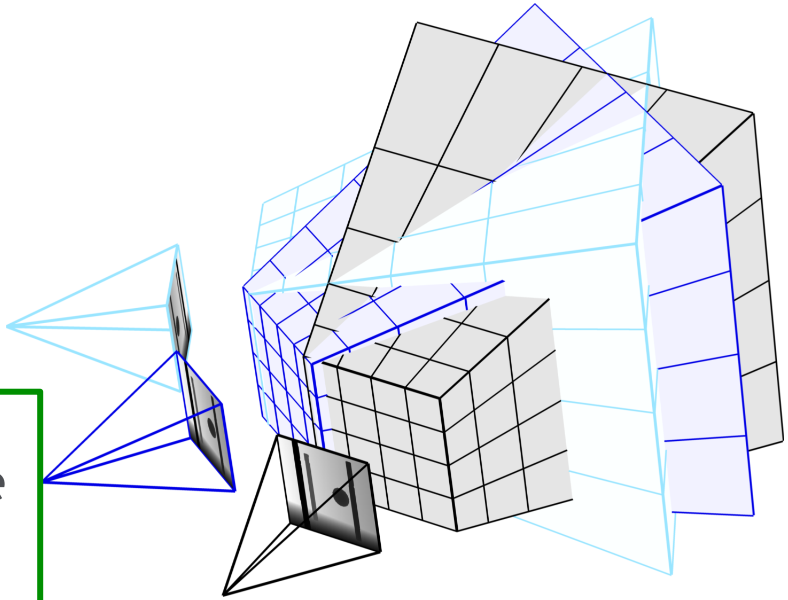
George Vogiatzis (Aston University) and
Oliver Woodford (Toshiba Research)

Background

- Plethora of depth-measuring technologies
  - binocular/multi-view stereo
  - Structured light stereo (e.g. Kinect)
  - Sonar
  - Time-of-flight
  - Laser
  - …

Depth-map fusion

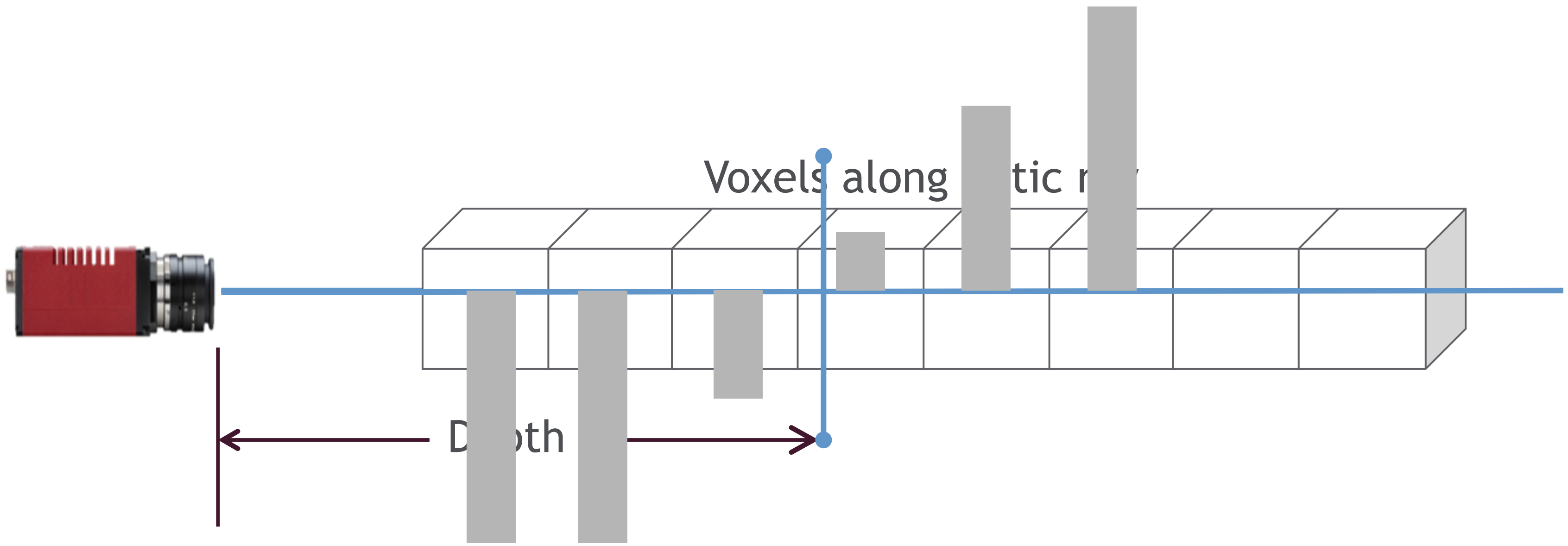
- Convert depth-maps to scene geometry
  - Crucial problem
- Offline fusion: collect all depth-maps THEN merge
  - Point-cloud
  - Octree
  - TSDF
- Online fusion: merge each incoming depth-map into state
  - Forward/inverse sensor modeling (Robotics)
  - TSDF (KinectFusion)



Aston University
Birmingham

TOSHIBA
Leading Innovation >>>

Truncated Signed Distance Functions



- Shown to be equivalent to accumulating probabilistic *evidence* of visibility (log-odds)
- Under a logistic sensor noise model ($\mathrm{sech}(x)^2$)
- No account of outlier measurements

Aston University
Birmingham

TOSHIBA
Leading Innovation

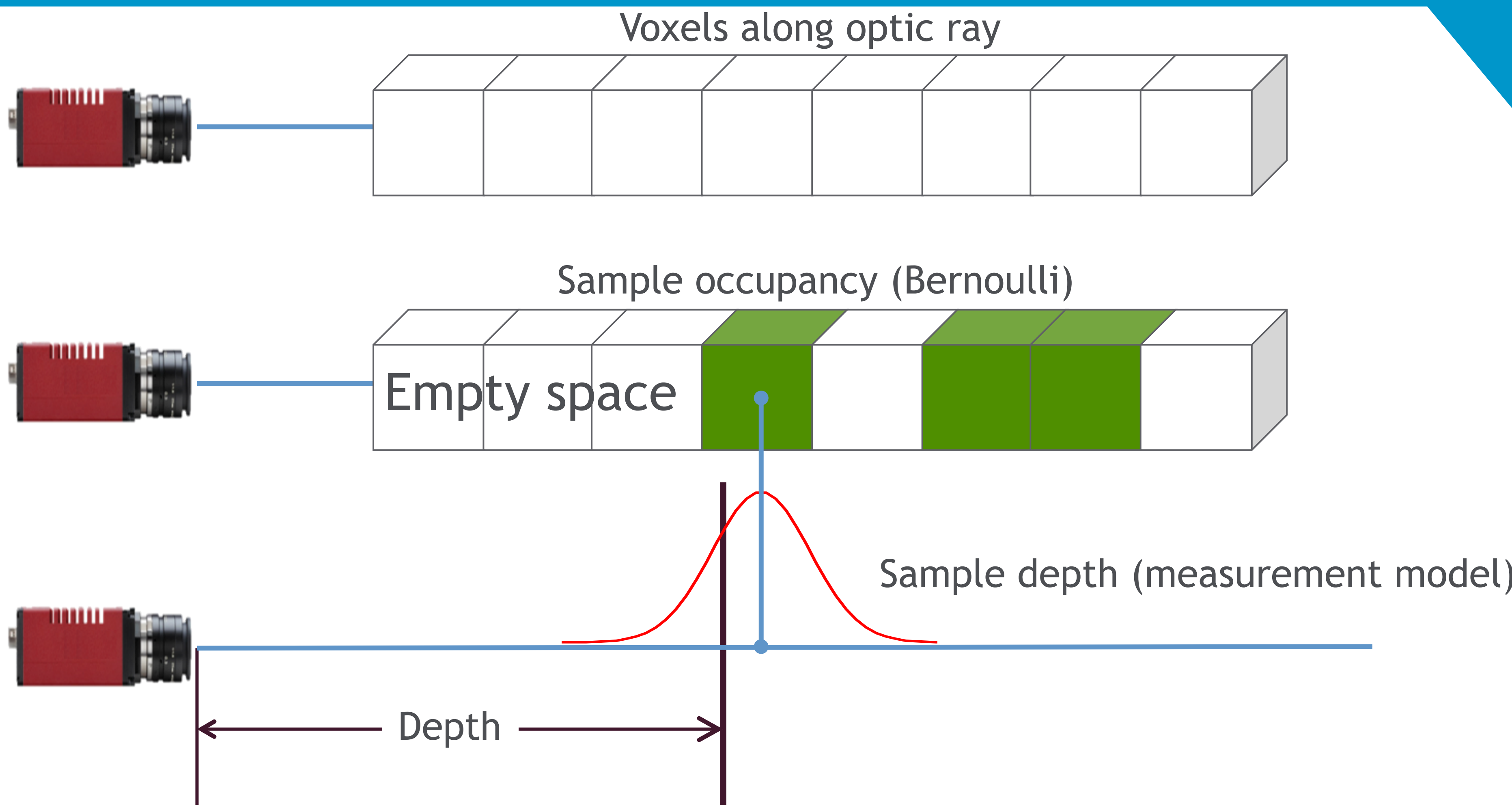
Robotics online depth fusion

- Inverse models (Elfes & Matthies 87, Konolige 97)
  - Model directly p(occupancy|depth)
  - No inter-depencdency of occupancy variables along an optic ray (free-space constraints)
- Forward models (Thrun 01, Pathak 07)
  - Model p(depth|visibility)
  - Assume occupancy is driven by visibility
  - Cannot model occlusion

Aston University
Birmingham

TOSHIBA
Leading Innovation

Generative model of depth measurement



Aston University
Birmingham

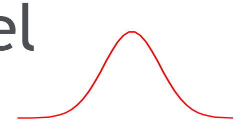
TOSHIBA
Leading Innovation

Generative model of depth measurement

- Occupancy $\mathbf{x} = \{x_i\}_{i=1}^{N},\ x_i \in \{0, 1\}$
- Depth measurement $y \in \mathbb{R}$
- Visible voxel index $v = 1, 2, \ldots$

$$p\left(y|\mathbf{x}\right) = \sum_v \boxed{p\left(y|v\right)}\,\boxed{p\left(v|\mathbf{x}\right)}$$

Measurement model



Is 1 iff v is index of first occupied voxel

Aston University
Birmingham

TOSHIBA
Leading Innovation

Inferring occupancy

- Prior $p(x)$ factorizes
- BUT, posterior $p(x|y)$ doesn't!
- Not feasible to maintain full covariance in online fusion
- Our approach:
  - Assume factored approximation $q(x)$
  - Minimize $KL(p||q)$ (Expectation Propagation)
  - Amounts to computing $p(x_i|y)$ marginals
    - Tractable!

Inferring occupancy

- Depth likelihood depends on occupancies through index of visible voxel

$$p\left(v\right) = p\left(x_v\right) \prod_{i=1}^{v-1} \left(1 - p\left(x_i\right)\right)$$

$$p\left(v|y\right) = p\left(y|v\right) \frac{p\left(v\right)}{p\left(y\right)}$$

$$p\left(x_i|v\right) = \begin{cases} 1 & v = i \\ 0 & v > i \\ p\left(x_i\right) & v < i \end{cases}$$

$$\begin{aligned} p\left(x_i|y\right) &= \sum_v p\left(x_i v|y\right) \\ &= \sum_v \underline{p\left(x_i|v\right)}\, \underline{p\left(v|y\right)} \\ &= \underline{p\left(v = i|y\right)} + p\left(x_i\right) \sum_{v=1}^{i-1} \underline{p\left(v|y\right)} \end{aligned}$$

Aston University
Birmingham

TOSHIBA
Leading Innovation

Outlier measurements

- Can use a simple noise+outlier model

$$p\left(y|v\right)$$

Aston University
Birmingham

TOSHIBA
Leading Innovation >>>

Outlier measurements

- Can use a simple noise+outlier model

$$p(y|v,\underbrace{\boxed{\omega}}_{\text{Outlier ratio}}) = \omega \cdot \underbrace{\boxed{\mathcal{C}(y)}}_{\text{Clutter dist.}} + (1-\omega) \cdot \underbrace{\boxed{\mathcal{M}_v(y)}}_{\text{Noise dist.}}$$

- Assume ω comes from Beta(α,β) hyper-prior

Full model

- Factorized prior $$p\left(\mathbf{x},\omega\right)=p\left(\mathbf{x}\right)p\left(\omega\right)$$

$$p(\mathbf{x})=\prod_{i=1}^{N}p(x_i|\gamma_i),\qquad p(x|\gamma)=\gamma^x(1-\gamma)^{1-x},$$

$$p(\boldsymbol{\omega})=\prod_{i=1}^{N}p(\omega_i|\alpha_i,\beta_i),\qquad p(\omega|\alpha,\beta)=\frac{\omega^{\alpha-1}(1-\omega)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$$

- Noise + outlier likelihood

$$p(y|v,\omega)=\omega\cdot\mathcal{C}(y)+(1-\omega)\cdot\mathcal{M}_v(y)$$

Inference

- Assume factored approximation

$$q(\mathbf{x}, \boldsymbol{\omega}) = \prod_{i=1}^{N} q_i(x_i) \prod_{j=1}^{N} q_j(\omega_j)$$

- Minimize KL divergence between

  $p(\mathbf{x}, \boldsymbol{\omega}|y)$ and $q(\mathbf{x}, \boldsymbol{\omega})$

- Matching sufficient statistics (~EP)

$$\mathbb{E}_{q(\mathbf{x},\boldsymbol{\omega})}[x_i] = \mathbb{E}_{p(\mathbf{x},\boldsymbol{\omega}|y)}[x_i],$$
$$\mathbb{E}_{q(\mathbf{x},\boldsymbol{\omega})}[\ln \omega_i] = \mathbb{E}_{p(\mathbf{x},\boldsymbol{\omega}|y)}[\ln \omega_i],$$
$$\mathbb{E}_{q(\mathbf{x},\boldsymbol{\omega})}[\ln(1-\omega_i)] = \mathbb{E}_{p(\mathbf{x},\boldsymbol{\omega}|y)}[\ln(1-\omega_i)]$$

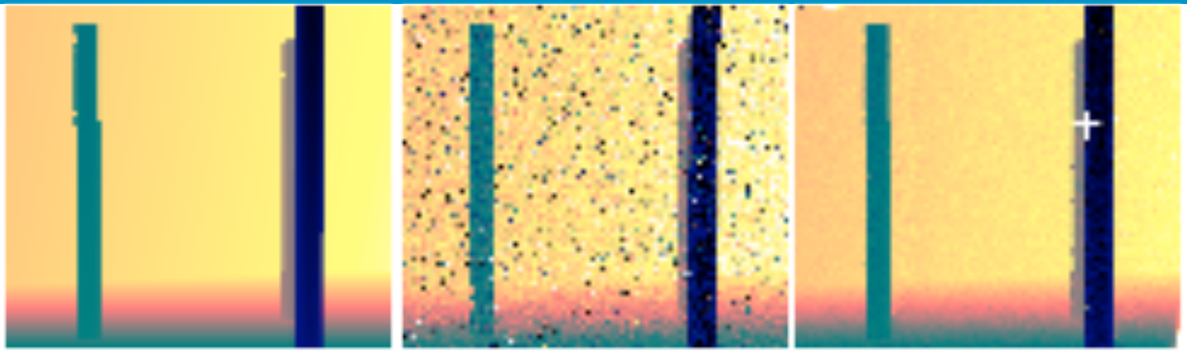
Modelling options

- Outlier ratio
  - One fixed ω (generative1)
  - One ω that is estimated from data (generative2)
  - Multiple ω, one per optic ray or per voxel, estimated from data (generative3)

- Online fusion
  - First order independence assumptions plus appearing and disappearing surfaces
  - Similar to ‘forgetting factor’ in TSDF

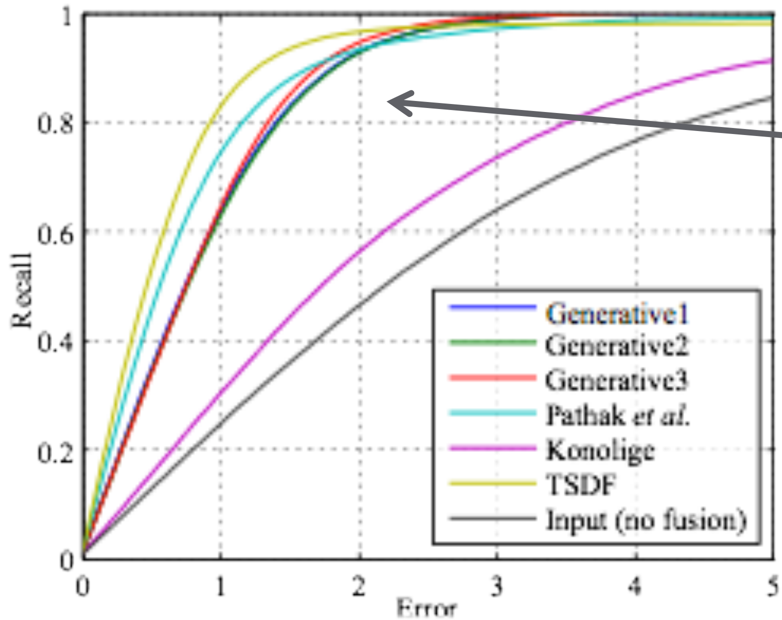
- How to output depth: use q(v)

Ground truth benchmarking



Ground truth Noise added Fused output

Error/recall curve

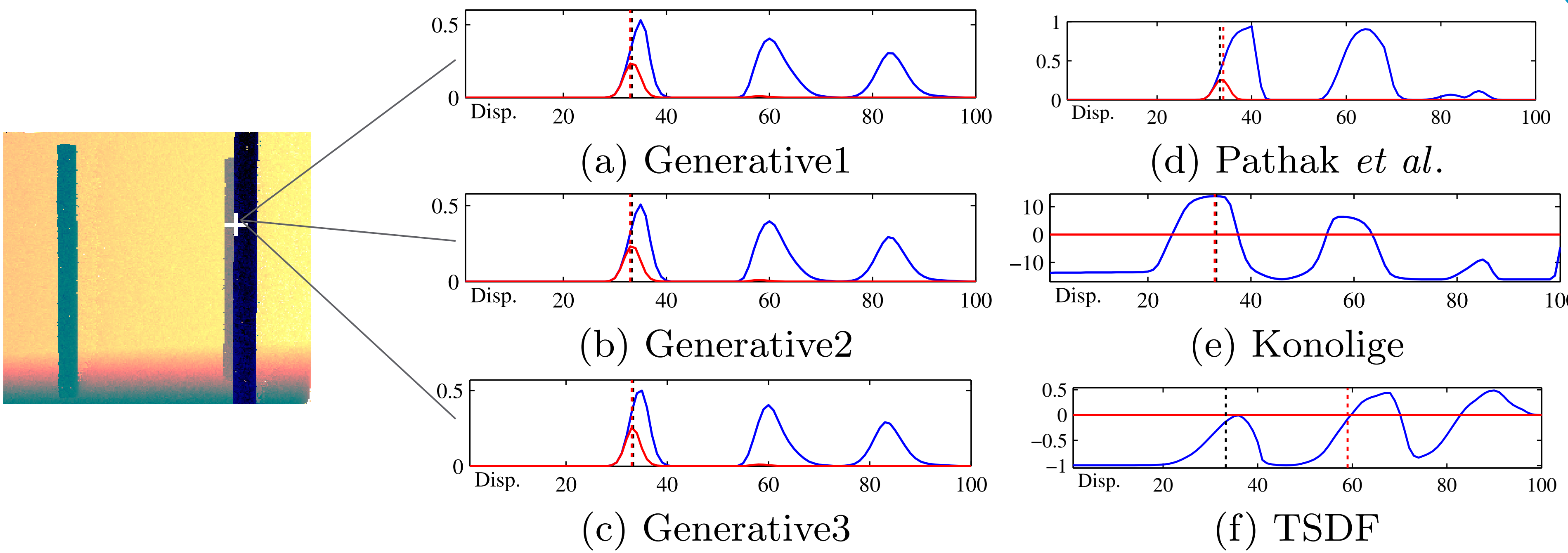


Score=area under curve normalized to 1.0

Aston University
Birmingham

TOSHIBA
Leading Innovation

Occupancy vs visible surface along a ray

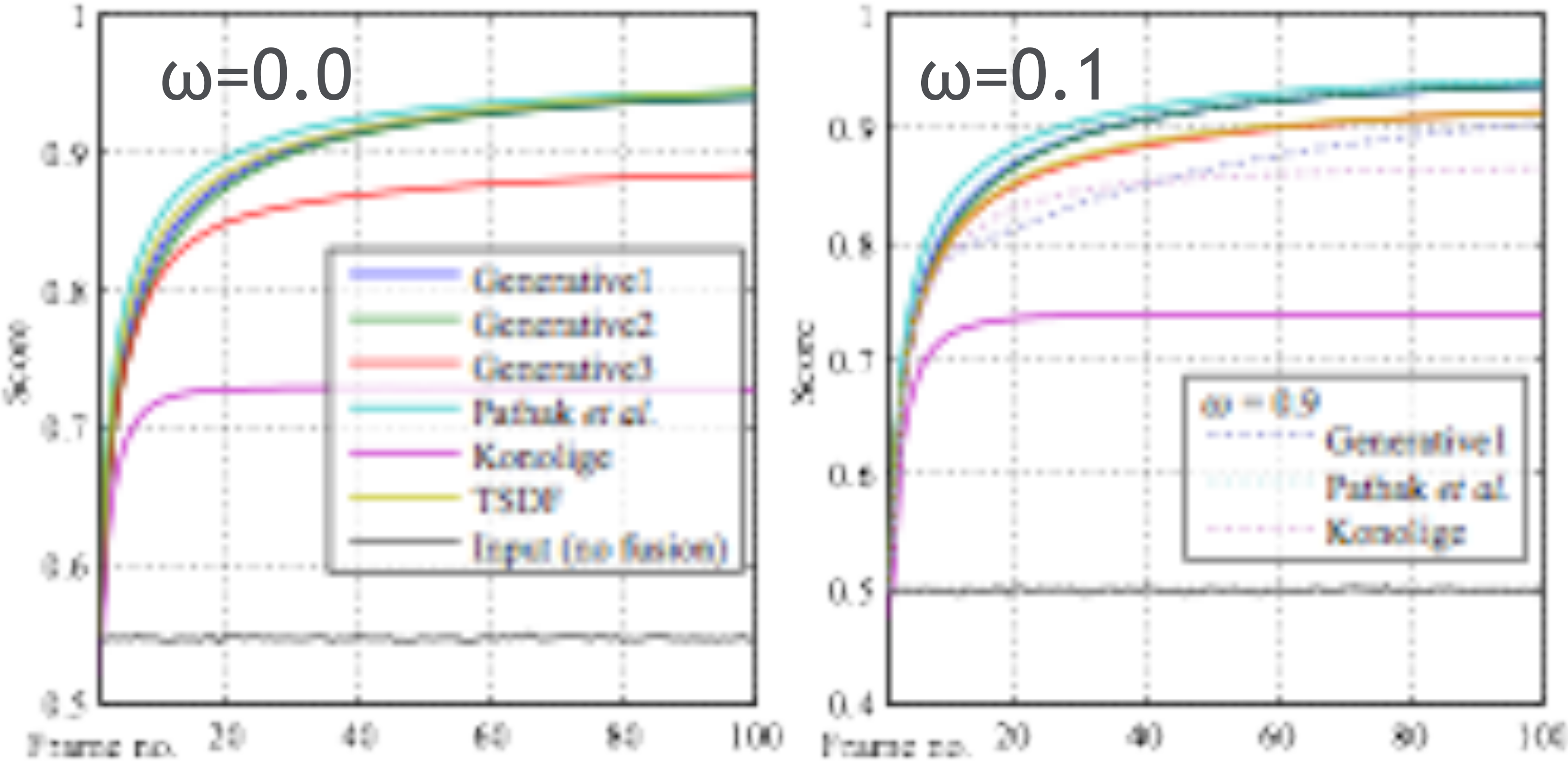


- More meaningful occupancy values, with better defined maxima

Aston University Birmingham

TOSHIBA Leading Innovation

Effect of outlier modelling

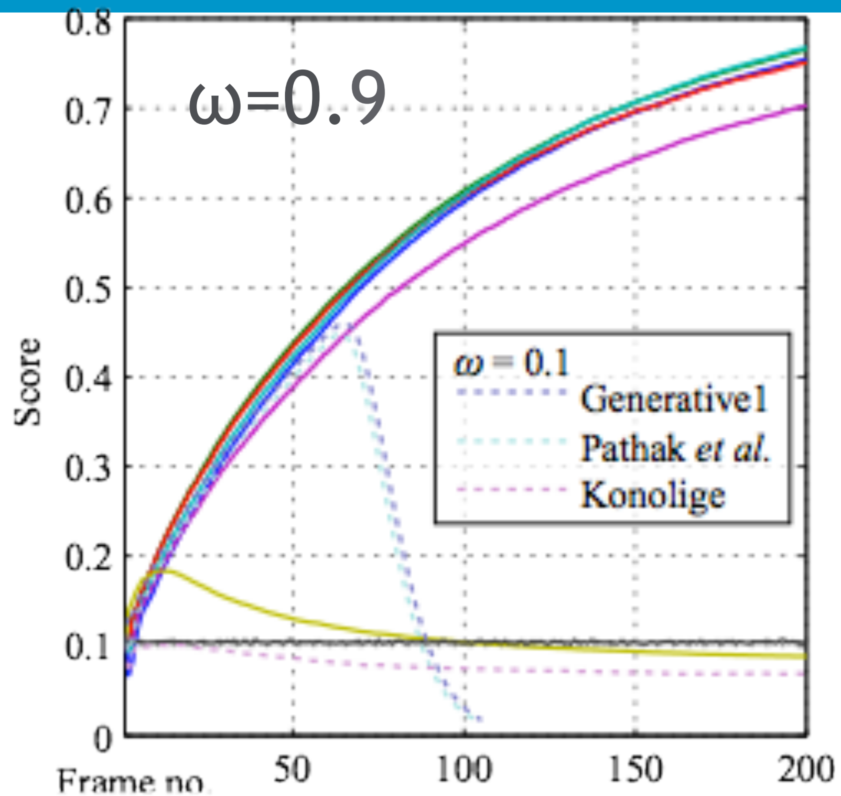


- Not much effect in low outlier regimes
- Generative2 (estimate one ω) is good performance/ computation compromise

Aston University
Birmingham

TOSHIBA
Leading Innovation >>>

Effect of outlier modelling

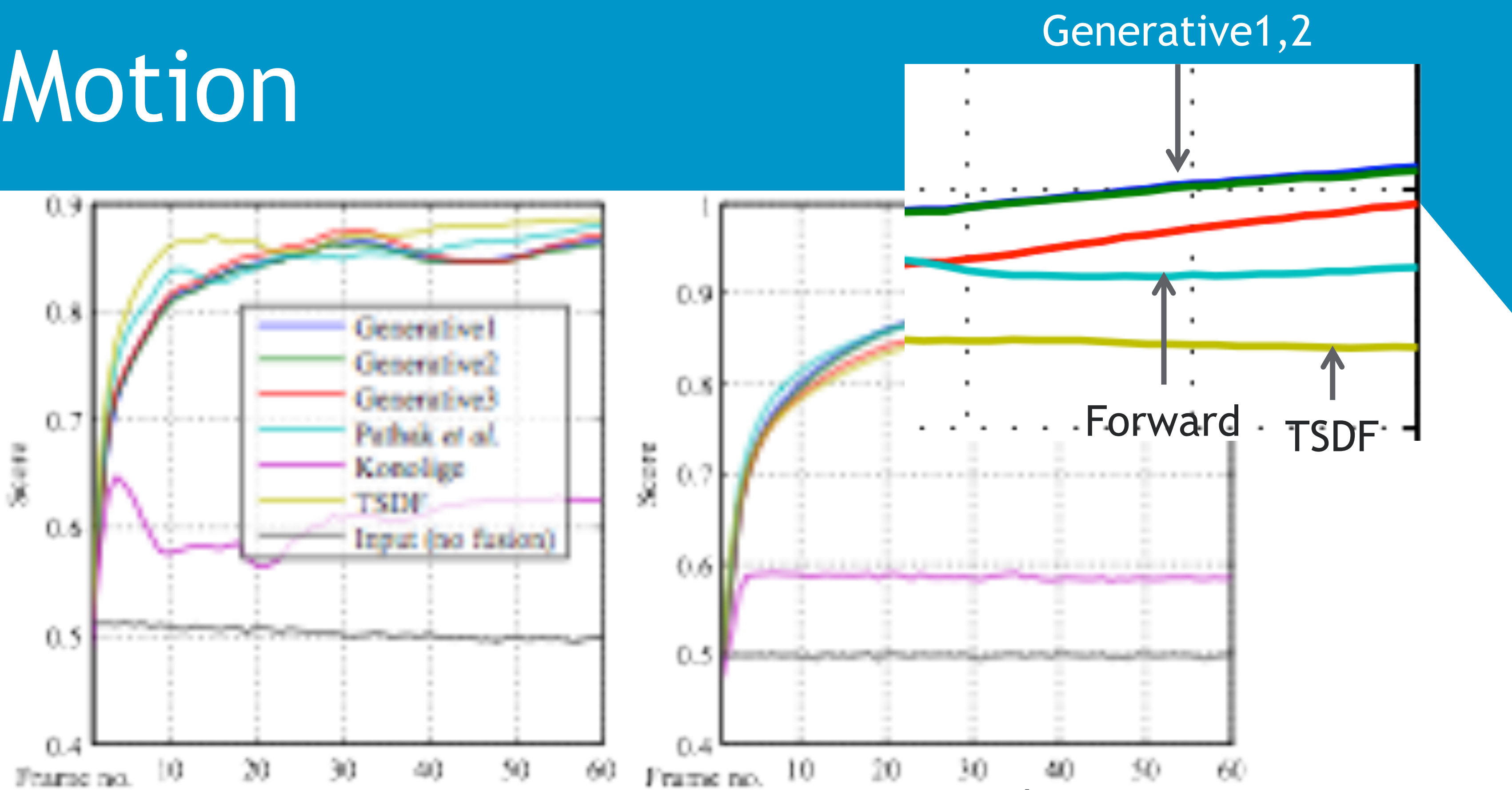


- Quite significant in heavy-outlier regime
- TSDF cannot cope
- Methods that **model** outliers but do not **estimate** ω also do poorly

Aston University
Birmingham

TOSHIBA
Leading Innovation >>>

Motion



Moving sensor

Moving object

- Faster reaction time to occlusions/disocclusions
- Due to more accurate occlusion reasoning

Aston University
Birmingham

TOSHIBA
Leading Innovation >>>

Ground truth benchmarking

Our method: Fixed ω | Estimate one ω | Estimate multiple ω

Forward model | Inverse model | Kinect fusion

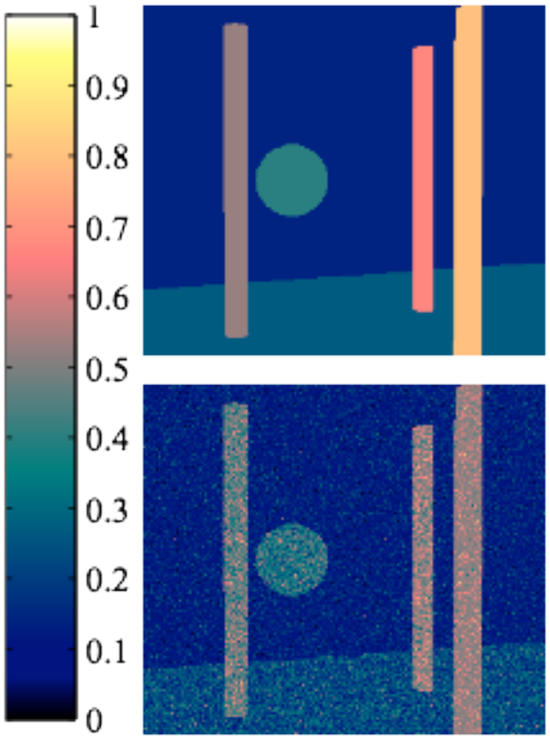
	ϖ	Generative1	Generative2	Generative3	Pathak *et al.*	Konolige	TSDF
Static sensor	0	0.065±0.34	0.069±0.33	0.36±0.62	**0.000**±0.33	1.37±0.25	-0.17±0.23
	0.1	0.064±0.35	0.059±0.34	0.19±0.45	**-0.002**±0.33	1.32±0.25	-0.37±0.31
	0.4	0.090±0.39	0.073±0.39	0.070±0.40	**-0.002**±0.37	1.18±0.33	-7.33±18.0
	0.9	0.60±7.08	**0.24**±6.79	0.45±6.94	0.79±7.91	3.15±11.0	-23.7±20.0
Moving sensor	0	0.35±0.97	0.33±1.09	0.35±1.43	**0.27**±0.86	2.09±8.49	-0.59±3.53
	0.1	0.35±2.13	**0.34**±2.16	0.36±2.20	0.42±2.77	3.69±14.1	-0.69±4.35
	0.4	0.50±3.85	**0.41**±4.09	0.48±3.87	0.74±4.86	6.90±21.1	-1.52±6.87
	0.9	**2.96**±11.1	8.00±18.4	3.39±12.7	3.11±11.1	20.9±32.2	-10.1±12.6

- Smaller errors at high ω regime
- TSDF struggles

Aston University
Birmingham

TOSHIBA
Leading Innovation

Estimating different outlier ratios



Ground truth ω

Estimated ω

- Can identify regions in the scene that produce different sensor response (e.g. shiny/textureless)
- First stage of scene classification scheme

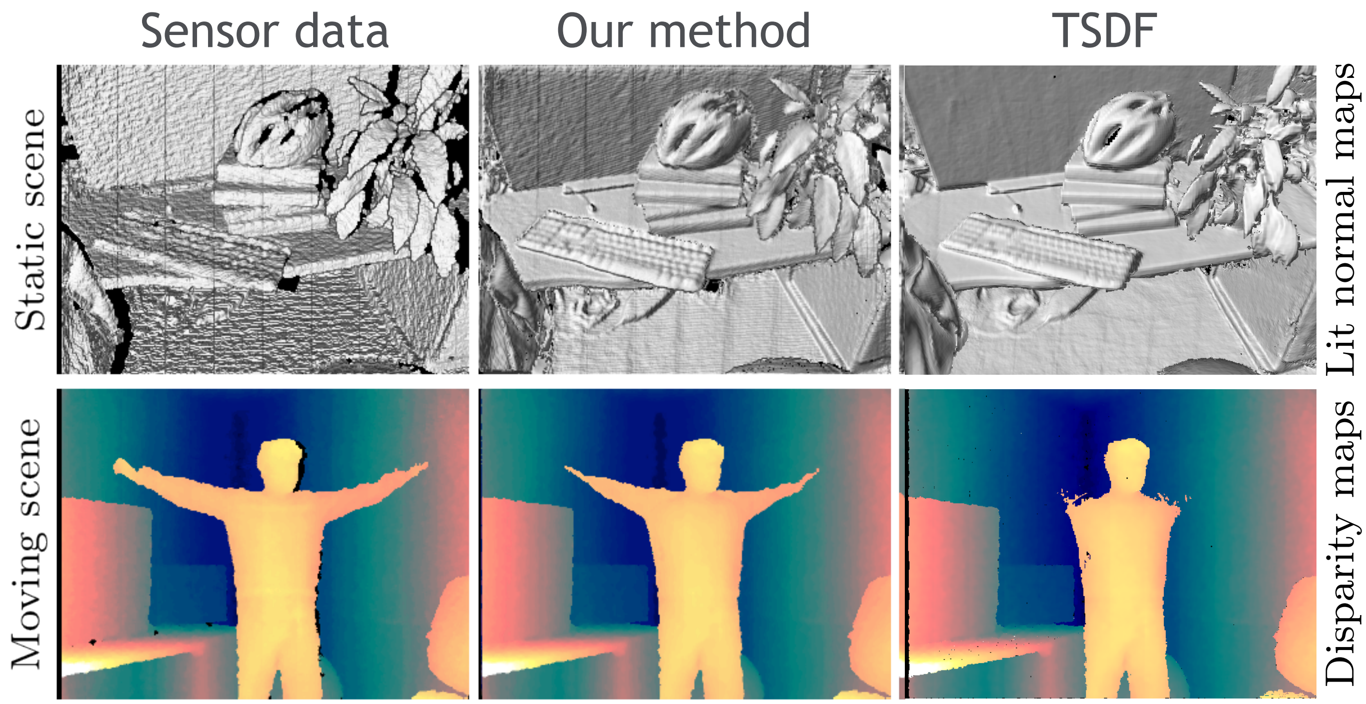
Aston University
Birmingham

TOSHIBA
Leading Innovation

Kinect – static scene/moving sensor

Kinect – moving scene/static sensor

Kinect data

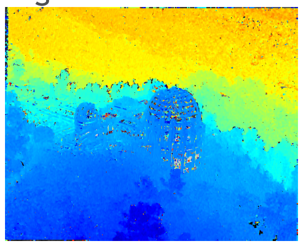

Sensor data
Our method
TSDF
Static scene
Moving scene
Lit normal maps
Disparity maps

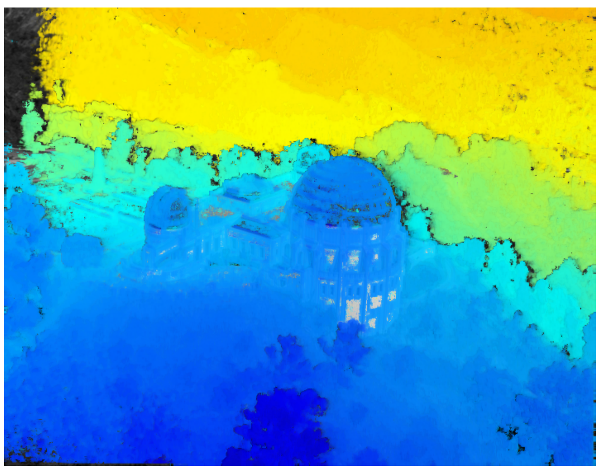
Multi-view stereo data

Multi-view stereo data

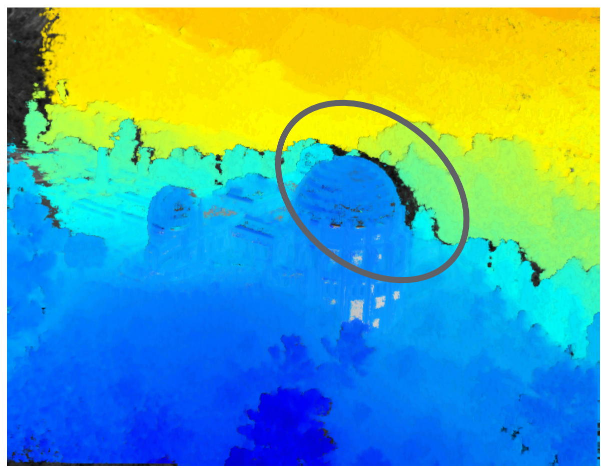
Input

Estim. Depth using Vogiatzis&Hernandez 2011

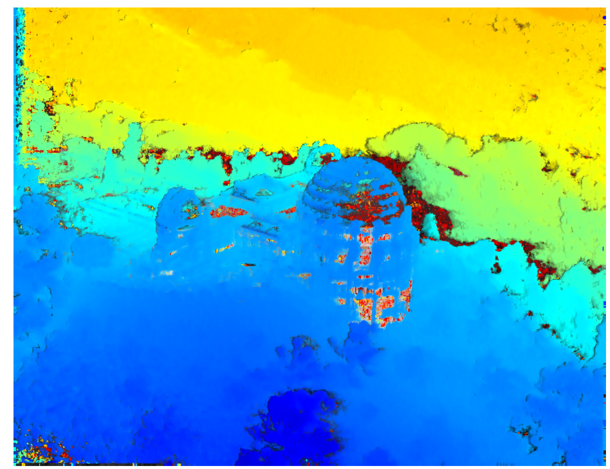




Our method



Forward model

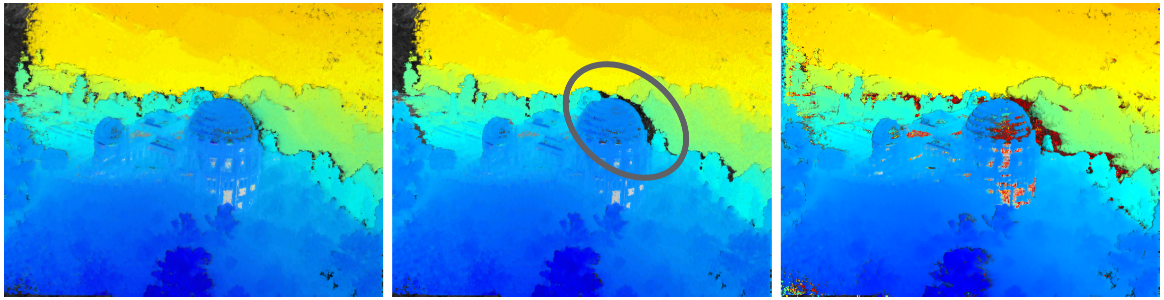


TSDF

Aston University Birmingham

TOSHIBA Leading Innovation

Take home messages



Our method Forward model TSDF

- Use more realistic sensor modelling (e.g. outliers)
- Generative models react faster to scene changes
- Our occlusion model => better reconstruction at depth discontinuities
- Inferring outlier ratios helps, but significantly only in extreme cases
  - interesting potential for scene classification
- Volume resampling is slow (~9fps in KinectFusion) but allows zooming

Many more experimental results in ECCV'12 paper

Thanks!

Aston University
Birmingham

TOSHIBA
Leading Innovation >>>