EXTRINSIC CAMERA PARAMETERS ESTIMATION FOR SHAPE-FROM-DEPTHS

Jonathan Ruttle, Claudia Arellano and **Rozenn Dahyot**

Trinity College Dublin, Ireland

European Signal Processing Conference

August 30th , 2012

Problem: 3D reconstruction of an object from Kinect scans I

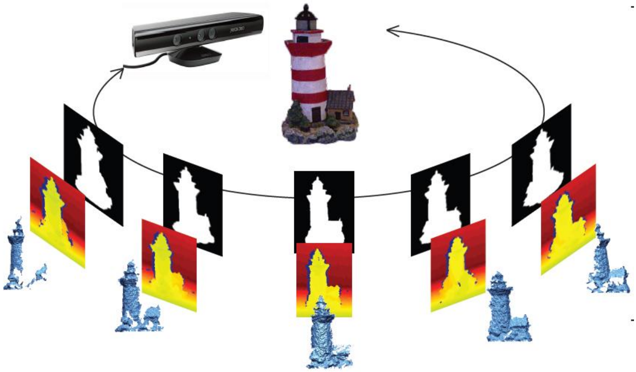


Figure: 3D reconstruction of an object using the kinect and turning table.

Problem: 3D reconstruction of an object from Kinect scans II

Objective: Inferring an accurate 3D surface of the object from a set of C noisy kinect scans

Strategy:

- Define a smooth differentiable cost function to estimate the surface of the object,
- The cost function should be robust,
- The uncertainty associated with each observation should be taken into account.

Standard approach I

Observations: $\mathbf{x}_c^{(i)} = (x_{1,c}^{(i)}, x_{2,c}^{(i)}, x_{3,c}^{(i)})$ have been collected:

- $x_{1,c}^{(i)}$ x-coordinate of the pixel i in the scan recorded by camera c
- $x_{2,c}^{(i)}$ y-coordinate of the pixel i in the scan recorded by camera c
- $x_{3,c}^{(i)}$ depth value of the pixel i in the scan recorded by camera c

Camera model

$$P(\Psi) = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \left[\begin{array}{c|c} & \psi_4 \\ R(\Psi) & \psi_5 \\ & \psi_6 \end{array} \right] \tag{1}$$

Standard approach II

The coordinate of the centre of the camera $C(\Psi)$ can then be computed with:

$$C(\Psi) = -\begin{bmatrix} \quad R(\Psi) \quad \end{bmatrix}' \begin{bmatrix} \psi_4 \\ \psi_5 \\ \psi_6 \end{bmatrix} \tag{2}$$

Standard approach III

Link function: The function $F(\mathbf{x}, \Psi, \Theta)$ that links the observation $\mathbf{x}$, the corresponding 3D position Θ, and the camera parameters Ψ is defined as:

$$F(\mathbf{x}, \Psi, \Theta) = \begin{pmatrix} F_1(\mathbf{x}, \Psi, \Theta) \\ F_2(\mathbf{x}, \Psi, \Theta) \\ F_3(\mathbf{x}, \Psi, \Theta) \end{pmatrix} =$$

$$\begin{pmatrix} x_1 - \frac{\theta_1 P(\Psi)_{11} + \theta_2 P(\Psi)_{12} + \theta_3 P(\Psi)_{13} + P(\Psi)_{14}}{\theta_1 P(\Psi)_{31} + \theta_2 P(\Psi)_{32} + \theta_3 P(\Psi)_{33} + P(\Psi)_{34}} \\ x_2 - \frac{\theta_1 P(\Psi)_{21} + \theta_2 P(\Psi)_{22} + \theta_3 P(\Psi)_{23} + P(\Psi)_{24}}{\theta_1 P(\Psi)_{31} + \theta_2 P(\Psi)_{32} + \theta_3 P(\Psi)_{33} + P(\Psi)_{34}} \\ x_3 - \sqrt{(C(\Psi)_1 - \theta_1)^2 + (C(\Psi)_2 - \theta_2)^2 + (C(\Psi)_3 - \theta_3)^2} \end{pmatrix} = 0$$

F is non linear in Θ and Ψ.

Standard approach IV

Standard approach: [e.g. KinectFusion]

- Filtering of the kinect scans (removing the noise)
- Estimation of camera parameters,
- Conversion of the scans to point clouds: $\mathbf{x}_c^{(i)} \rightarrow \Theta_c^{(i)}$

Problem:

- Filtering remove noise (good) and smooth the surface (bad)
- Difficult to model the uncertainty on $\Theta_c^{(i)}$ while uncertainty about $\mathbf{x}_c^{(i)}$ is known.

Our Robust Modelling I

Stochastic equation:

$$\lambda + F(\mathbf{x}, \Theta, \Psi) = \epsilon \sim p_\epsilon(\epsilon)$$

- λ is an additive auxiliary variable and the conditional density of λ given $\mathbf{x}$, Θ and Ψ can then be written:

$$p_{\lambda|\Theta\Psi\mathbf{x}}(\lambda|\Theta, \Psi, \mathbf{x}) = p_\epsilon(\lambda + F(\mathbf{x}, \Theta, \Psi))$$

  The case of interest is when $\lambda = 0$.

- $\mathbf{x}$ is the random variable associated with the observations $\{\mathbf{x}_c^{(i)}\}$

Our Robust Modelling II

- $\Theta \in \mathbb{R}^3$ is the latent information of interest.
- Ψ are the camera parameters (nuisance parameters) and F is the link function.
- $\epsilon \in \mathbb{R}^3$ is the noise and p_ϵ is chosen Gaussian with a diagonal covariance matrix with bandwidths $h_1 = h_2 = 1$ (pixel precision) and $h_3 = 0.002m$ (depth precision).

Our Robust Modelling III

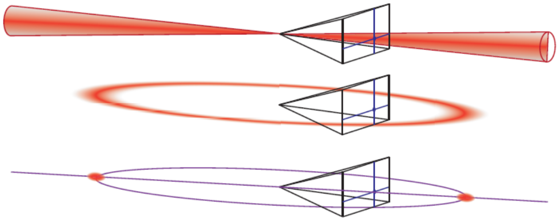
Cost function: The p.d.f. $p_{\lambda\Theta\Psi}$ can be computed by:

$$
\begin{aligned}
p_{\lambda\Theta\Psi}(\lambda,\Theta,\Psi) &= p_\Theta(\Theta)\; p_\Psi(\Psi) \int p_\epsilon(\lambda + F(\mathbf{x},\Theta,\Psi))\; p_\mathbf{x}(\mathbf{x})\; d\mathbf{x} \\
&= p_\Theta(\Theta)\; p_\Psi(\Psi)\; \underbrace{\mathbb{E}\left[p_\epsilon(\lambda + F(\mathbf{x},\Theta,\Psi))\right]}_{p_{\lambda|\Theta\Psi}(\lambda|\Theta,\Psi)}
\end{aligned}
$$

Using observations from the first camera scan, the empirical average can be computed (with $\lambda = 0$):

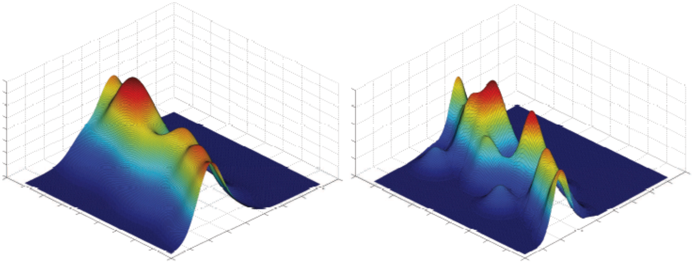
$$
\mathbb{E}\left[p_\epsilon(F(\mathbf{x},\Theta,\Psi_1))\right] \simeq \frac{1}{N_1}\sum_{i=1}^{N_1} p_\epsilon(F(\mathbf{x}_1^{(i)},\Theta,\Psi_1)) = \overline{\mathrm{lik}}(\Theta,\Psi_1)
$$

Our Robust Modelling IV



Our Robust Modelling V

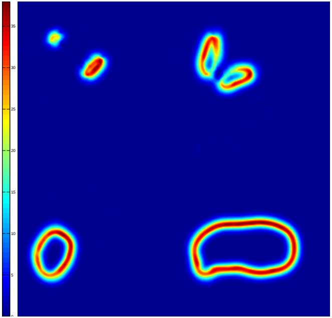
Comparing an isotropic KDE fitted on 3D point clouds $\{\Theta^{(i)}\}$ (right) with our modeling $\overline{\text{lik}}$ (left) in the $\Theta-$space.

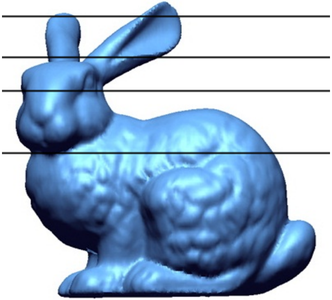


Estimation of the extrinsic camera parameters I

The extrinsic camera parameters can be estimated (noted $\hat{\Psi}_c, \forall c$) via calibration:

$$\overline{\text{lik}}(\Theta) = \sum_{c=1}^{C} \overline{\text{lik}}(\Theta, \hat{\Psi}_c)$$





Estimation of the extrinsic camera parameters II

In practice, the extrinsic camera parameter needs to be refined and re-estimated from the data.

Choosing camera $c = 1$ as a reference camera ($\hat{\Psi}_1$ is available), we formulate the problem as follow:

$$\forall c = 2, \cdots C, \quad \hat{\Psi}_c = \arg\max_{\Psi_c} \int \overline{\text{lik}}(\Theta, \hat{\Psi}_1)\ \overline{\text{lik}}(\Theta, \Psi_c)\ d\Theta$$

Estimation is performed thanks to a gradient ascent algorithm.

Estimation of the extrinsic camera parameters III

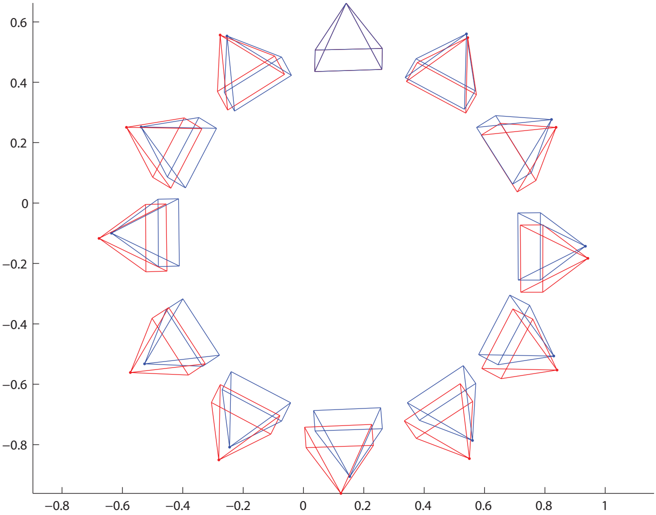
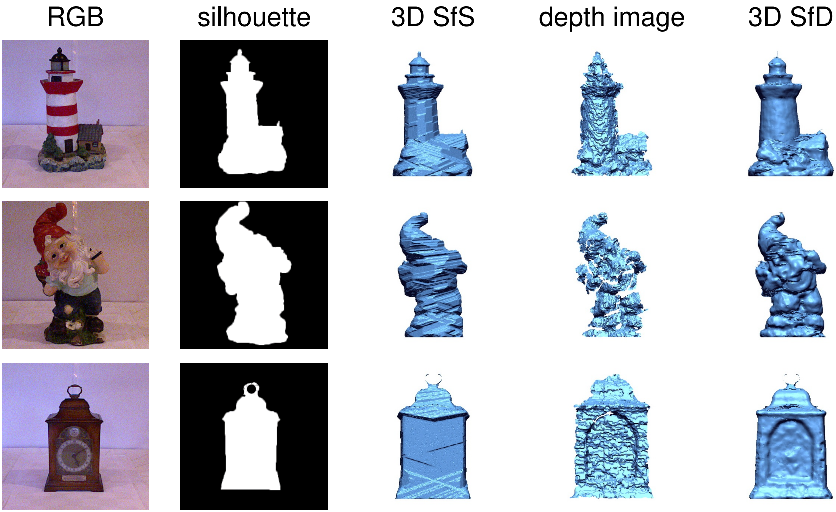
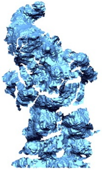


Figure: Camera parameter refinement using a Kinect and a turning table. The red cameras show the original position and orientation of the cameras. The blue cameras show the position and orientation after refinement. Note the reference camera at the top.

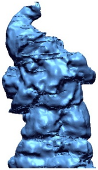
Comparison with Shape from silhouettes (SfS)



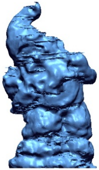
Impact of the number of camera views C I



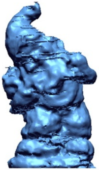
1



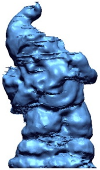
3



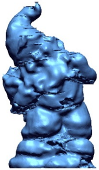
6



9

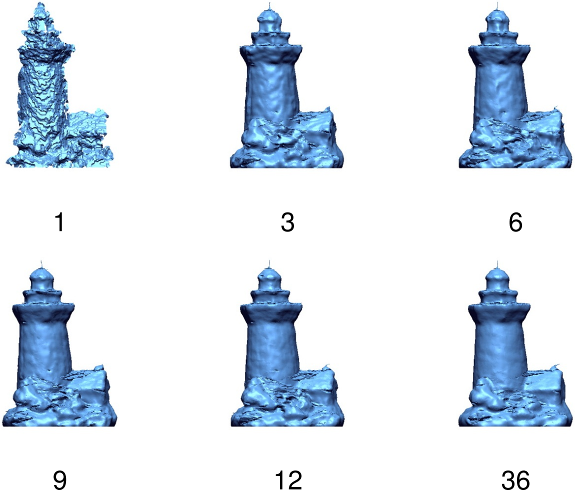


12

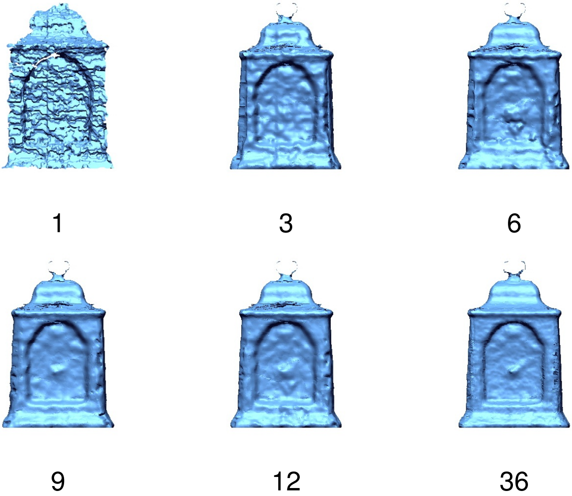


36

Impact of the number of camera views C II

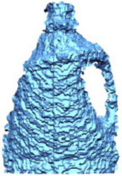


Impact of the number of camera views C III

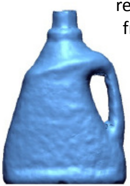


Comparison with Laser scan (Minolta vivid 700)

One Kinect scan

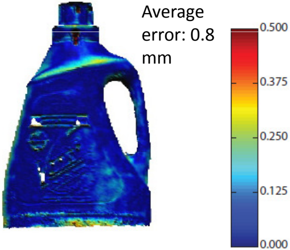


Our reconstruction from multiple kinect scans



One laser scan

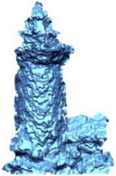
Average error: 0.8 mm



Surface error between laser scan and our reconstruction

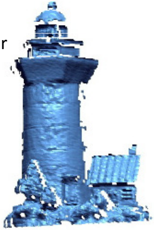
Comparison with Laser scan (Minolta vivid 700)

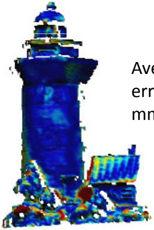
One Kinect scan



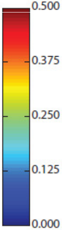
Our reconstruction from multiple kinect scans

One laser scan





Average error: 1.4 mm



Scale in cm

Surface error between laser scan and our reconstruction

Conclusion

We propose

- a smooth kernel objective function for 3D reconstruction from depth scans.
- with a robust estimation of nuisance parameters (i.e. extrinsic camera parameters)

Any question?

Related work:

- *Statistical Framework for Multi Sensor Fusion and 3D reconstruction*, Jonathan Ruttle, PhD thesis Trinity College Dublin Ireland, 2012.
- *Extrinsic camera Parameter estimation for shape-from-depths*, Jonathan Ruttle et al., Eusipco 2012.
- *Smooth Kernel Density Estimate for Multiple View Reconstruction*, Jonathan Ruttle et al., CVMP 2010.