DTAM:Dense Tracking and Mapping in Real-Time

NewCombe, Lovegrove & Davision ICCV11

Outline

- Introduction
- Related Work
- System Overview
- Dense Mapping
- Dense Tracking
- Evaluation and Results
- Conclusions and Future Work

Outline

- **Introduction**
- Related Work
- System Overview
- Dense Mapping
- Dense Tracking
- Evaluation and Results
- Conclusions and Future Work

Introduction

- **Dense** Tracking and Mapping in **Real-Time**

- DTAM is a system for real-time camera tracking and reconstruction which relies not on feature extraction but dense, every pixel methods.

- Simultaneous frame-rate Tracking and Dense Mapping

Outline

- Introduction
- **Related Work**
- System Overview
- Dense Mapping
- Dense Tracking
- Evaluation and Results
- Conclusions and Future Work

Related Work

- Real-time SFM(Structure from Motion)
- PTAM(G. Klein and D. W. Murray. ISMAR 2007)
- Improving the agility of keyframe-based SLAM(G. Klein and D. W. Murray. ECCV 2008)
- Live dense reconstruction with a single moving camera(R. A. Newcombe and A. J. Davison. CVPR 2010)
- Real-time dense geometry from a handheld camera(J. Stuehmer et.al. 2010)

Outline

- Introduction
- Related Work
- **System Overview**
- Dense Mapping
- Dense Tracking
- Evaluation and Results
- Conclusions and Future Work

System Overview

- Input

  -Single hand held RGB Camera

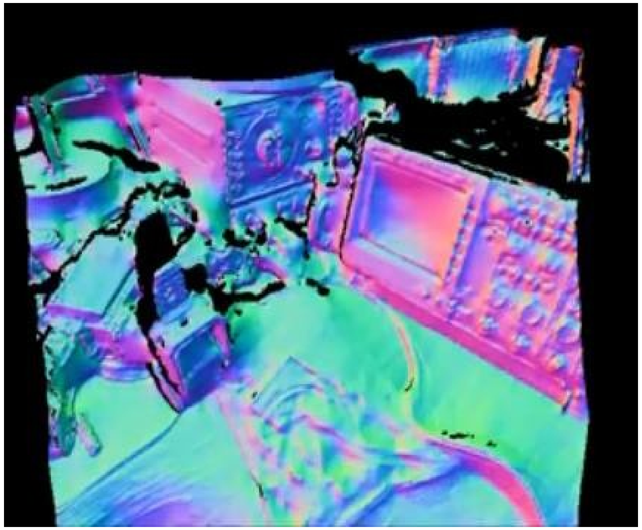
-

- Objective:

  -Dense Mapping

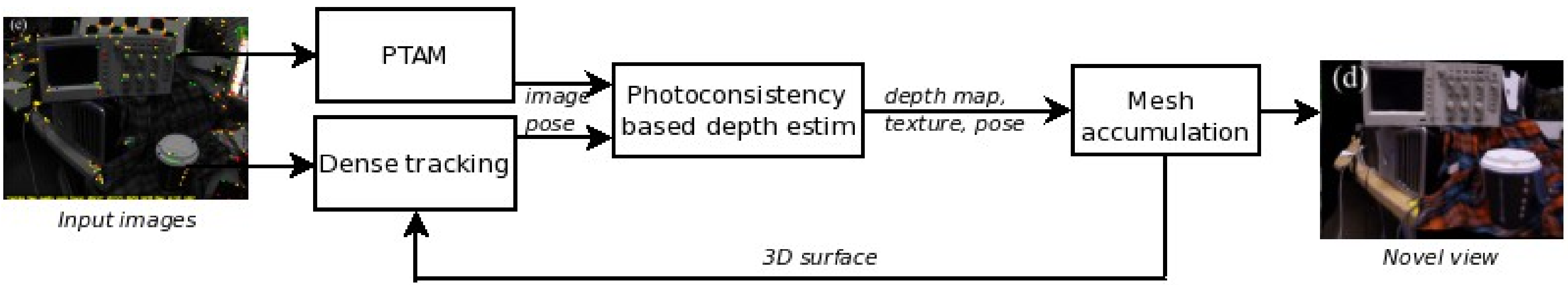
  -Dense Tracking

Input Imgage



3D Dense Map

System Overview

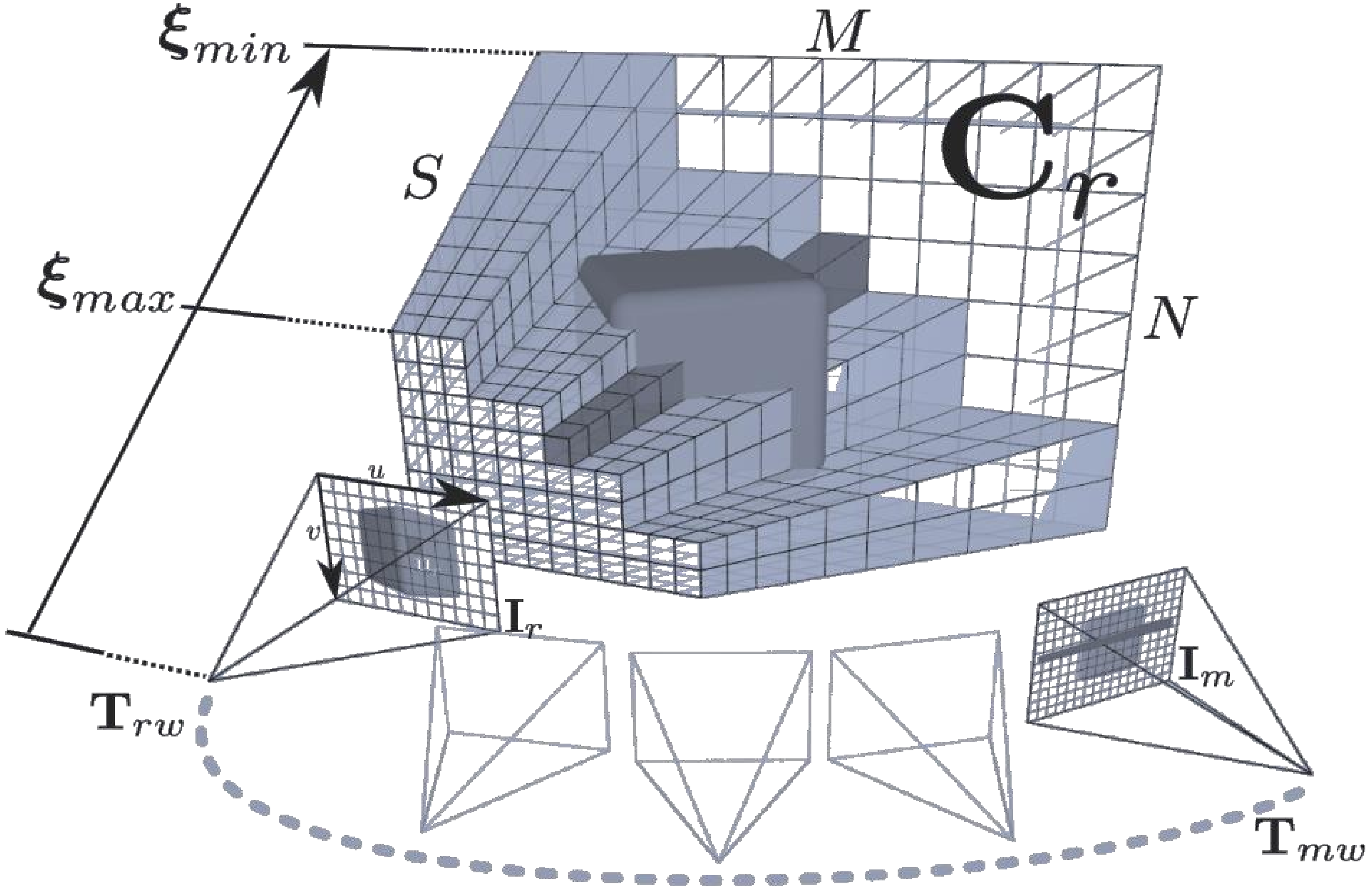


Outline

- Introduction
- Related Work
- System Overview
- **Dense Mapping**
- Dense Tracking
- Evaluation and Results
- Conclusions and Future Work

Dense Mapping

- Estimate inverse depth map from bundles of images



Photometric error

- Total cost

$$\mathbf{C}_r(\mathbf{u}, d) = \frac{1}{|\Gamma(r)|} \sum_{m \in \Gamma(r)} ||\rho_r(\mathbf{I}_m, \mathbf{u}, d)||_1$$

- Photometric error

$$\rho_r(\mathbf{I}_m, \mathbf{u}, d) = \mathbf{I}_r(\mathbf{u}) - \mathbf{I}_m(\pi(KT_{mr}\pi^{-1}(\mathbf{u}, d)))$$

- Where:

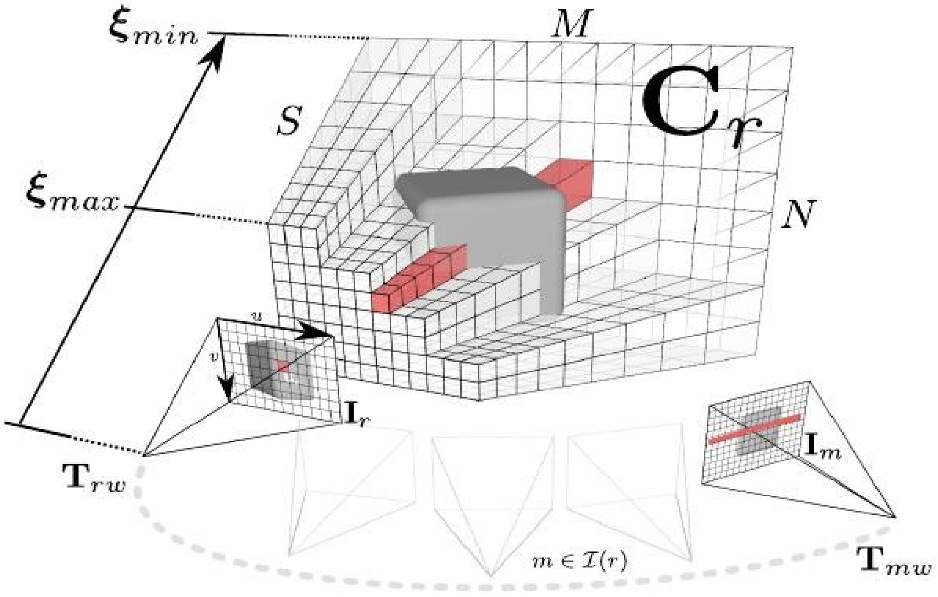
- K ...intrinsic matrix

- Tmr ...transformation from m to r

- $\pi(\mathbf{x}_c) = (x/z, y/z)^T$

- $\pi^{-1}(\mathbf{u}, d) = \frac{1}{d} K^{-1} \dot{\mathbf{u}}$

Depth map estimation



- Principle:
  - **S** depth hypothesis are considered for each pixel of the reference image **Ir**
  - Each corresponding 3D point is projected onto a bundle of images **Im**
  - Keep the depth hypothesis that best respects the color consistency from the reference to the bundle of images

- Formulation: $$\mathbf{C}_r(\mathbf{u}, d) = \frac{1}{|\mathcal{I}(r)|} \sum_{m \in \mathcal{I}(r)} \| \rho_r \left(\mathbf{I}_m, \mathbf{u}, d \right) \|_1$$

  - $\mathbf{u}, d$:pixel position and depth hypothesis
  - $|\mathcal{I}(r)|$:number of valid reprojection of the pixel in the bundle
  - ρ_r :photometric error between reference and current image

$$\rho_r \left(\mathbf{I}_m, \mathbf{u}, d \right) = \mathbf{I}_r \left(\mathbf{u} \right) - \mathbf{I}_m \left(\pi \left(\mathrm{K} \mathbf{T}_{mr} \pi^{-1} \left(\mathbf{u}, d \right) \right) \right)$$

Depth map estimation

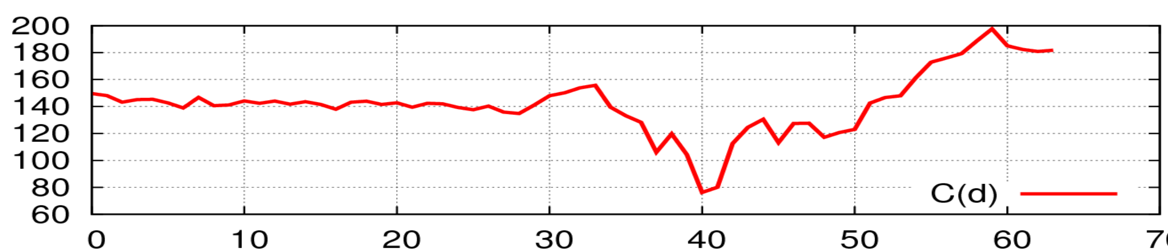
$$\mathbf{C}_r(\mathbf{u}, d) = \frac{1}{|\mathcal{I}(r)|} \sum_{m \in \mathcal{I}(r)} \| \rho_r \left(\mathbf{I}_m, \mathbf{u}, d \right) \|_1$$

Example reference image pixel

Rerojection in Image bundle

Reprojection of depth Hypotheses on one image of bundle

Photomtric error



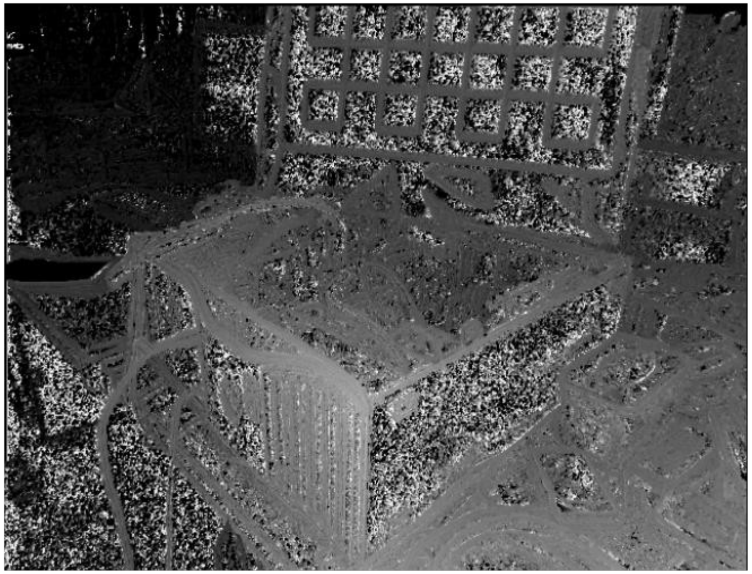
Depth Hypotheses

Inverse Depth Map Computation

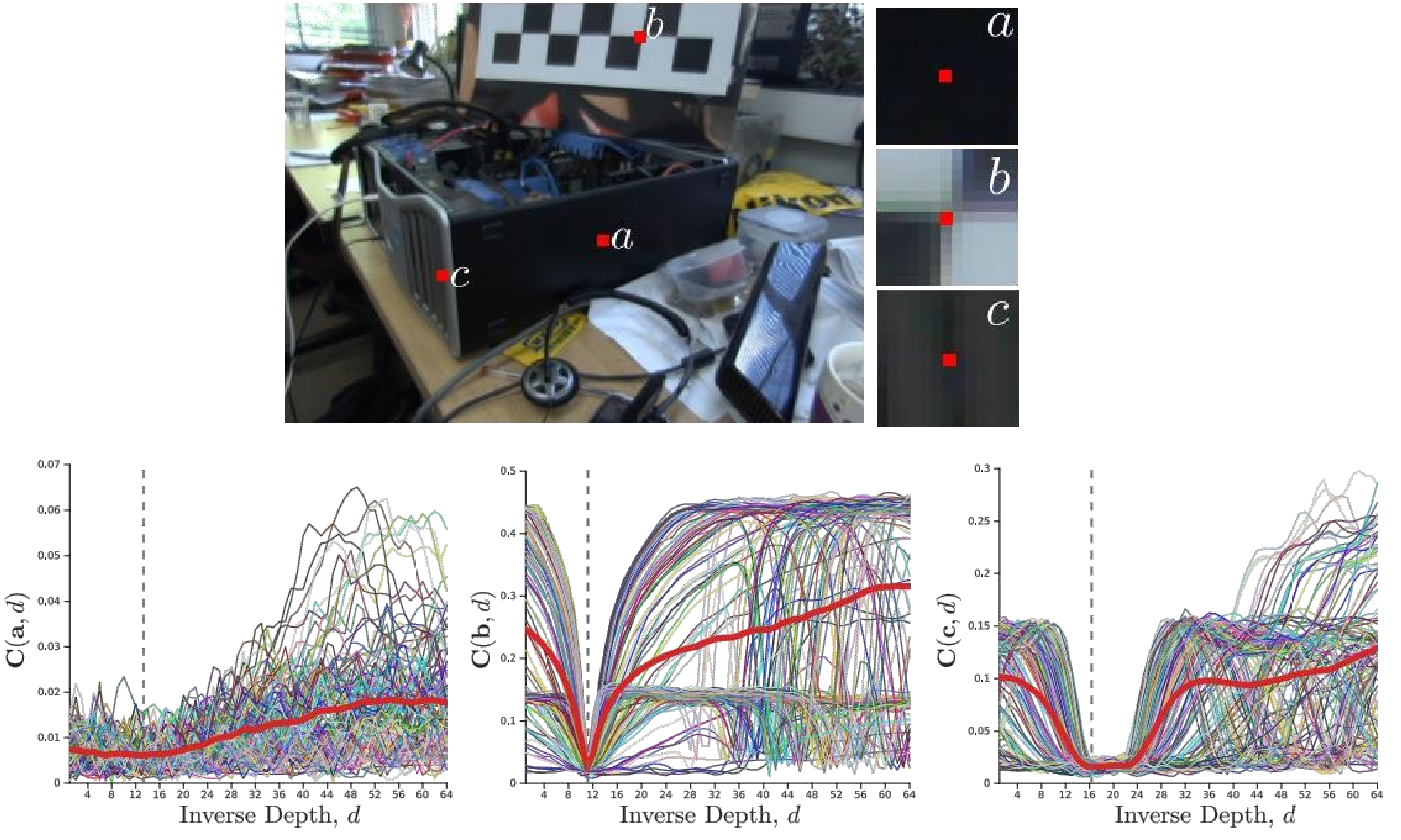
- Inverse depth map can be computed by minimizing the photometric error(exhaustive search ove the volume):

$$min_d \mathbf{C}(\mathbf{u}, d)$$

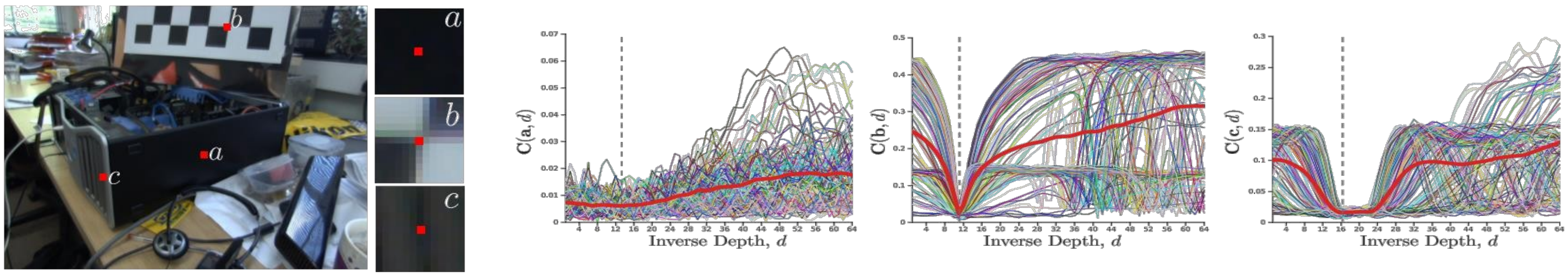
- But featureless regions are prone to false minima



Inverse Depth Map Computation



Depth map filtering approach



- Problem:

  - Uniform regions in reference image do not give discriminative enough photometric error

- Idea:

  - Assume the depth is smooth on uniform regions

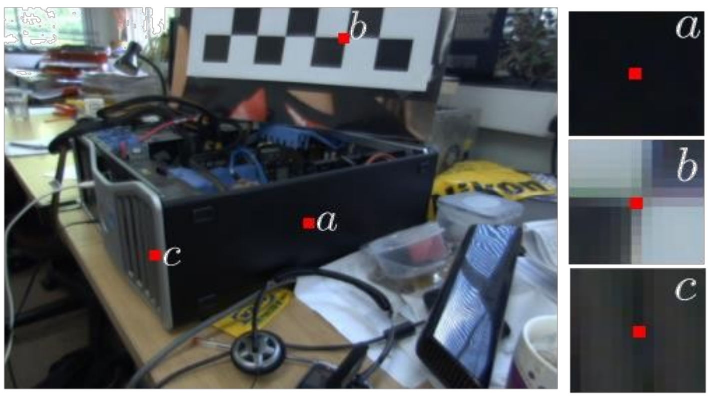
  - Use total variational approach where depth map is the functional to optimize:

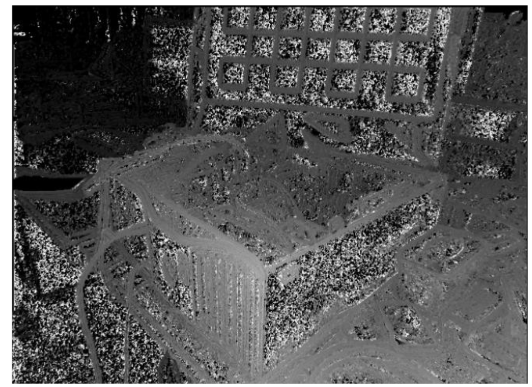
    *photometric error defines the data term

    *the smoothness constraint defines the regularization

Inverse Depth Map Computation

- Featureless regions are prone to false minima





- Solution:Regularization term

  - We want to penalize deviation from spatially smooth solution

  - But preserve edges and discontinuities

Depth map filtering approach

- Formulation of the variational approach

$$E_{\boldsymbol{\xi}} = \int_{\Omega} \left\{ g(\mathbf{u}) \|\nabla \boldsymbol{\xi}(\mathbf{u})\|_{\epsilon} + \lambda \mathbf{C}\left(\mathbf{u}, \boldsymbol{\xi}(\mathbf{u})\right) \right\} \mathrm{d}\mathbf{u}$$

- First term: regularization constraint, g is defined as 0 for image gradients and 1 for uniform regions. So that gradient on depth map is penalized for uniform regions

- Second term: data term defined by the photometric error

- Huber norm: differentiable replacement to L1 norm that better preserve discontinuities compared to L2

Energy Functional

- Regularised cost

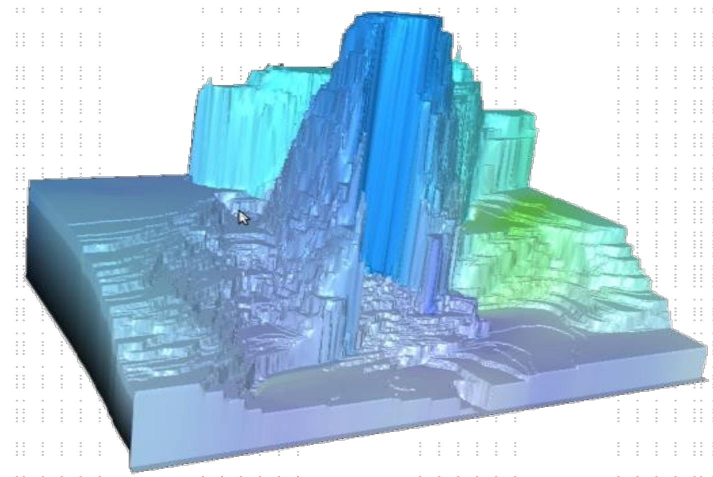
$$E_\xi = \int_\Omega \{ \underbrace{g(\mathbf{u}) \, \|\nabla \xi(\mathbf{u})\|_\epsilon}_{\text{Regularization term}} + \underbrace{\lambda \mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))}_{\text{Photometric cost term}} \} \, d\mathbf{u}$$

$$\underbrace{||x||_\epsilon}_{\text{Huber norm}} = \begin{cases} \frac{||x||_2^2}{2\epsilon} & ||x||_2 \leq \epsilon \\ ||x||_1 - \frac{\epsilon}{2} & else \end{cases}$$

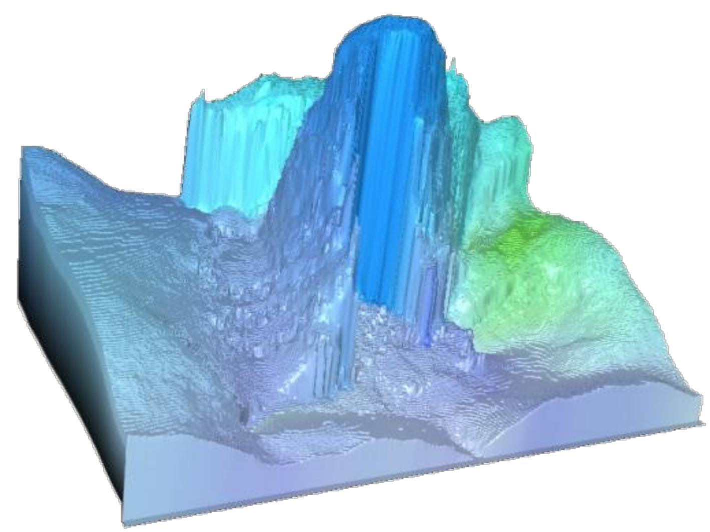
$$\underbrace{g(\mathbf{u})}_{\text{Weight}} = e^{-\alpha ||\nabla I_r(\mathbf{u})||_2^\beta}$$

Total Variation(TV) Regularization

- L1 penalization of gradient magnitudes
  - Favors sparse, piecewise-constant solutions
  - Allows sharp discontinuities in the solution



- Problem
  - Staircasing
  - Can be reduced by using quadratic penalization for small gradient magnitudes

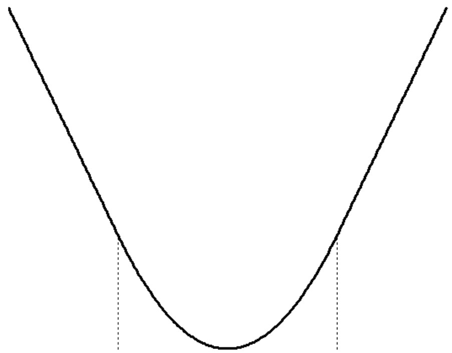


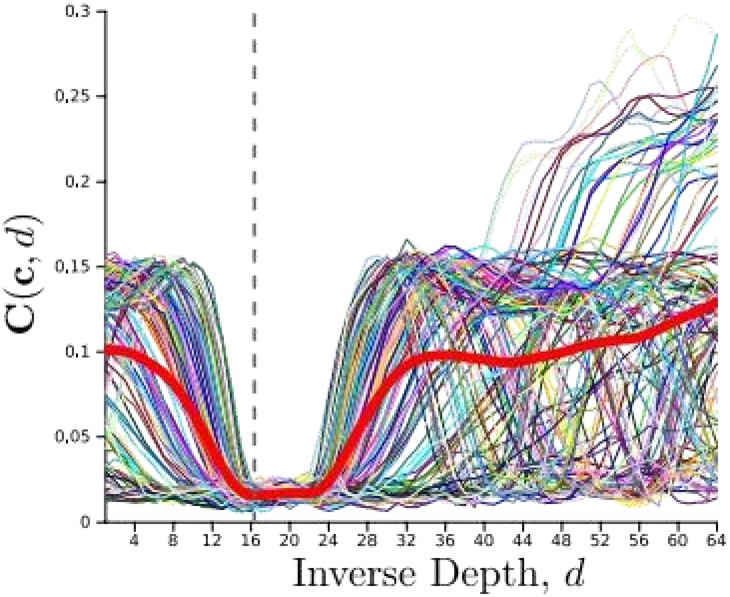
Energy Functional Analysis

$$E_\xi = \int_\Omega \left\{ \boxed{g(\mathbf{u}) \left\| \nabla \xi(\mathbf{u}) \right\|_\epsilon} + \boxed{\lambda \mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))} \right\} d\mathbf{u}$$

Convex

Not convex





Energy Minimization

- Composition of both terms is **non-convex** fuction

- Possible solution

  - Linearize the cost volume to get a convex approximation of the data term

  - Solve approximation iteratively within coarse-to-fine warping scheme

   * Can lead in loss of the reconstruction details

- Better solution?

Key observation

- Data term can be globally optimized by exhaustive search(point-wise optimization)
- Convex regularization term can be solved efficiently by convex optimization algorithms
- And we can approximate the energy functional by decoupling data and regularity term following the approach described in [1][2]

[1]F.Steinbrucker et.al: Large displacement optical flow computation without warping
[2]A.Chambolle et.al: An Algorithm for Total Variation Minimization and Applications

Alternating two Global Optimizations

$$E_\xi = \int_\Omega \{g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_\epsilon + \lambda \mathbf{C}(\mathbf{u}, \xi(\mathbf{u}))\} \, d\mathbf{u}$$

$\alpha = \Omega \to \mathbb{R}$

$$E_{\xi,\alpha} = \int_\Omega \left\{ g(\mathbf{u}) \|\nabla \xi(\mathbf{u})\|_\epsilon + \boxed{\frac{1}{2\theta}(\xi(\mathbf{u}) - \alpha(\mathbf{u}))^2} + \lambda \mathbf{C}(\mathbf{u}, \alpha(\mathbf{u})) \right\} d\mathbf{u}$$

*Drives original and aux. Variables together
*Minimizing functional above equivalent to minimizing original formulation as $\theta \to 0$

*Data and regularity terms are decoupled via aux. Variable α
*Optimization process is split into two sub-prolems

Alternating two Global Optimizations

$$E_{\xi,\alpha} = \int_\Omega \left\{ g(\mathbf{u}) \left\| \nabla \xi(\mathbf{u}) \right\|_\epsilon + \frac{1}{2\theta}(\xi(\mathbf{u}) - \alpha(\mathbf{u}))^2 + \lambda \mathbf{C}(\mathbf{u}, \alpha(\mathbf{u})) \right\} d\mathbf{u}$$

- Energy functional can be globally minimized w.r.t ξ
  * Since it is convex in ξ
  * E.g. gradient descent

- Energy functional can be globally minimized w.r.t α
  * Not convex w.r.t α, but trivially point-wise optimizable
  * Exhaustive search

Algorithm

- Initialization

  - Compute $\alpha_u^0 = \xi_u^0 = min_d C(u,d)$

  - θ = large_value

- Until $\theta^n > \theta_{end}$

  - Compute ξ_u^n

    * Minimize $E_{\xi,\alpha}$ with fixed α

    * Use convex optimization tools, e.g. gradient descent

  - Compute α_u^n

    * Minize $E_{\xi,\alpha}$ with fixed ξ

    * Exhaustive search

  – Decrement θ

Even better

- Problem

  - optimization badly conditioned as $\nabla_u \rightarrow 0$ (uniform regions)

  - expensive when doing exhaustive search

  - accuracy is not good enough

- Solution

  - **Primal-Dual** approach for convex optimization step

  - Acceleration of non-convex search

  - Sub-pixel accuracy

Primal-Dual Approach

- General class of energy minimization problems:

$$\min_{x \in X} \{ F(Kx) + G(x) \}$$

* Usually regularization term
* Often a norm:||Kx||

* Data term

- Can obtain dual form by replacing F(Kx) by its convex conjugate F*(y)
- Use duality principles to arrive at the primal-dual form of $g(u)\|\nabla \xi(u)\|_{\epsilon} + Q(u)$ following [1][2][3]

[1] J.-F. Aujol. Some first-order algorithms for total variation based image restoration
[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging
[3] M.Zhu. Fast numerical algorithms for total variation based image restoration

Primal-Dual Approach

- General problem formulation:

$$\min_{x \in X} \{F(Kx) + G(x)\}$$

$$F^*(y) = \max_{x \in X} \{\langle y, Kx \rangle - F(Kx)\}$$

- By definition(Legendre-Fenchel transform):

$$F(Kx) = \max_{y \in Y} \{\langle Kx, y \rangle - F^*(y)\}$$

- Dual Form(Saddle-point problem):

$$\min_{x \in X} \max_{y \in Y} \{\langle Kx, y \rangle - F^*(y) + G(x)\}$$

Primal-Dual Approach

- Conjugate of Huber norm(obtained via Legendre-Fenchel transform)

$$||x||_\epsilon = \begin{cases} \frac{||x||_2^2}{2\epsilon} & ||x||_2 \leq \epsilon \\ ||x||_1 - \frac{\epsilon}{2} & else \end{cases}$$

$$\delta(p) = f^*(p) = \begin{cases} 0 & ||p|| \leq 1 \\ \infty & else \end{cases}$$

$$\boxed{f^*(p) = \frac{\epsilon}{2}||p||_2^2}$$

$$\Rightarrow \quad \min_{x \in X} \max_{y \in Y} \left\{ \langle Kx, y \rangle - \delta(y) - \boxed{\frac{\epsilon}{2}||y||_2^2} + G(x) \right\}$$

Minimization

- Solving a saddle point problem now!
- Condition of optimality met when $\partial_{x,y}(E(x,y)) = 0$
- Compute partial derivatives
  - $\partial_x E(x,y)$
  - $\partial_y E(x,y)$

- Perform gradient descent
  - Ascent on y(maximization)
  - Descent on x(minimization)

$$\min_{x \in X} \max_{y \in Y} \left\{ \langle Kx, y \rangle - \delta(y) - \frac{\epsilon}{2} \|y\|_2^2 + G(x) \right\}$$

Discretisation

- First some notation:

  - Cost volume is discretized in M X N X S array

    * M X N ... reference image resolution

    * S ... number of points linearly sampling the inverse depth range

  - Use MN X 1 stacked rasterised column vector

    * **d** ... vector version of ξ

    * **a** ... vector version of α

    * **g** ... MN X 1 vector with per-pixel weights

    * G=diag(**g**) ... element-wise weighting matrix

  - A**d** computes 2MN X 1 gradient vector

Implementation

- Replace the weighted Huber regularizer by its conjugate

$$\underbrace{\boxed{\|AG\mathbf{d}\|_{\epsilon}}}_{F(AGd)} = \max_{\mathbf{q},\|\mathbf{q}\|_2 \leq 1} \left\{ \langle AG\mathbf{d}, \mathbf{q} \rangle - \underbrace{\boxed{\delta(\mathbf{q}) - \frac{\epsilon}{2} \|\mathbf{q}\|_2^2}}_{F^*(q)} \right\}$$

- Saddle-point problem
  - Primal variable d and dual variable q
  - Coupled with data term
    * Sum of convex and non-convex functions

$$\max_{\mathbf{q},\|\mathbf{q} \leq 1\|} \; \min_{\mathbf{d},\mathbf{a}} \; E(\mathbf{d}, \mathbf{a}, \mathbf{q})$$

$$E(\mathbf{d}, \mathbf{a}, \mathbf{q}) = \left\{ \langle AG\mathbf{d}, \mathbf{q} \rangle + \frac{1}{2\theta} \|\mathbf{d} - \mathbf{a}\|_2^2 + \lambda \mathbf{C}(\mathbf{a}) - \underbrace{\boxed{\delta(\mathbf{q}) - \frac{\epsilon}{2} \|\mathbf{q}\|_2^2}}_{F^*(q)} \right\}$$

Algorithm

- Compute partial derivatives
  - $\partial_{\mathbf{q}} E(\mathbf{d}, \mathbf{a}, \mathbf{q})$
  - $\partial_{\mathbf{d}} E(\mathbf{d}, \mathbf{a}, \mathbf{q})$

- For fixed **a**, gradient ascent w.r.t **q** and gradient descent w.r.t **d** is performed
- For fixed **d**, exhaustive search w.r.t **a** is performed
- θ is decremented
- Until $\theta^n > \theta_{end}$

[1] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging

Outline

- Introduction
- Related Work
- System Overview
- Dense Mapping
- **Dense Tracking**
- Evaluation and Results
- Conclusions and Future Work

Dense Tracking

- Inputs:

  - 3D texture model of the scene

  - Pose at previous frame

- Tracking as a registration problem

  - First inter-frame rotation estimation: the previous image is aligned on the current image to estimate a coarse inter-frame rotation

  - Estimated pose is used to project the 3D model into 2.5D image

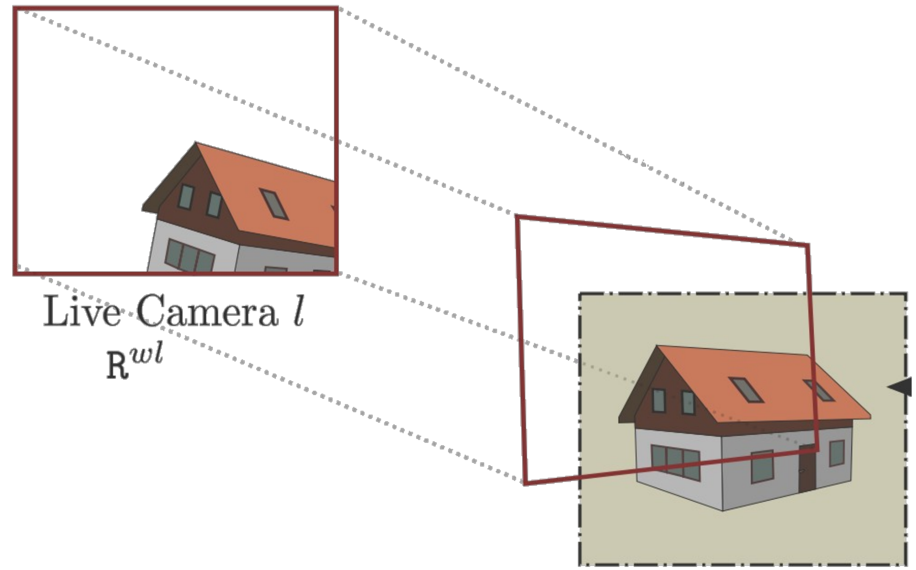
  - The 2.5D image is registered with the current frame to find the current pose

↓

Template matching problem

Tracking Strategy and Algorithm

- Based on image alignment against dense model
- Coarse-to-fine strategy
  - Pyramid hierarchy of images
- Lucas-Kanade algorithm
  - Estimate “warp” between images
  - Iterative minimization of a cost function
  - Parameters of warp correspond to dimensionality of search space



Tracking in Two Stages

- Two stages

  - Constrained rotation estimation

    * Use coarser scales $\hat{T}_{wl}$

    * Rough estimate of pose

  - Accurate 6-DOF pose refinement

    * Set virtual camera ν at location $T_{w\nu}=\hat{T}_{wl}$

      Project dense model to the virtual camera

      Image I_ν ,inverse depth image ξ_ν

    * Align live image I_l and I_ν to estimate $T_{l\nu}$

    * Final pose estimation $T_{wl}=T_{w\nu}T_{l\nu}$

SSD optimization

- Problem:

  - Align template image T(x) with input image I(x)

- Formulation:

  - Find the transform $W(x;p)$ that best maps the pixels of the templates into the ones of the current image minimizing:

$$\sum_x \left[I\left(W\left(x;p \right) \right) - T\left(x \right) \right]^2$$

  - $p = \left(p_1, \ldots, p_n \right)^T$ are the displacement parameters to be optimized

- Hypothesis:

  - Known a coarse approximation of the template position $\left(\boldsymbol{p_0} \right)$

SSD optimization

- Problem:
  - minimize $\sum_x \left[I(W(x;p)) - T(x) \right]^2$
  - The current estimation of **p** is iteratively updated to reach the minimum of the function.

- Formulations:
  - Direct additional

$$\sum_x \left[I(W(x;p+\Delta p)) - T(x) \right]^2$$

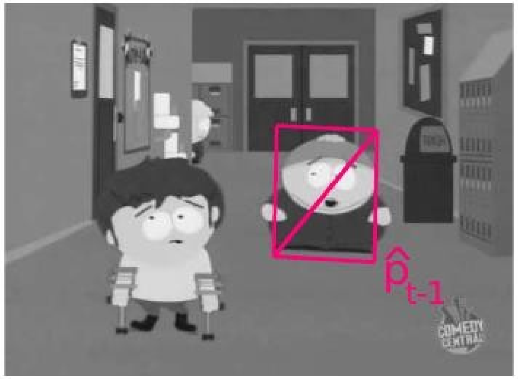
  - Direct compositional

$$\sum_x \left[I(W(W(x;\Delta p);p)) - T(x) \right]^2$$

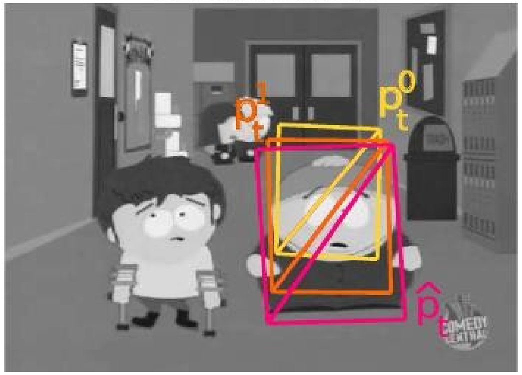
  -Inverse

$$\sum_x \left[I(W(x;\Delta p)) - I(W(x;p)) \right]^2$$

Previous image I_{t-1}



Current image I_t



SSD optimization

- Example: Direct additive method

  - Minimize:

$$\sum_x \left[I\left(W\left(x;p+\Delta p\right)\right) - T\left(x\right)\right]^2$$

  - First order Taylor expansion:

$$\sum_x \left[I\left(W\left(x;p\right)\right) + \nabla I \frac{\partial W}{\partial p} \Delta p - T\left(x\right)\right]^2$$

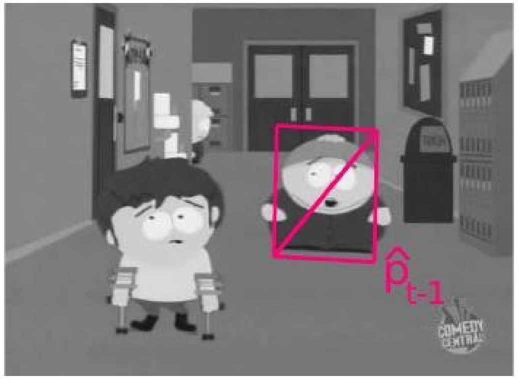
  - Solution:

$$\Delta p = \sum_x H^{-1} \left[\nabla I \frac{\partial W}{\partial p}\right]^T \left[T\left(x\right) - I\left(W\left(x;p\right)\right)\right]$$

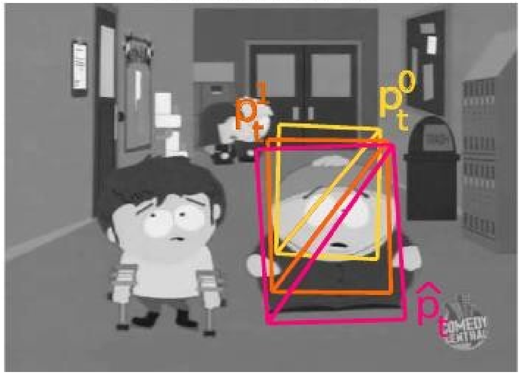
  - with:

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p}\right]^T \left[\nabla I \frac{\partial W}{\partial p}\right]$$

Previous image I_{t-1}



Current image I_t



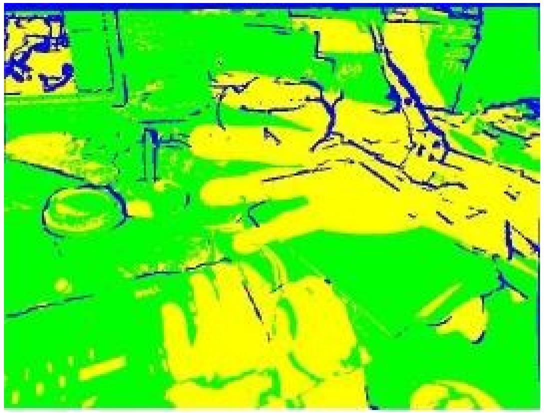
SSD robustified

- Formulation:

$$\Delta p = \sum_x H^{-1} \left[\nabla I \frac{\partial W}{\partial p} \right]^T [T(x) - I(W(x;p))]$$

- Problem: In case of occlusion, the occluded pixels cause the optimum of the function to be changed. The occluded pixels have to be ignored from the optimization

- Method
  - Only the pixels with a difference $[T(x) - I(W(x;p))]$ lower than a threshold are selected
  - Threshold is iteratively updated to get more selective as the optimization reaches the optimum



Template matching in DTAM

- Inter-frame rotation estimation

  - the template is the previous image that is matched with current image. Warp is defined on $SO(3)$. The initial estimate of **p** is identity.

- Full pose estimation

  - template is 2.5D, warp is defined by full 3D motion estimation, which is on $SE(3)$

  - The initial pose is given by the pose estimated at the previous frame and the inter-frame rotation estimation

6 DOF Image Alignment

- Gauss-Newton gradient descent non-linear optimization

$$F(\psi) = \frac{1}{2} \sum_{u \in \Omega} \left(f_u(\psi) \right)^2$$

$$f_u(\psi) = I_l\left(\pi\left(KT_{lv}(\psi)\pi^{-1}\left(u, \xi_v(u)\right)\right)\right) - I_v(u)$$

$$T_{lv}(\psi) = \exp\left(\sum_{i=1}^{6} \psi_i \underset{SE(3)}{gen_i} \right)$$

$$\psi \in R^6$$

Belongs to Lie Algebra $\varsigma\varrho 3$

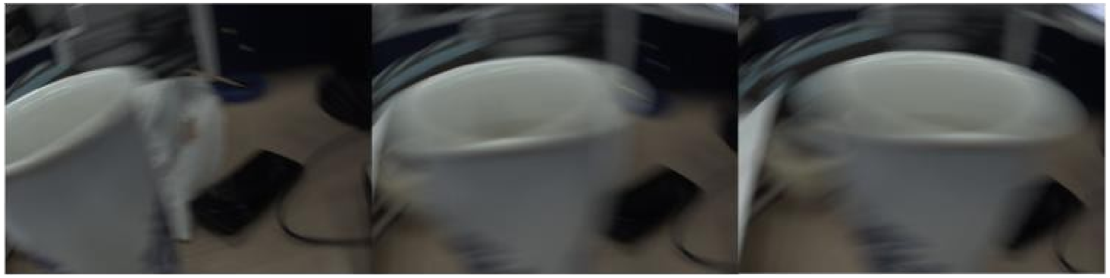
- Non-linear expression linearized by first-order Taylor expansion

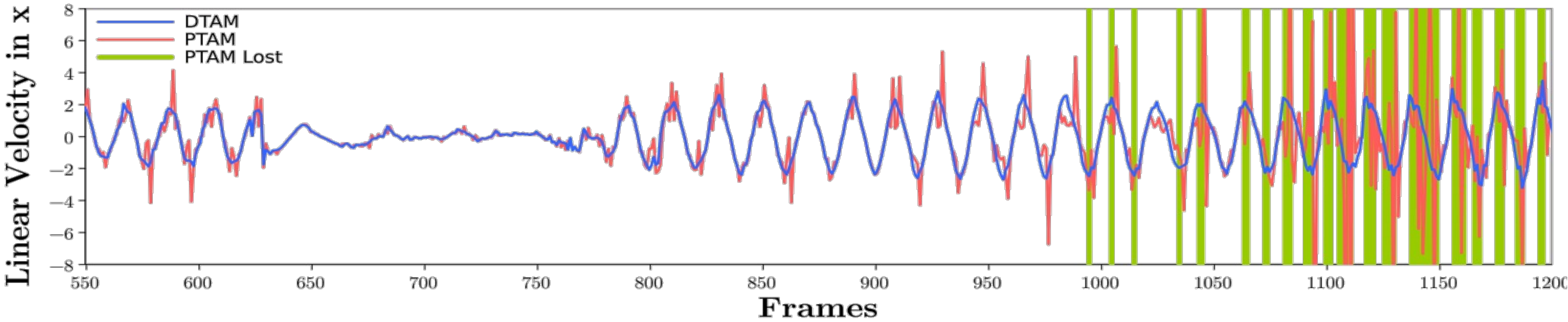
Outline

- Introduction
- Related Work
- System Overview
- Dense Mapping
- Dense Tracking
- **Evaluation and Results**
- Conclusions and Future Work

Evaluation and Results

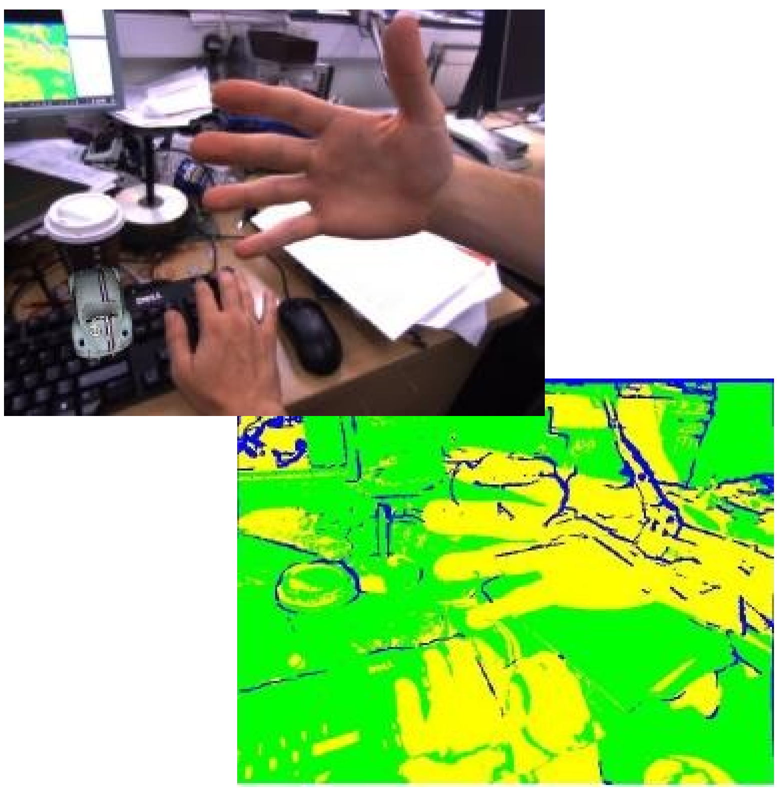
- Runs in real-time
  - NVIDIA GTX 480 GPU
  - i7 quad-core GPU
  - Grey Flea2 camera
    * Resolution 640*480
    * 30Hz
- Comparison with PTAM
  - a challenging high acceleration back-and-forth trajectory close to a cup
  - with DTAM's relocaliser disabled

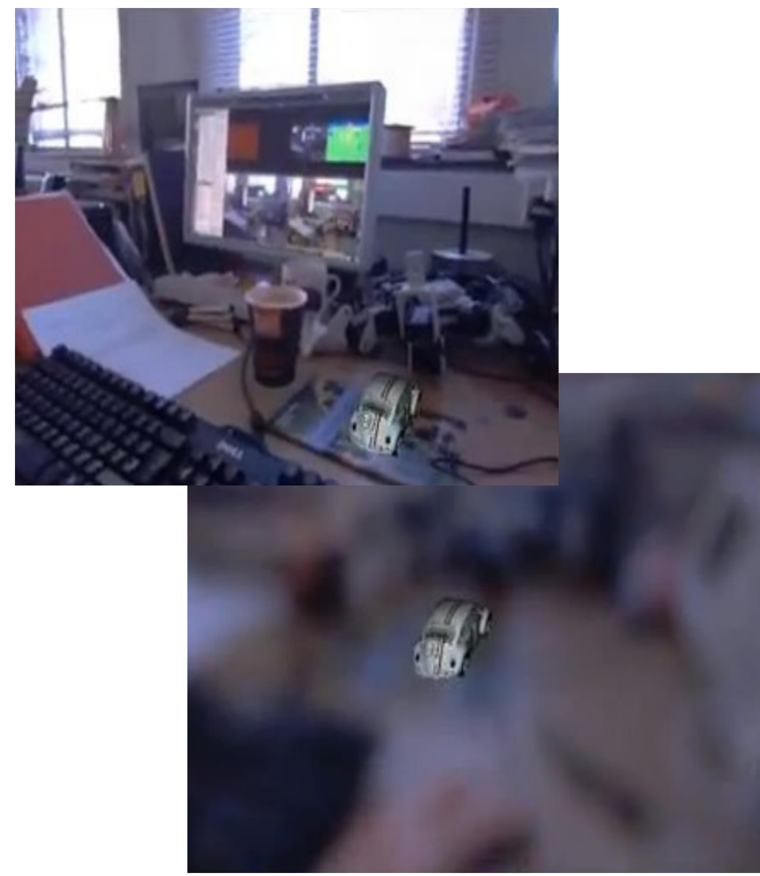




Evaluation and Results

- Unmodelled objects
- Camera defocus





Outline

- Introduction
- Related Work
- System Overview
- Dense Mapping
- Dense Tracking
- Evaluation and Results
- **Conclusions and Future Work**

Conclusions

- First live full dense reconstruction system
- Significant advance in real-time geometrical vision
- Robust

  - rapid motion

  - cemera defocus
- Dense modelling and dense tracking make the system beat any point-based method with modelling and tracking performance

Future Work

- Short comings

  - Brightness constancy assumption

    * often violated in real-world

    * not robust to global illumination changes

  - Smoothness assumption on depth

- Possible solutions

  - integrate a normalized cross correlation measure into the objective function for more robustness to local and global lighting changes

  - joint modelling of the dense lighting and reflectance properties of the scene to enable moe accurate photometric cost functions(the authors are more interested in this approach)

Thank You!