Learning 3D models of non-rigid scenes from video

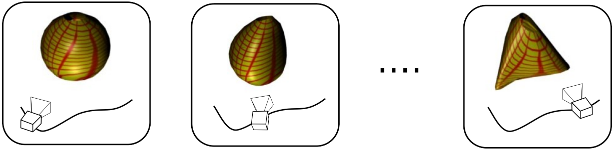
Lourdes Agapito
University College London

Acknowledgements: Anastasios Roussos, Chris Russell, Sara Vicente, Ravi Garg, Nikolaos Pitelis, Rui Yu, Joao Fayad, Marco Paladini

October 1, 2013

Non-Rigid Structure from Motion

Problem statement: Point a camera at a deforming object and generate a detailed 3D model by exploiting 2D motion.

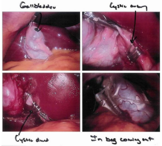


- No given models, no training data, no extra sensors.

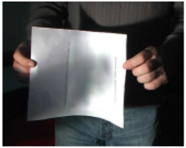
Only input: 2D video data

Why is it important?



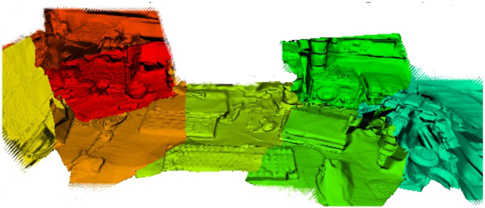
Motion capture/Animation Gesture analysis Laparoscopy

Augmented reality (Pilet et al. 2008) Cloth animation

Rigid Structure from Motion

- The **rigidity prior** is enough to make the problem well posed.
- Commercial applications exist.
- Mature problem / Solved?
- Not quite: Real time, large scale, dense – open problems.



(Newcombe et al. ICCV'11)

(Agarwal et al. ICCV'09)

But the Real World is Non-Rigid

Non-Rigid

Dynamic

Articulated

Piecewise-Rigid

NRSfM: An ill-posed problem $\rightarrow$ Priors

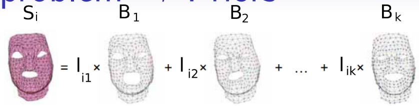
NRSfM: An ill-posed problem → Priors

Statistical priors

Low rank shape basis.

Gaussian priors.

Dynamics.



Bregler *et al.* CVPR 2000.

NRSfM: An ill-posed problem → Priors

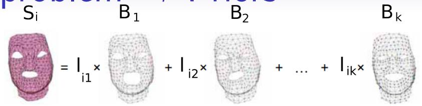
Statistical priors

Low rank shape basis.

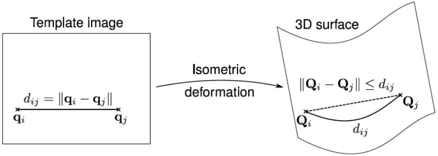
Gaussian priors.

Dynamics.

Known 3D template



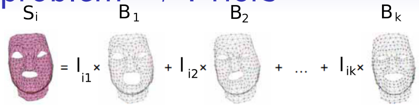
Bregler *et al.* CVPR 2000.



NRSfM: An ill-posed problem → Priors

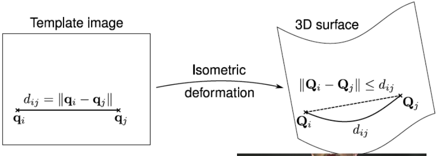
Statistical priors

Low rank shape basis.
Gaussian priors.
Dynamics.

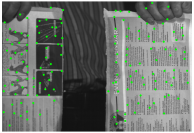


Bregler *et al.* CVPR 2000.

Known 3D template



Physical priors

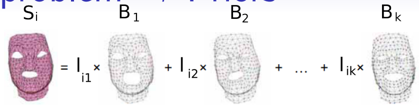


Single object (C_0 continuity).
Smooth surface (C_1 continuity).
Inextensibility, isometry.
Temporal and spatial smoothness.
Piecewise planar/quadratic/rigid.

NRSfM: An ill-posed problem → Priors

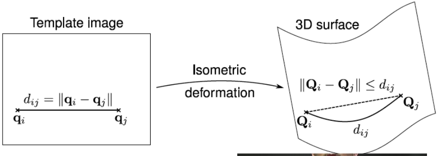
Statistical priors

Low rank shape basis.
Gaussian priors.
Dynamics.

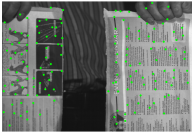


Bregler *et al.* CVPR 2000.

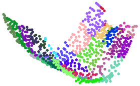
Known 3D template



Physical priors

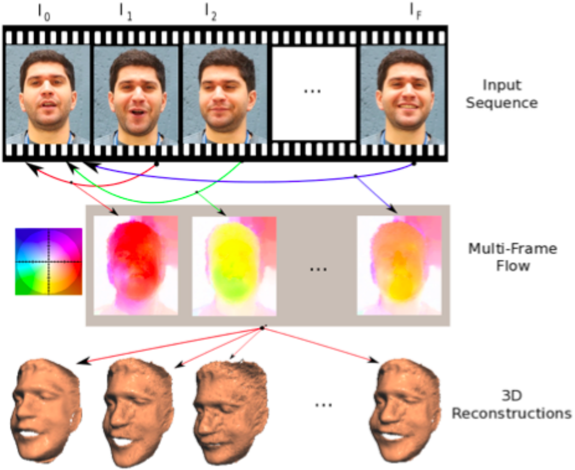


Single object (C_0 continuity).
Smooth surface (C_1 continuity).
Inextensibility, isometry.
Temporal and spatial smoothness.
Piecewise planar/quadratic/rigid.



Current Research Threads I

1. Non-Rigid 3D Reconstruction with a Global Model



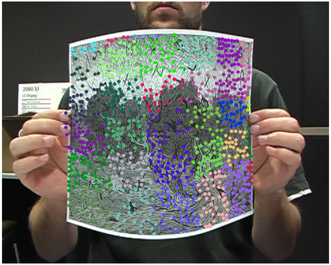
Del Bue-Agapito IJCV'04, Del Bue et al. CVPR'06, Paladini et al. CVPR'09, Paladini et al. ECCV'10, Fayad et al. ECCV'10

Russell et al. CVPR'11, Paladini et al. IJCV'12, Del Bue et al. PAMI'12, **Garg et al. IJCV'13, Garg et al. CVPR'13**

Current Research Threads II

2. Reconstructing Extreme Deformations with Connected Local Models

Articulated



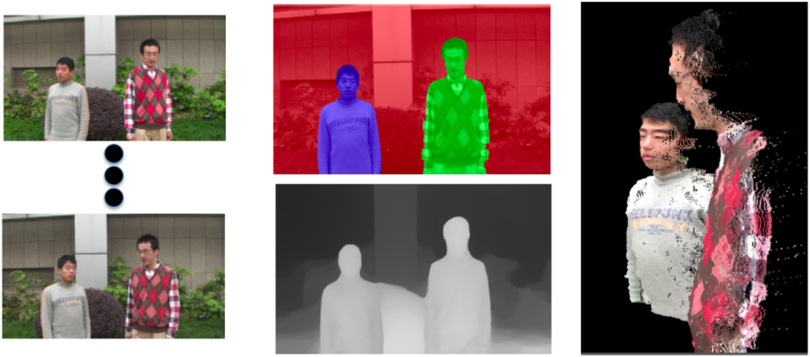
Non-Rigid

Russell et al. CVPR'11, Fayad et al. ICCV'11, Pitelis et al. CVPR'13

Current Research Threads III

3. Towards Direct Non-Rigid Modelling from Pixel Intensities

- Jointly solving for dense matching, segmentation and 3D reconstruction directly from pixel intensities for multi-rigid scenes.



Input: Video sequence   Output: Segmentation + dense 3D models

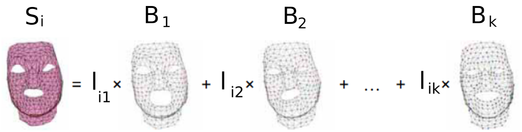
Roussos-Russell-Garg-Agapito ISMAR'12

Roadmap

- **Non-Rigid Reconstruction with a Global Low Rank Model**
- Reconstructing Extreme Deformations with Connected Local Models
- Direct Dense Multibody SfM from Pixel Intensities.

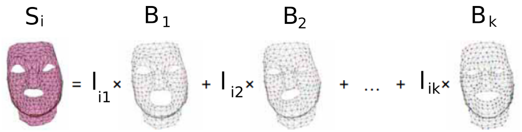
Non-Rigid SFM: Low Rank Shape Model

The 3D shape is represented as a linear combination of basis shapes.
(Bregler et al. CVPR2000)



Non-Rigid SFM: Low Rank Shape Model

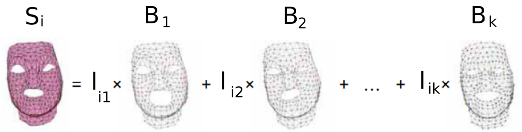
The 3D shape is represented as a linear combination of basis shapes.
(Bregler et al. CVPR2000)



$$B_k = \begin{bmatrix} \begin{matrix} X_{K1} \\ Y_{K1} \\ Z_{K1} \end{matrix} & \begin{matrix} X_{K2} \\ Y_{K2} \\ Z_{K2} \end{matrix} & \dots & \begin{matrix} X_{KP} \\ Y_{KP} \\ Z_{KP} \end{matrix} \end{bmatrix}$$

Non-Rigid SFM: Low Rank Shape Model

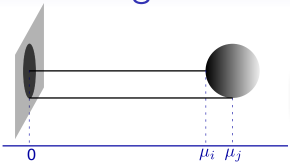
The 3D shape is represented as a linear combination of basis shapes.
(Bregler et al. CVPR2000)



$$B_k = \begin{bmatrix} X_{K1} & X_{K2} & \dots & X_{KP} \\ Y_{K1} & Y_{K2} & & Y_{KP} \\ Z_{K1} & Z_{K2} & & Z_{KP} \end{bmatrix}$$

$$\mathrm{S}_i = \sum_{d=1}^{k} l_{id}\mathrm{B}_d$$

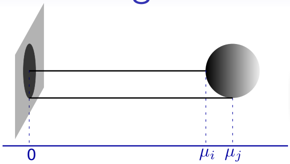
Non-Rigid Structure from Motion



$$\begin{bmatrix} u_{11} \\ v_{11} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}$$

3D points are projected onto the image via an orthographic camera.

Non-Rigid Structure from Motion



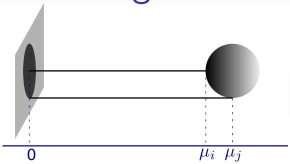
$$\begin{bmatrix} u_{11} \\ v_{11} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}$$

3D points are projected onto the image via an orthographic camera.

$$\mathtt{W}_i = \mathtt{R}_i \; \mathtt{S}_i$$

$$\begin{bmatrix} u_{11} & \cdots & u_{1P} \\ v_{11} & \cdots & v_{1P} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix}$$

Non-Rigid Structure from Motion



$$\begin{bmatrix} u_{11} \\ v_{11} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}$$

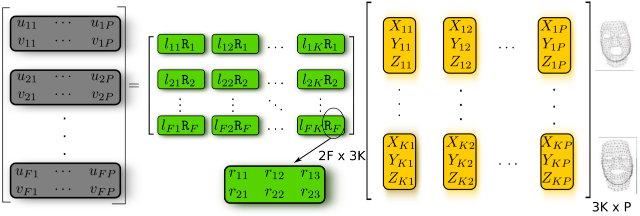
3D points are projected onto the image via an orthographic camera.

$$\mathtt{W}_i \quad = \quad \mathrm{R}_i \quad \mathtt{S}_i$$

$$\begin{bmatrix} u_{11} & \cdots & u_{1P} \\ v_{11} & \cdots & v_{1P} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix}$$

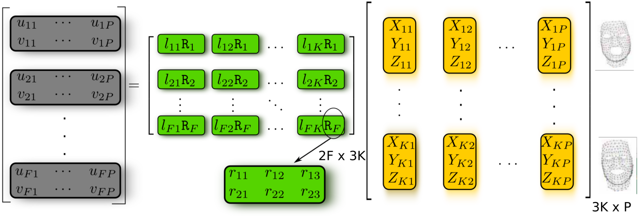
$$\mathtt{W}_i = \mathrm{R}_i\left(\sum_{d=1}^{k} l_{id}\mathrm{B}_d\right)$$

Non-Rigid Structure from Motion



Measurement matrix is **rank 3K** → Bregler et al. CVPR 2000

Non-Rigid Structure from Motion

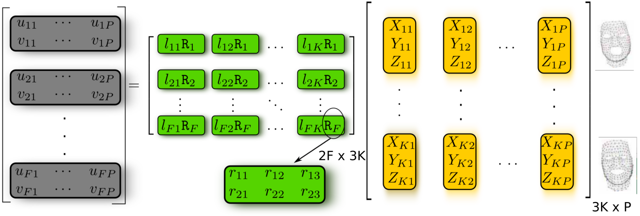


Measurement matrix is **rank 3K** → Bregler et al. CVPR 2000

- Closed form Solutions
  1. Factorization $\text{SVD}(\mathtt{W}) = \mathtt{MS} = \hat{\mathtt{M}}\mathtt{QQ}^{-1}\hat{\mathtt{S}}$.
  2. Metric upgrade: Estimation of mixing matrix $\mathtt{Q}_{3K\times 3K}$ is non-linear.

Bregler et al. CVPR'00, Brand CVPR'05, Xiao et al. IJCV'06, Akhter et al. CVPR'09, Dai et al. CVPR'12

Non-Rigid Structure from Motion



Measurement matrix is **rank 3K** → Bregler et al. CVPR 2000

- Closed form Solutions
  1. Factorization SVD($\mathtt{W}$) $= \mathtt{MS} = \hat{\mathtt{M}}\mathtt{Q}\mathtt{Q}^{-1}\hat{\mathtt{S}}$.
  2. Metric upgrade: Estimation of mixing matrix $\mathtt{Q}_{3K\times 3K}$ is non-linear.

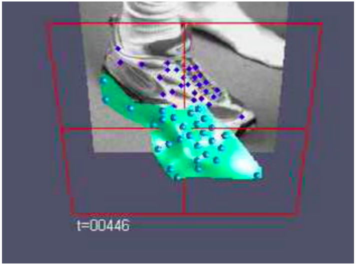
Bregler et al. CVPR'00, Brand CVPR'05, Xiao et al. IJCV'06, Akhter et al. CVPR'09, Dai et al. CVPR'12

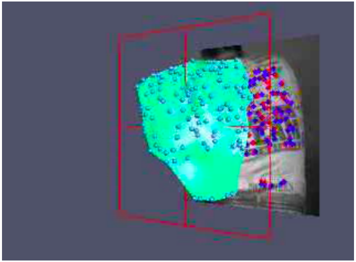
- Energy Optimization: $\sum_i^F \left\| \mathtt{W}_i - \mathrm{R_i}\left(\sum_{d=1}^{k} l_{id}\mathrm{B_d}\right)\right\|^2$.

Torresani et al. PAMI'08, **Paladini et al. IJCV'12, Del Bue et al. PAMI'12**

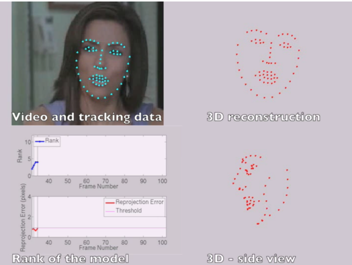
Low Rank Shape Model: Results

Torresani-etal PAMI'08



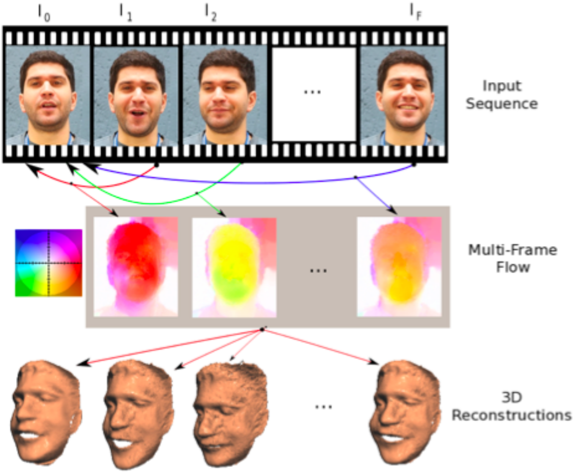


Paladini-Bartoli-Agapito ECCV'10



Dense Low-Rank Non-Rigid 3D Reconstruction

Garg-Roussos-Agapito CVPR'13



Orthographic Projection Model

Our geometric model is:

$$W = RS$$

- W is a known matrix of 2D optical flow tracking data
- R is a block diagonal stacking of projection matrices R_f
- S stacks the deformable 3D shapes for all F frames

$$\underbrace{W}_{2F\times N} = \left[\begin{array}{ccc} u_{11} & \cdots & u_{1N} \\ v_{11} & \cdots & v_{1N} \\ \hline \vdots & \ddots & \vdots \\ \hline u_{F1} & \cdots & u_{FN} \\ v_{F1} & \cdots & v_{FN} \end{array}\right], \underbrace{R}_{2F\times 3F} = \begin{bmatrix} R_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & R_F \end{bmatrix}, \underbrace{S}_{3F\times N} = \left[\begin{array}{ccc} X_{11} & \cdots & X_{1N} \\ Y_{11} & \cdots & Y_{1N} \\ Z_{11} & \cdots & Z_{1N} \\ \hline \vdots & \ddots & \vdots \\ \hline X_{F1} & \cdots & X_{FN} \\ Y_{F1} & \cdots & Y_{FN} \\ Z_{F1} & \cdots & Z_{FN} \end{array}\right]$$

Priors

We propose to use two priors:

- The 3D trajectories lie on a low dimensional linear subspace, i.e. the **shape matrix S is inherently low rank**.
- The 3D trajectories of the points are **spatially smooth**.

We cast the dense NRSfM problem as a variational energy minimisation problem:

$$E(R, S) = \lambda E_{data} + \tau E_{low\_rank} + E_{smooth}$$

Energy to minimise

We cast the dense NRSfM problem as a variational energy minimisation problem:

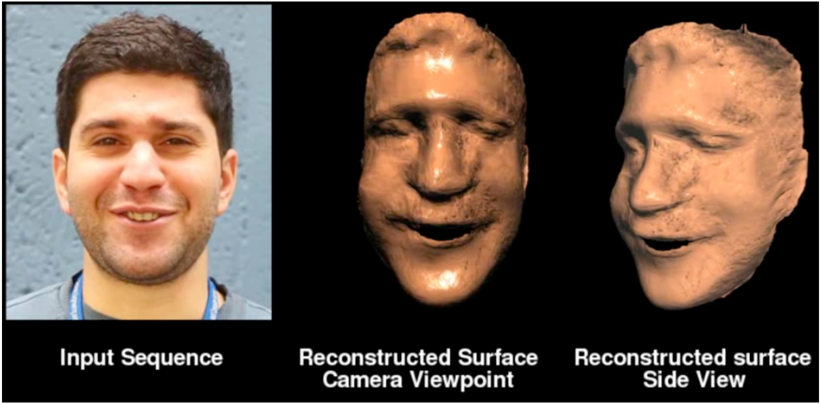
$$E(R,S) = \lambda E_{data} + \tau E_{trace} + E_{reg}$$

where

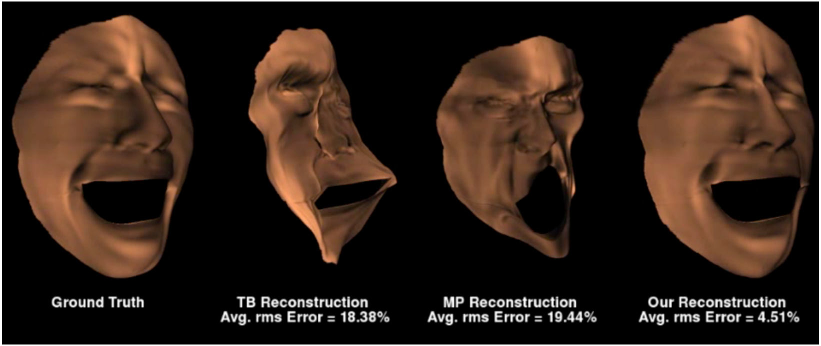
- $$E_{data} = \frac{1}{2}\|W - RS\|^2_{\mathcal{F}}$$
- $$E_{trace} = \|S\|_* = \sum_{j=1}^{min(3F,N)} \Lambda_j$$
- $$E_{reg} = \sum_{i=1}^{3F}\sum_{p=1}^{N} \|\nabla S_i(p)\|$$

OPTIMIZATION: Alternate between estimating R and S. R $\rightarrow$ bundle adjustment, S $\rightarrow$ primal-dual regularization + soft impute.
Efficient Primal-Dual Optimization. Highly parallel $\rightarrow$ GPU.

Experimental Results: Real Sequences



Experimental Results: Synthetic Sequences



How do we obtain dense correspondences?

Non-Rigid Video Registration with Subspace Constraints

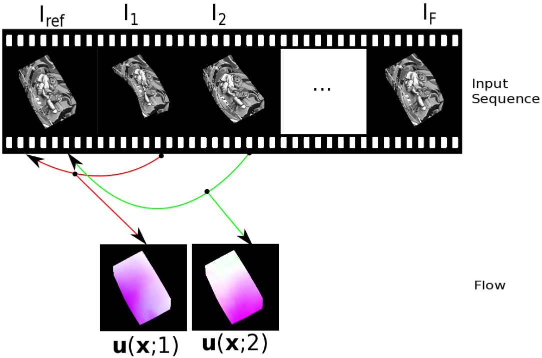
Garg-Roussos-Agapito (IJCV'13)

How do we obtain dense correspondences?

Non-Rigid Video Registration with Subspace Constraints

Garg-Roussos-Agapito (IJCV'13)

Register video $(I_1, I_2, \ldots I_F)$ to a single reference template I_{ref}.

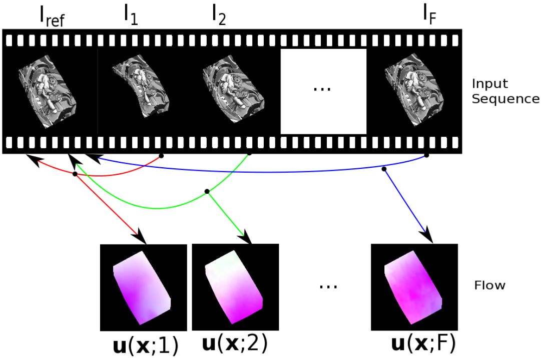


How do we obtain dense correspondences?

Non-Rigid Video Registration with Subspace Constraints

Garg-Roussos-Agapito (IJCV'13)

Register video $(I_1, I_2, \ldots I_F)$ to a single reference template I_{ref}.

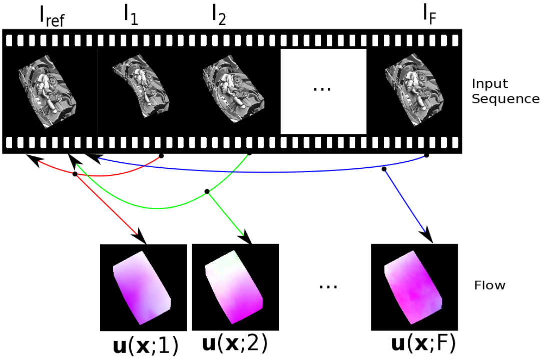


How do we obtain dense correspondences?

Non-Rigid Video Registration with Subspace Constraints

Garg-Roussos-Agapito (IJCV'13)

Register video $(I_1, I_2, \ldots I_F)$ to a single reference template I_{ref}.



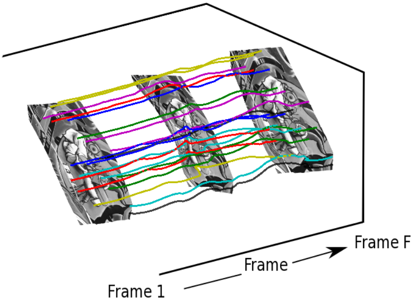
such that $I_n(\mathbf{x} + \mathbf{u}(\mathbf{x}; n)) \approx I_{ref}(\mathbf{x}) \;\; \forall\, n \in \{1, \ldots, F\}$

Assuming

- Presence of non-rigid motion.
- Smooth deformations.

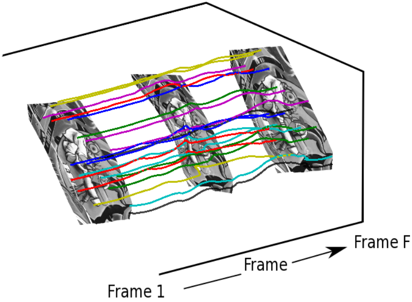
Trajectory Basis

$$\underbrace{\begin{bmatrix} \mathbf{u}(\mathbf{x};1) \\ \mathbf{u}(\mathbf{x};2) \\ \vdots \\ \mathbf{u}(\mathbf{x};F) \end{bmatrix}}_{\mathbf{x} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}}$$

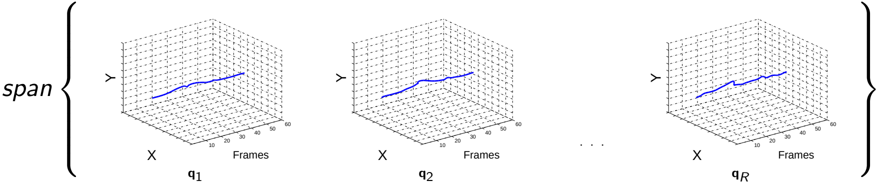


Trajectory Basis

$$\underbrace{\begin{bmatrix} \mathbf{u}(\mathbf{x};1) \\ \mathbf{u}(\mathbf{x};2) \\ \vdots \\ \mathbf{u}(\mathbf{x};F) \end{bmatrix}}_{\mathbf{x}\in\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}}$$



$\downarrow$ lie in



$$R \ll 2F$$

Optical flow with soft constraint

Our proposed multi-frame optical flow energy:

$$E\left[u(x;n),\, L(x)\right] = \alpha \overbrace{\int_{\Omega} \sum_{n=1}^{F} \left| I\left(x + u(x;n)\;;\; n\right) - I(x;n_0)\right| \, dx}^{Brightness\;\; Constancy}$$

Optical flow with soft constraint

Our proposed multi-frame optical flow energy:

$$E\left[u(x;n)\,,\,L(x)\right] = \alpha \overbrace{\int_\Omega \sum_{n=1}^{F} \left| I\left(x + u(x;n)\;;\;n\right) - I(x;n_0)\right|\, dx}^{Brightness\;\;Constancy}$$

$$+ \underbrace{\beta \int_\Omega \sum_{n=1}^{F} \left\| u(x;n) - \sum_{i=1}^{R} \mathbf{q}_i(n) L_i(x) \right\|^2 dx}_{Subspace\;\;Constraint}$$

Optical flow with soft constraint

Our proposed multi-frame optical flow energy:

$$E\left[u(x;n)\,,\,L(x)\right] = \alpha \overbrace{\int_\Omega \sum_{n=1}^{F} \left| I\left(x + u(x;n)\;;\;n\right) - I(x;n_0)\right| \,dx}^{Brightness\;\;Constancy}$$

$$+ \beta \underbrace{\int_\Omega \sum_{n=1}^{F} \left\| u(x;n) - \sum_{i=1}^{R} \mathbf{q}_i(n) L_i(x) \right\|^2 dx}_{Subspace\;\;Constraint} + \underbrace{\int_\Omega \sum_{i=1}^{R} \|\nabla L_i(x)\| \,dx}_{Spatial\;\;Regularisation}$$

Optical flow with soft constraint

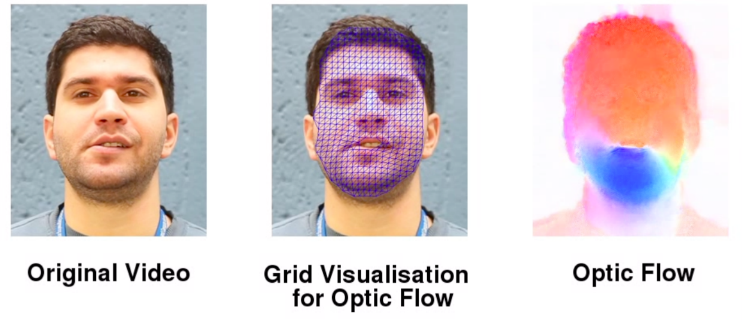
Our proposed multi-frame optical flow energy:

$$E\left[u(x;n)\,,\;L(x)\right]=\alpha\overbrace{\int_{\Omega}\sum_{n=1}^{F}\left|I\left(x+u(x;n)\;;\;n\right)-I(x;n_0)\right|\,dx}^{Brightness\;\;Constancy}$$
$$+\,\beta\underbrace{\int_{\Omega}\sum_{n=1}^{F}\left\|u(x;n)-\sum_{i=1}^{R}\mathbf{q}_i(n)L_i(x)\right\|^2dx}_{Subspace\;\;Constraint}\;+\;\underbrace{\int_{\Omega}\sum_{i=1}^{R}\|\nabla L_i(x)\|\,dx}_{Spatial\;\;Regularisation}$$

Efficient Primal-Dual Optimization. Highly parallel $\rightarrow$ GPU

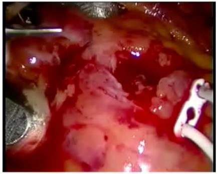
Results: Dense Video Registration – Face sequence

- Robust Subspace Constraints for Video Registration.

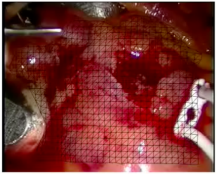


Results: Dense Video Registration – Heart sequence

- Robust Subspace Constraints for Video Registration.



Original Video

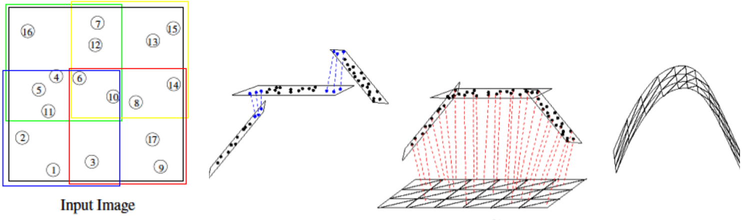


Grid Visualisation for Optic Flow

Optic Flow

Piecewise Reconstruction for Extreme Deformations

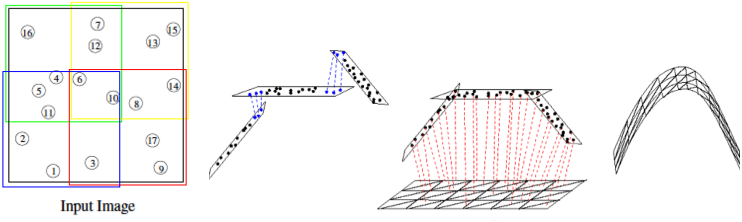
- **Intuition:** Local models have fewer parameters therefore they are easier to optimise and less prone to overfitting.



(Varol et al. ICCV'09)

Piecewise Reconstruction for Extreme Deformations

- **Intuition:** Local models have fewer parameters therefore they are easier to optimise and less prone to overfitting.



(Varol et al. ICCV'09)

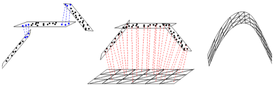
- **Simple idea:**

1. The first image is divided into overlapping patches.
2. The patches are reconstructed independently in 3D.
3. Overlapping points are used to enforce spatial consistency stitch patches to create a continuous global surface.

Piecewise Reconstruction of Deformable Surfaces

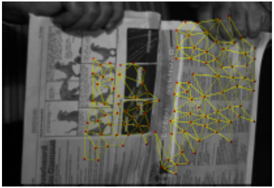
Existing approaches

Piecewise Planar



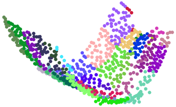
(Varol *et al.* ICCV'09)

Local Isometry



(Taylor *et al.* CVPR'10)

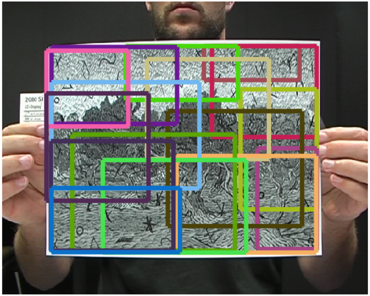
Piecewise Quadratic



(Fayad *et al.* ECCV'10)

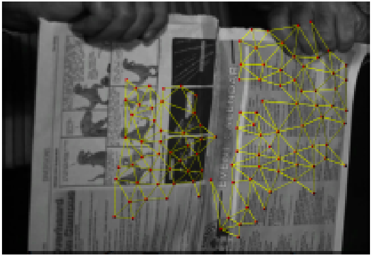
Division into overlapping patches

Fayad-Del Bue-Agapito ECCV'10



Regular grid

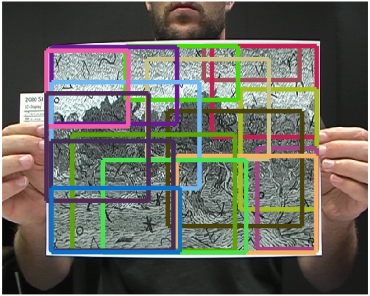
Taylor *et al.* CVPR'10



Too granular

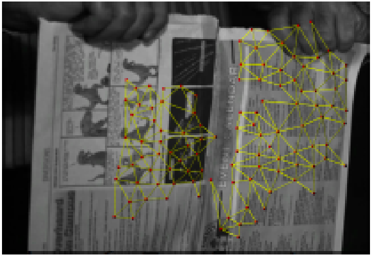
Division into overlapping patches

Fayad-Del Bue-Agapito ECCV'10



Regular grid

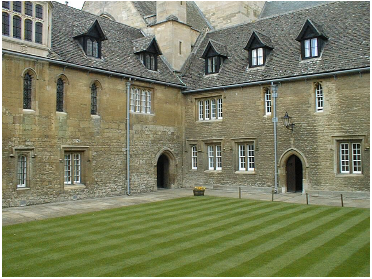
Taylor *et al.* CVPR'10

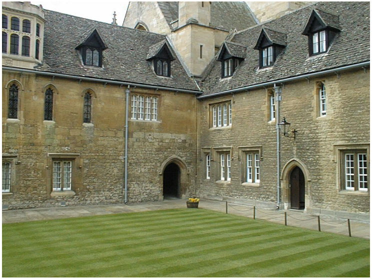


Too granular

- What is the best division of points into patches?
- What is the **best assignment of points to models?**

Example: Piecewise Reconstruction





Reconstruct as piecewise planar from 2 frames.

Geometric Multiple Model Fitting

Isack and Boykov IJCV '11 PEARL

- The problem of multiple model fitting is formulated as a labelling problem using a Markov Random Field.
- The labels x_i represent the parameters of a model that must be fitted to the data.

Geometric Multiple Model Fitting

Isack and Boykov IJCV '11 PEARL

- The problem of multiple model fitting is formulated as a labelling problem using a Markov Random Field.
- The labels x_i represent the parameters of a model that must be fitted to the data.

1. Propose an excess of models
2. Assign best model to each point – minimize reprojection error

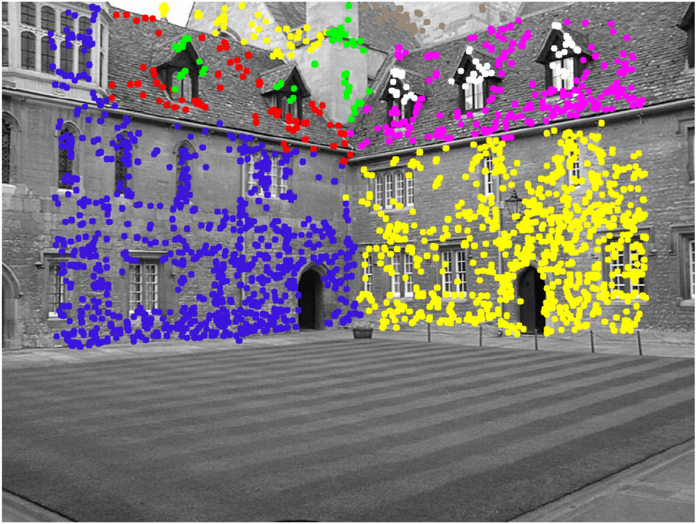
$$\min_{\mathbf{x}} E(\mathbf{x}) = \sum_{i \in \mathcal{P}} U_i(x_i)$$

3. Add Pairwise Potentials and Sparsity inducing MDL prior

$$\min_{\mathbf{x}} E(\mathbf{x}) = \sum_{i} U_i(x_i) + \sum_{(i,j) \in \mathcal{N}} P_{i,j}(x_i, x_j) + \mathrm{MDL}(\mathbf{x})$$

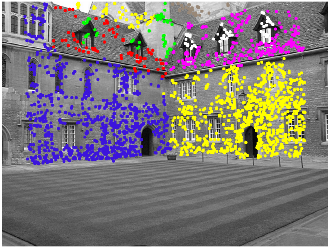
4. Alternate between assigning points to models and re-estimating the parameters of the models.

Multiple Model Fitting



Isack and Boykov IJCV '11

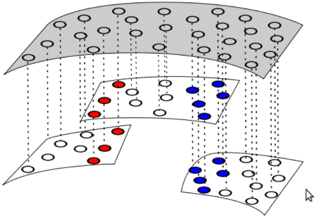
What if we want to do reconstruction?



If every model is independent each plane has a separate depth ambiguity.

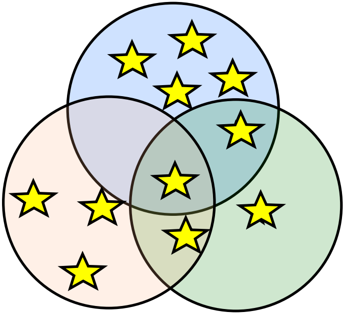
- Resolve depth by matching camera motion for each model.
- Only works for rigid scenes.

Overlapping Models

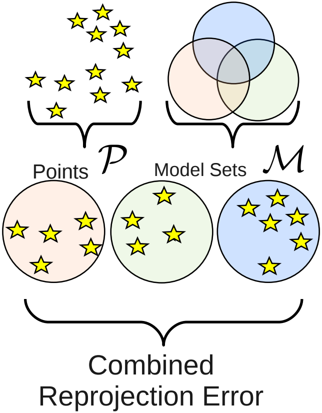


- Instead, we need to segment *overlapping* models, and force reconstructions to agree about the regions of overlap.
- Double count points in overlap - penalises long/jagged boundaries
- Implicit higher-order smoothing comes from forcing models to explain the same data.

Making Models Overlap



$$\sum_{p \in \mathcal{P}} \left(\sum_{\alpha \in m_p} U_p(\alpha) \right)$$



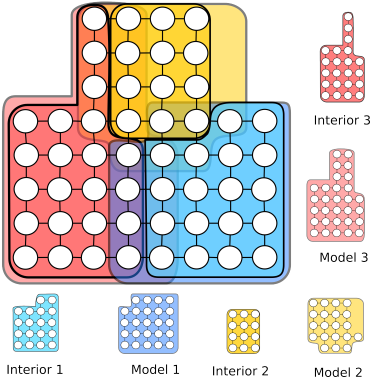
Points are assigned to a *sub-set* of models rather than just one

Trick: Use topological constraints

Russell-Fayad-Agapito CVPR'11

New labels: A point is an **interior point** of a model, if all of its neighbours belong to the same model

Key idea: Each point must be an interior point of at least 1 model



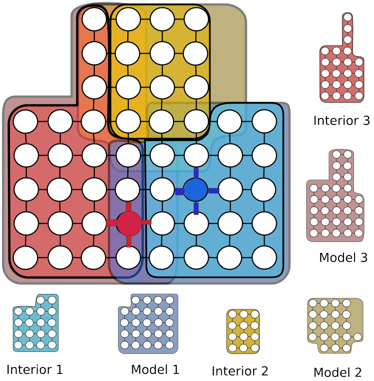
The cost of a point taking an interior label requires adding the cost of all its neighbours taking the same label.

Trick: Use topological constraints

Russell-Fayad-Agapito CVPR'11

New labels: A point is an **interior point** of a model, if all of its neighbours belong to the same model

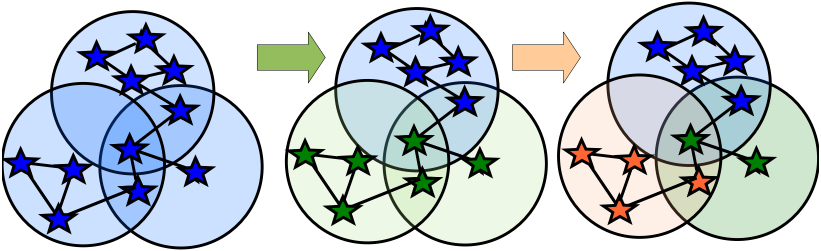
Key idea: Each point must be an interior point of at least 1 model



The cost of a point taking an interior label requires adding the cost of all its neighbours taking the same label.

We solve this with α-expansion

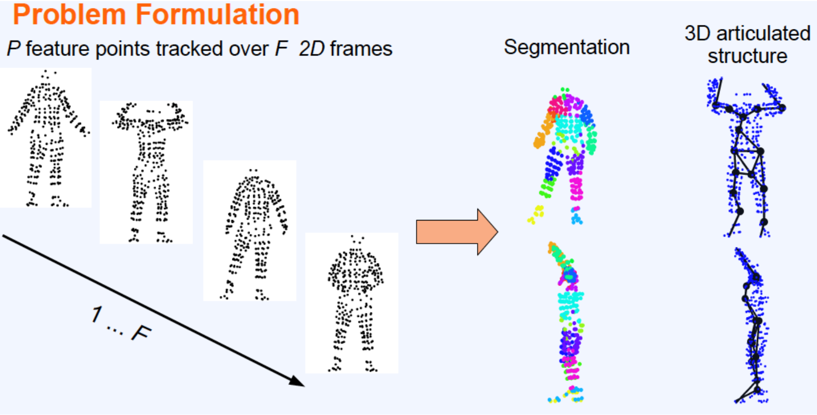
Each point belongs to the interior of exactly 1 model
"Only" $\mathcal{M}^{\mathcal{P}}$ possible solutions instead of $2^{\mathcal{M}^{\mathcal{P}}}$



Our code is publicly available!

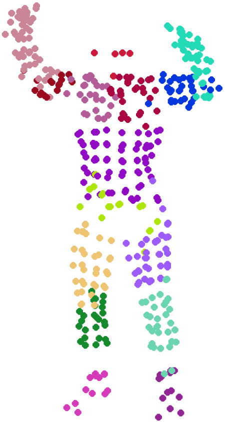
Applications: Articulated Structure from Motion

Goal: To recover the 3D shape and pose of an articulated object from a monocular video using only 2D tracking data as input.



Articulated Structure from Motion (ICCV '11)

Fayad-Russell-Agapito ICCV 2011



Articulated motion is piecewise rigid motion.
Areas of overlap are joints.

Articulated Motion

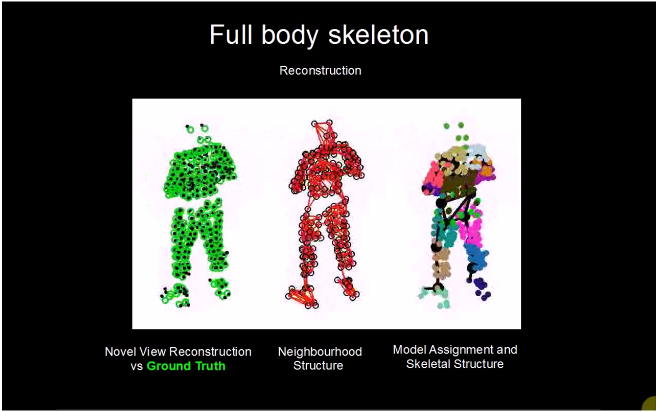
- Simultaneous segmentation of the body into rigid segments and 3D reconstruction.
- Estimating the model parameters $\mathtt{R}_i$, t_i and $\mathtt{S}$ for each rigid segment can be formulated as the factorization problem which minimizes image reprojection error:

$$\arg\min_{\mathtt{R}_i, t_i, \mathtt{S}} \sum_{i=1}^{F} ||\tilde{\mathtt{W}_\mathtt{i}} - \mathtt{R}_i \mathtt{S} - t_i||^2$$

- Decomposing the points into rigid pieces, allows us to use pre-existing robust methods to reconstruct partial tracks (Marques+Costeira'08).
- We create a skeleton by 'joining up' the joint centres

Articulated Motion: Results

Fayad-Russell-Agapito ICCV 2011

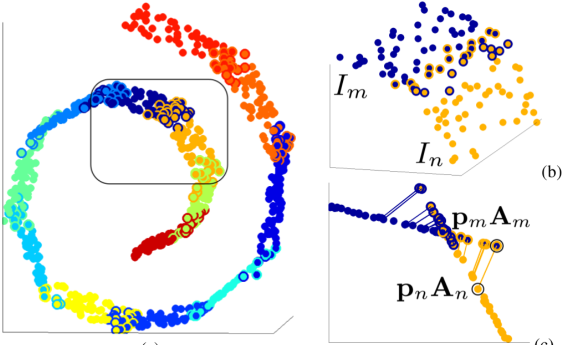


Piecewise Rigid Reconstruction: Input

Piecewise Rigid Reconstruction: Results

Learning Manifolds as Connected Linear Subspaces

Pitelis-Russell-Agapito CVPR'13

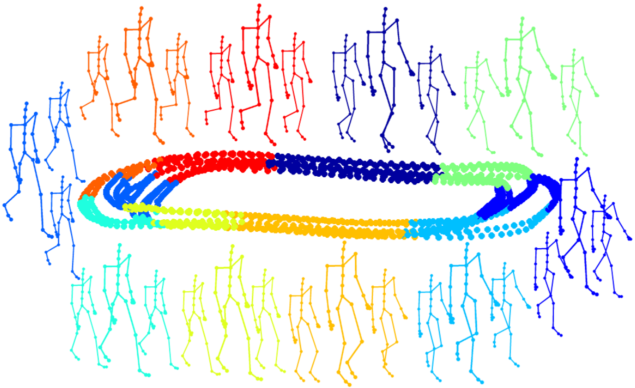


(a) (b) (c)

- Alternate between
  - Assigning points to overlapping local linear subspaces.
  - Fitting subspaces to points using PCA.
- Advantages:
  - Can model closed manifolds.
  - No need to unfold.

Results: 3D Motion Reconstruction

1. Manifold is learnt on 3D mocap data from 1 individual.
2. Input: 2D data from a different individual.
3. Inverse problem: reconstruct 3D pose given 2D data + manifold.



Direct Dense Multibody SfM from Pixel Intensities

Roussos-Russell-Garg-Agapito ISMAR'12

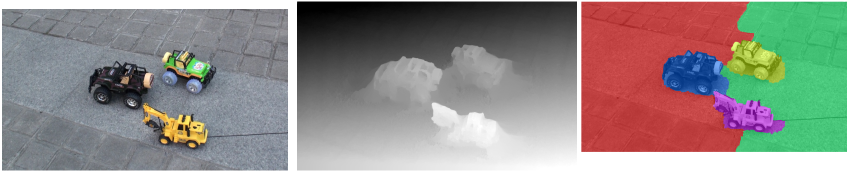
Problem statement: Given a scene with an unknown number of independently moving objects estimate automatically:

1. Segmentation of the scene into objects.
2. Dense 3D model of each object.
3. Rigid motion of the objects.

Directly from pixel intensities not from tracks or optical flow.

Multiple Model Fitting

- Strength of our approach: to optimize a single cost function to estimate all the variables:
  1. Dense depth map: $d(\mathbf{x}) : \Omega \to \mathbb{R}$
  2. Labelling: $L(\mathbf{x}) : \Omega \to \{1, \dots, N\}$
  3. Rigid transformations: $T_{\ell f}$
  4. Number of regions: N



Multiple Model Fitting

$$E[d(\mathbf{x}), L, \mathcal{T}] = \lambda E_{data} + E_{reg} + \beta E_{potts} + \gamma \mathrm{MDL}.$$

- E_{data}: average photometric cost.
- E_{reg}: TV regulariser over the depth map.
- E_{potts}: regulariser over the segmentation.
- MDL: induces sparsity.

Multiple Model Fitting

$$E[d(\mathbf{x}), L, \mathcal{T}] = \lambda E_{data} + E_{reg} + \beta E_{potts} + \gamma \mathrm{MDL}.$$

- E_{data}: average photometric cost.
- E_{reg}: TV regulariser over the depth map.
- E_{potts}: regulariser over the segmentation.
- MDL: induces sparsity.

$$E = \lambda \int_\Omega C(\mathbf{x}, d(\mathbf{x}), \mathcal{T}_{L(\mathbf{x})}) + \int_\Omega \|\nabla d\|_\sigma \ \mathrm{d}\mathbf{x} + \beta \sum_{\ell=1}^{N} \mathrm{Per}(\Omega_\ell) + \gamma \mathrm{MDL}.$$

Optimization

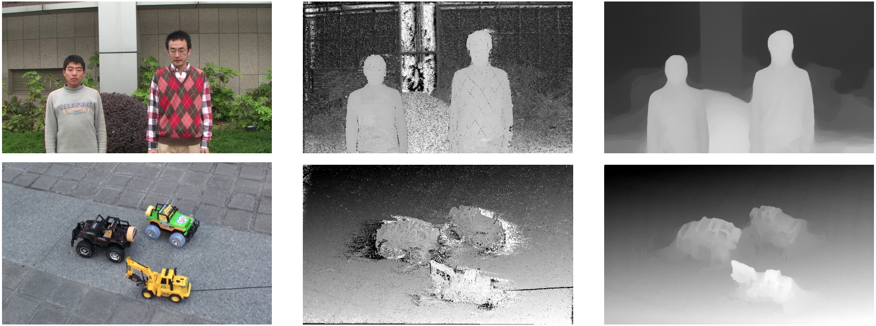
Our algorithm is a **hill-climbing** approach that alternates between:

1. Estimating **depth-map** given a segmentation and rigid motions. We use a dense variational approach (DTAM Newcombe-Lovegrove-Davison ICCV'11)
2. Estimating **segmentation** given rigid motions and a depth-map. We use the inference approach of Isack and Boykov IJCV'11 (PEARL)
3. Estimating **rigid motions** given depth-map and segmentation. We use our novel dense tracking algorithm.

Until convergence.

Results: Dense Depth Map Estimation

- Strategy similar to **DTAM**[1].



Reference frames | Initialization of $d(\mathbf{x})$ | Final regularized $d(\mathbf{x})$

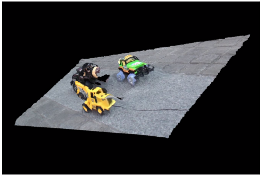
[1] R. Newcombe, S. Lovegrove, A. Davison. DTAM: Dense tracking and mapping in real-time. ICCV'11.

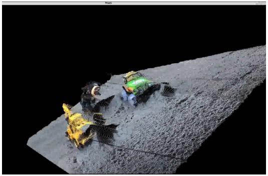
Results: Segmentation

Initialization

Segmentation using E_{potts} and MDL

Results: Dense 3D reconstructions





Augmented Reality applications

- Accurate estimation of the surfaces geometry.
- High quality occlusion reasoning.

Augmented Reality applications

- Accurate estimation of the 3D motions.
- High quality occlusion reasoning.

Conclusions: Future Directions

- **Unified Framework:** 3D shape estimation directly from pixel intensity information. Simultaneous 2D registration + segmentation + 3D reconstruction.
- **Reconstruction in the Wild:** Robustify approaches, dealing with multiple non-rigid objects, severe occlusions, introduction of new tracks, changes in lighting. Complete reconstruction of almost *any video*.
- **Towards Real-Time:** Online formulation of NRSfM – dense non-rigid models *on-the-fly*.
- **Focus on Domain-specific Applications:** laparoscopy, augmented reality, cinema post-production, human motion reconstruction. Semisupervised and unsupervised learning where additional 3D structural information could supplement limited annotation data.