Automatic 3D Object Segmentation in Multiple Views using Volumetric Graph-Cuts

Neill Campbell[1] George Vogiatzis[2]
Carlos Hernández[2] Roberto Cipolla[1]

[1]**Department of Engineering**
University of Cambridge
Trumpington Street
Cambridge
CB2 1PZ
UK

[2]**Toshiba Research Europe**
208 Cambridge Science Park
Milton Road
Cambridge
CB4 0GZ
UK

British Machine Vision Conference, 2007

Motivation

Motivation

State of the Art 3D Reconstruction

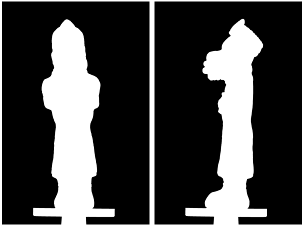
Automatic acquisition of 3D models from image sequences with known camera calibration

Motivation

State of the Art 3D Reconstruction

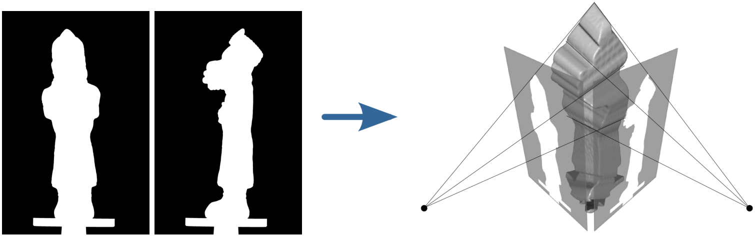
Automatic acquisition of 3D models from image sequences with known camera calibration

Motivation



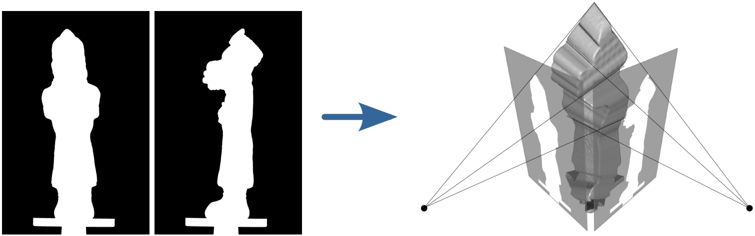
Object silhouettes (the visual hull) required to provide:

Motivation



Object silhouettes (the visual hull) required to provide:

Motivation



Object silhouettes (the visual hull) required to provide:

- Approximate initial reconstruction
- Outer bound for the reconstructed object
- Approximate occlusion reasoning

Motivation

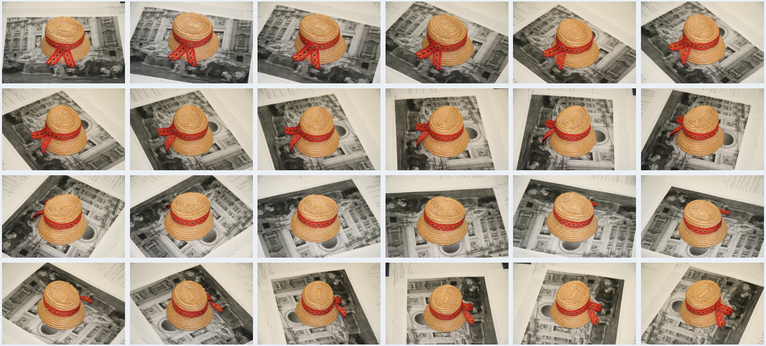
Automatic reconstruction requires automatic object segmentation across the multiple views of a calibrated image sequence

Motivation

Automatic reconstruction requires **automatic object segmentation** across the multiple views of a calibrated image sequence

Motivation

Automatic reconstruction requires **automatic object segmentation** across the multiple views of a calibrated image sequence



Motivation

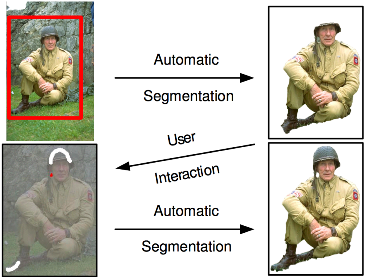
Automatic reconstruction requires **automatic object segmentation** across the multiple views of a calibrated image sequence

Overview

Overview: Segmentation

Current approach to image segmentation:

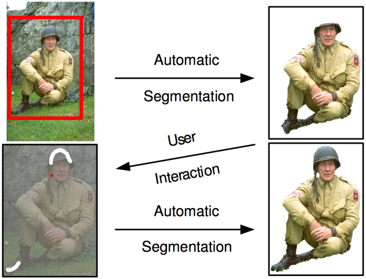
- User interactive approach



Overview: Segmentation

Current approach to image segmentation:

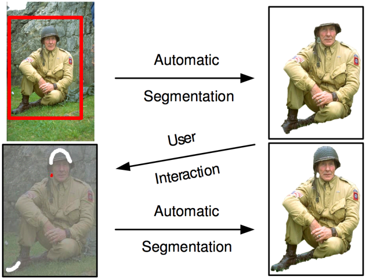
- User interactive approach
- Each image segmented independently (in 2D)



Overview: Segmentation

Current approach to image segmentation:

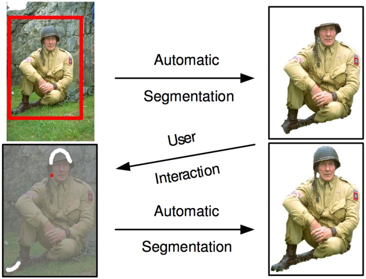
- User interactive approach
- Each image segmented independently (in 2D)
- Statistical models for object and background



Overview: Segmentation

Current approach to image segmentation:

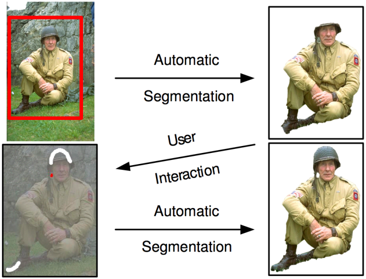
- User interactive approach
- Each image segmented independently (in 2D)
- Statistical models for object and background
- Energy cost function minimised
    - Graph-Cuts — globally optimal
    - Level-Sets — prone to local minima



Overview: Segmentation

Current approach to image segmentation:

- User interactive approach
- Each image segmented independently (in 2D)
- Statistical models for object and background
- Energy cost function minimised
  - Graph-Cuts — globally optimal
  - Level-Sets — prone to local minima



- This is a sizeable task for a large image sequence

Overview: Automatic Segmentation

An automatic algorithm needs to remove the demands placed on the user:

Exploit two constraints

Overview: Automatic Segmentation

An automatic algorithm needs to remove the demands placed on the user:

Exploit two constraints

1. Silhouette coherency constraint arising from object rigidity

Overview: Automatic Segmentation

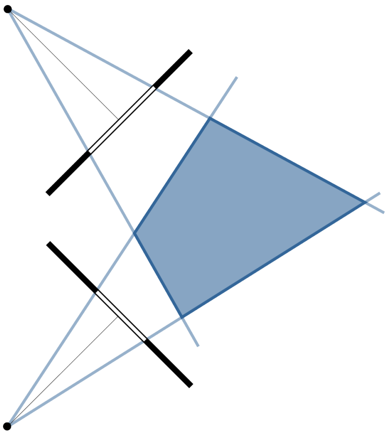
An automatic algorithm needs to remove the demands placed on the user:

Exploit two constraints

1. Silhouette coherency constraint arising from object rigidity
2. Fixation constraint from views focusing on object

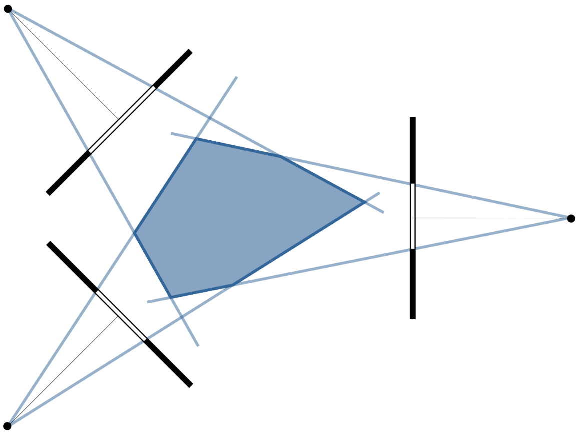
Overview: Silhouette Coherency

The object silhouettes may be combined to form a visual hull:

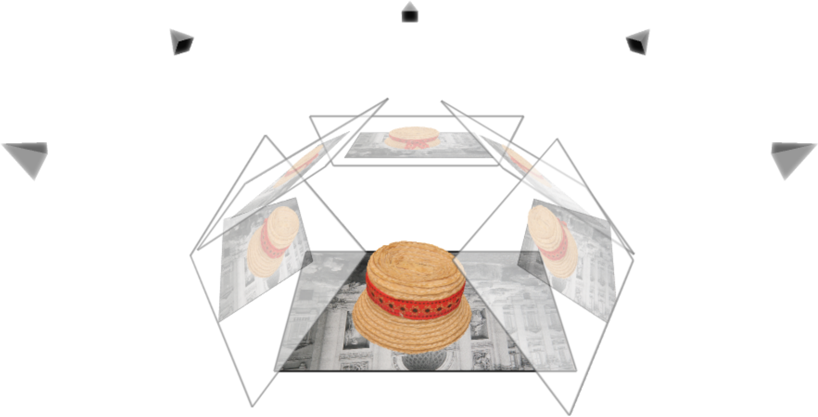


Overview: Silhouette Coherency

The object silhouettes may be combined to form a visual hull:

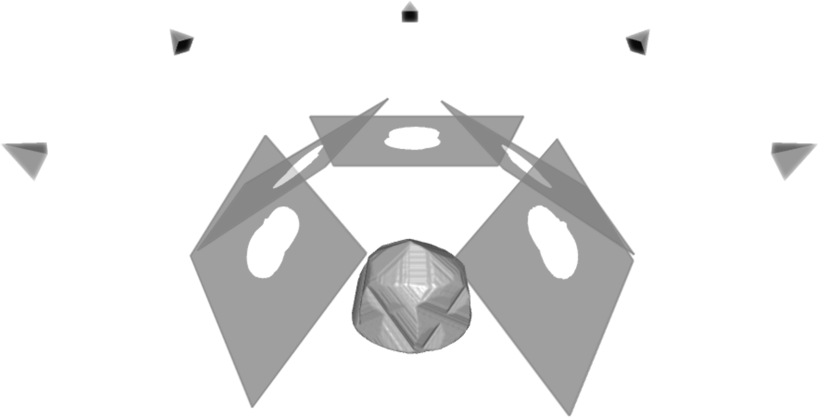


Overview: Silhouette Coherency



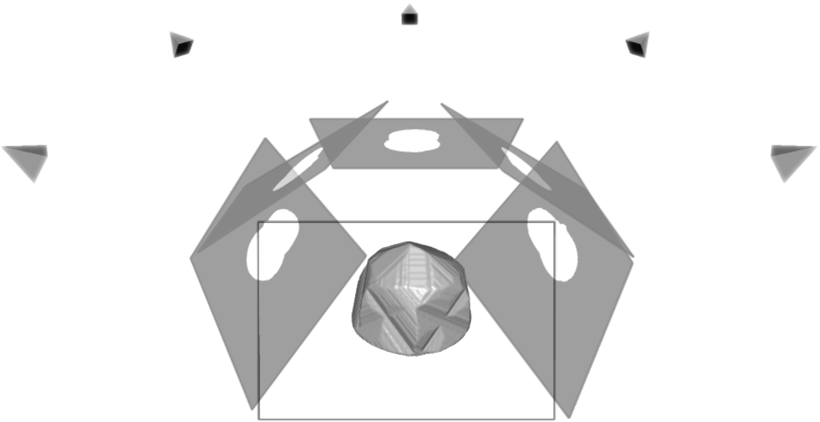
Take a set of views of the object

Overview: Silhouette Coherency



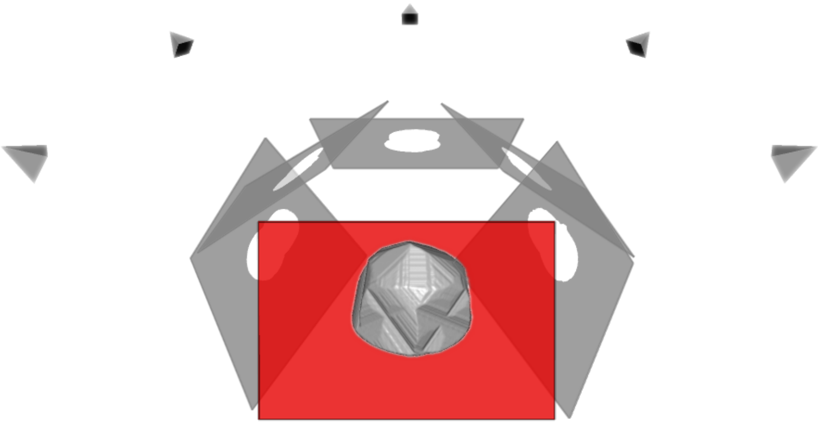
Form the visual hull from perfect segmentations

Overview: Silhouette Coherency



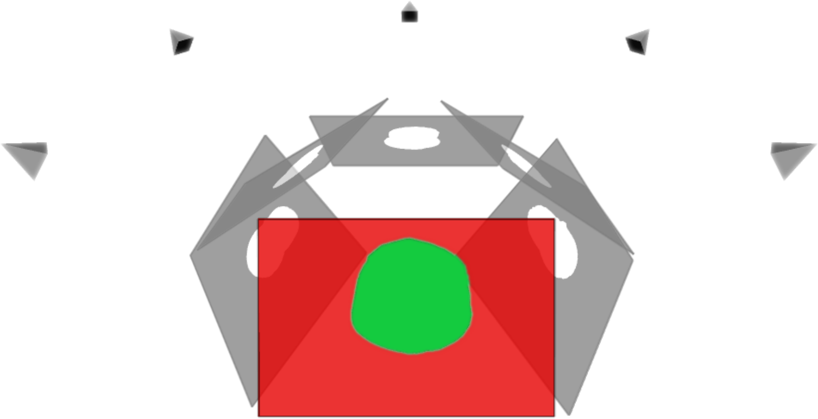
Now when we come to segment a new image

Overview: Silhouette Coherency



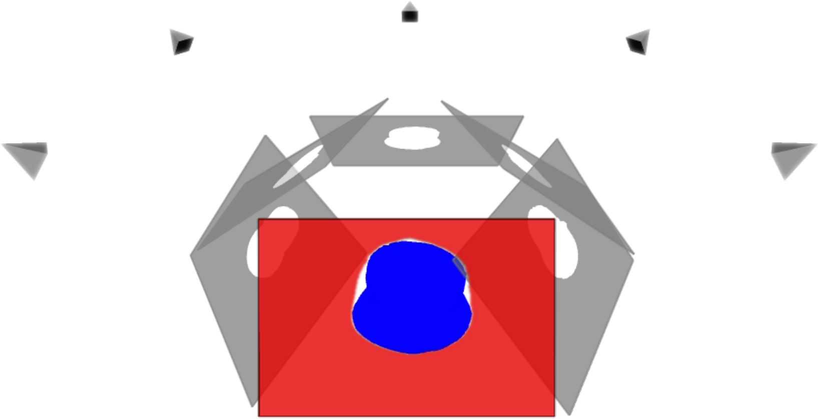
We know that the silhouette must lie inside the existing visual hull

Overview: Silhouette Coherency



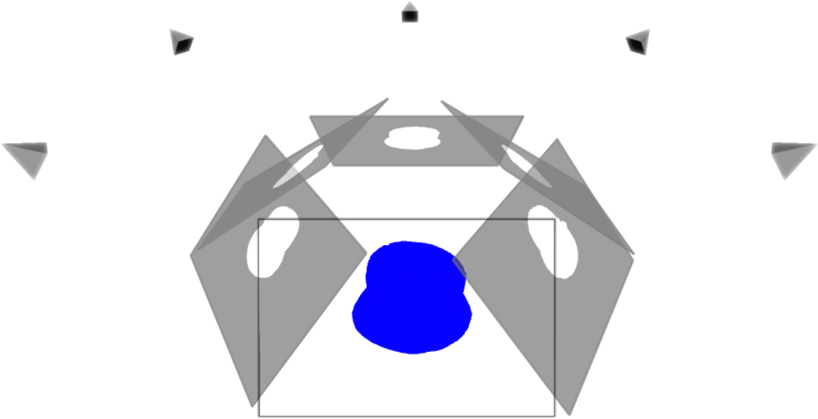
We know that the silhouette must lie inside the existing visual hull

Overview: Silhouette Coherency



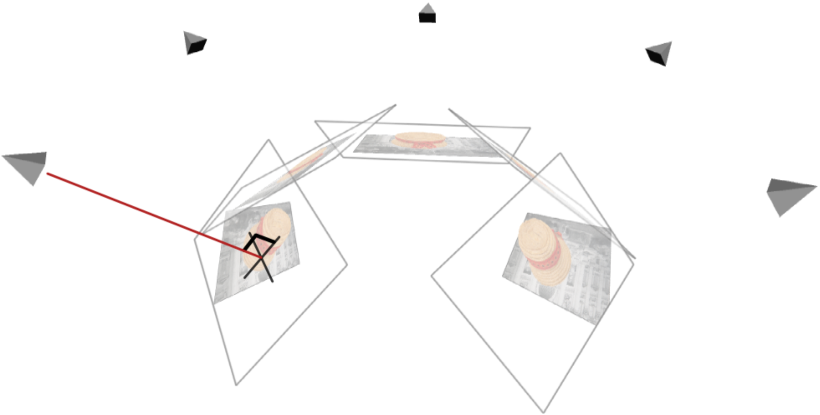
We know that the silhouette must lie inside the existing visual hull

Overview: Silhouette Coherency



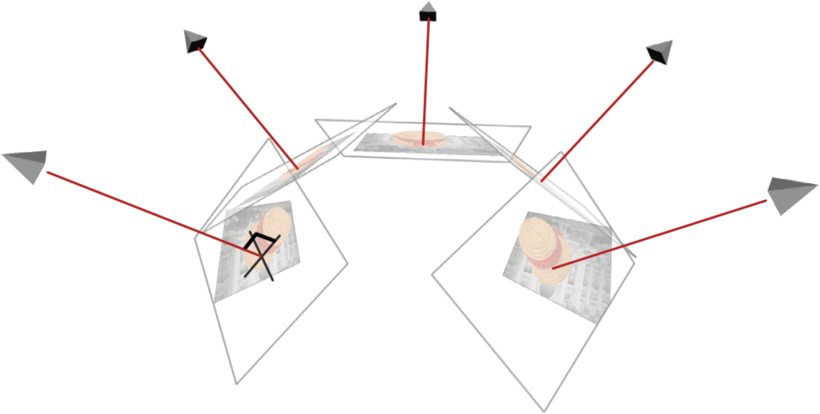
We know that the silhouette must lie inside the existing visual hull

Overview: Fixation Constraint



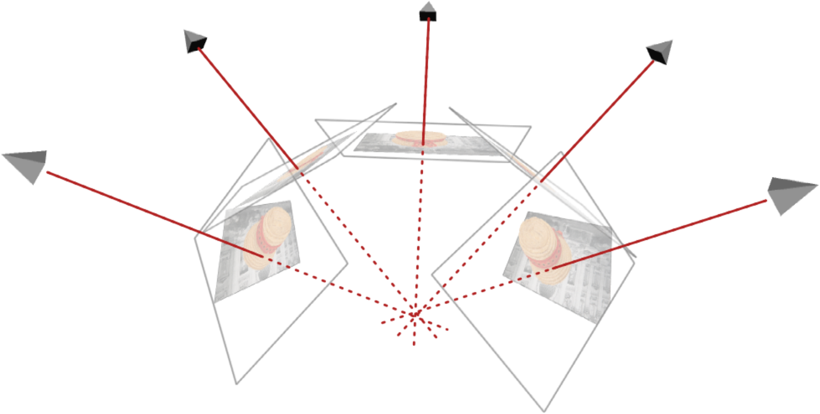
Locate the optical centre for each image (optical axes)

Overview: Fixation Constraint



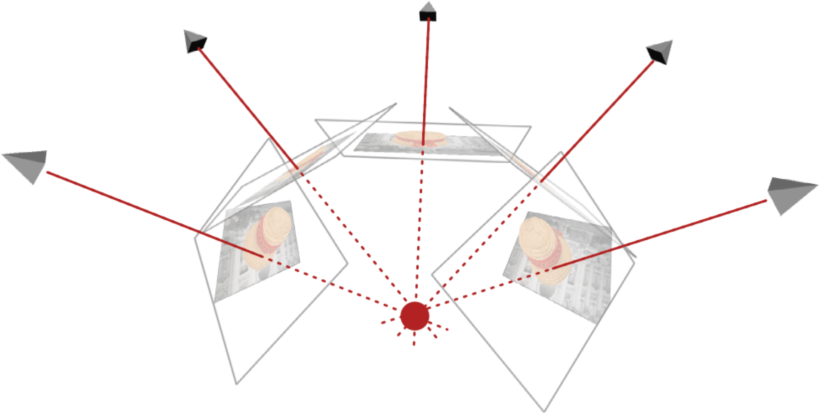
Locate the optical centre for each image (optical axes)

Overview: Fixation Constraint



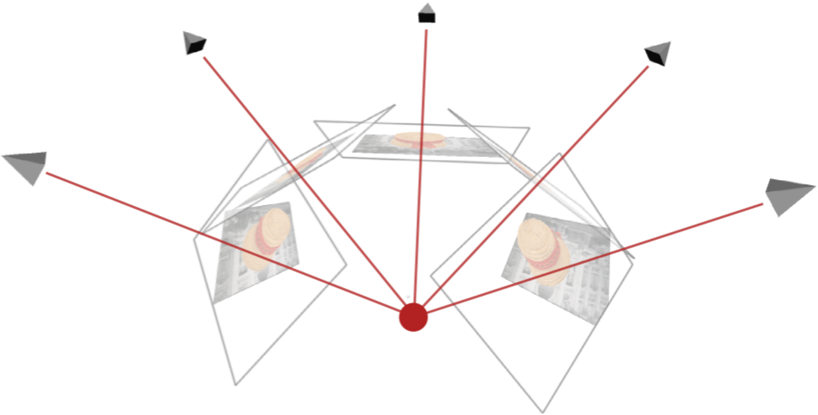
Project the optical axes into the volume

Overview: Fixation Constraint



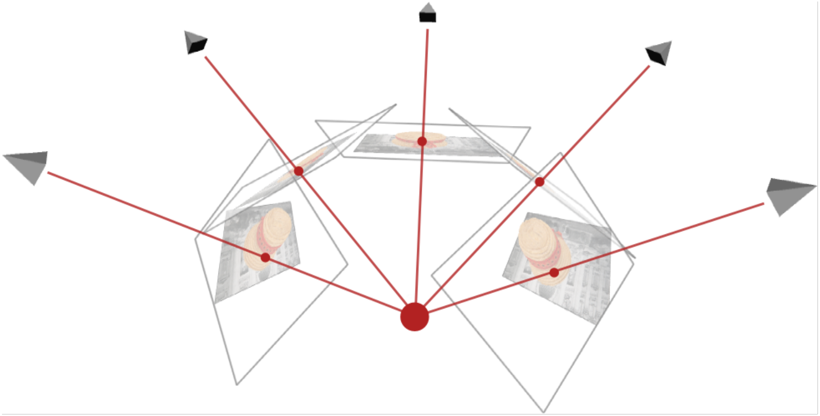
Locate the centroid of intersection (least squares fit)

Overview: Fixation Constraint



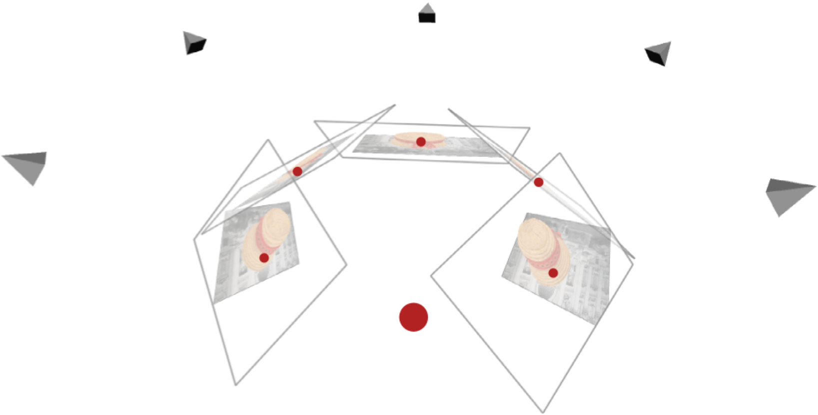
Use this centroid to back project into the original images...

Overview: Fixation Constraint



Use this centroid to back project into the original images...

Overview: Fixation Constraint



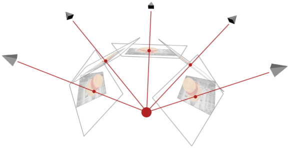
...to find initialisation locations in each image

Overview: Automatic Segmentation in 3D

Our approach offers the following advantages over the independent 2D approach:

Overview: Automatic Segmentation in 3D

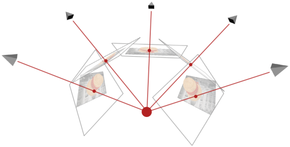
Our approach offers the following advantages over the independent 2D approach:



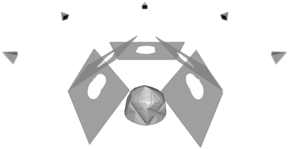
Exploit the fixation condition for automatic initialisation

Overview: Automatic Segmentation in 3D

Our approach offers the following advantages over the independent 2D approach:



Exploit the fixation condition for automatic initialisation



Perform segmentation across all images simultaneously in 3D, enforcing object rigidity via *silhouette coherency*

Overview: Automatic Segmentation in 3D

Our Approach:

Overview: Automatic Segmentation in 3D

Our Approach:

- Iteratively learn colour models for the object and background

Overview: Automatic Segmentation in 3D

Our Approach:

- Iteratively learn colour models for the object and background
- Adopt an energy based cost function

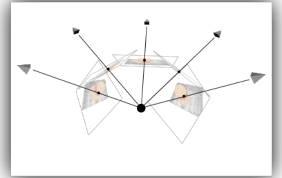
Overview: Automatic Segmentation in 3D

Our Approach:

- Iteratively learn colour models for the object and background
- Adopt an energy based cost function
- Use volumetric Graph-Cuts on a voxel array (direct estimation of the visual hull) to perform segmentation

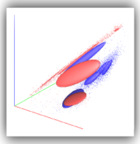
Automatic Segmentation Algorithm

Automatic Segmentation Algorithm: Initialisation



Calibrated Image Sequence → Use Fixation Points to Initialise Colour Models →

Automatic Segmentation Algorithm: Iteration Sequence

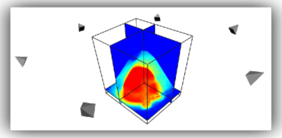


Perform Graph-Cut Segmentation to Produce Visual Hull

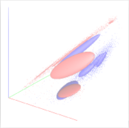
Update Object and Background Colour Models

Generate Classification Probabilities

Update Voxel Graph with new Values



Automatic Segmentation Algorithm: Convergence



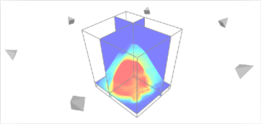
Perform Graph-Cut Segmentation
to Prod

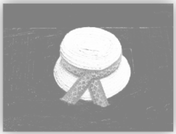
Update Object and
lour Models

**Converge to Accurate
Visual Hull and Silhouettes**

Updat
with new Values

ssification
Probabilities





Building Colour Models

Building Colour Models

The colour model is learnt iteratively and provides a likelihood that individual image pixels are part of the object or the background.

Building Colour Models

The colour model is learnt iteratively and provides a likelihood that individual image pixels are part of the object or the background.

- We use a K component Gaussian Mixture Model (GMM) in RGB colour-space

Building Colour Models

The colour model is learnt iteratively and provides a likelihood that individual image pixels are part of the object or the background.

- We use a K component Gaussian Mixture Model (GMM) in RGB colour-space

$$p(\mathbf{u} \,|\, \pi_k, \mu_k, \Sigma_k) = \sum_{k=1}^{K} p(k)\, p(\mathbf{u} \,|\, \mu_k, \Sigma_k) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\mathbf{u} \,|\, \mu_k, \Sigma_k)$$

Building Colour Models

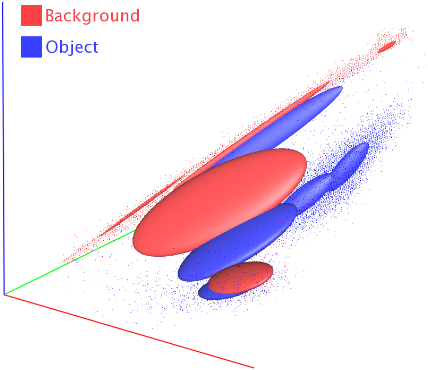
The colour model is learnt iteratively and provides a likelihood that individual image pixels are part of the object or the background.

- We use a K component Gaussian Mixture Model (GMM) in RGB colour-space

$$p(\mathbf{u} \mid \pi_k, \mu_k, \Sigma_k) = \sum_{k=1}^{K} p(k)\, p(\mathbf{u} \mid \mu_k, \Sigma_k) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(\mathbf{u} \mid \mu_k, \Sigma_k)$$

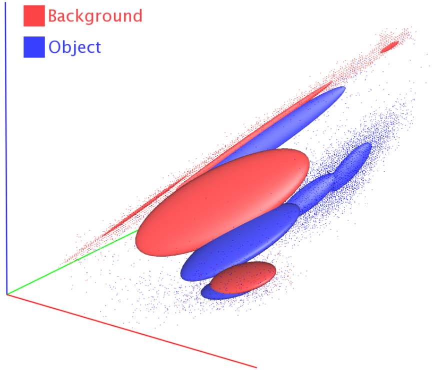
- The Expectation-Maximisation (EM) algorithm used to fit the model parameters

Building Colour Models



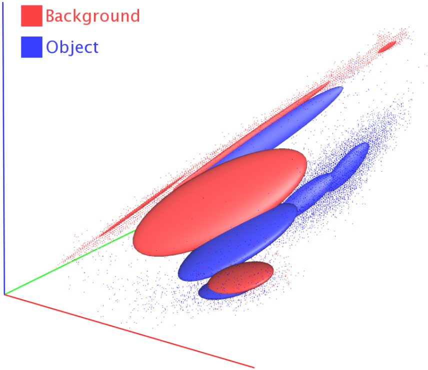
- Single object model, $p(\mathbf{u} \mid \mathrm{object})$

Building Colour Models



- Single object model, $p(\mathbf{u} \mid \text{object})$
- Per image background models, $p(\mathbf{u} \mid \text{background}, \text{image})$

Building Colour Models

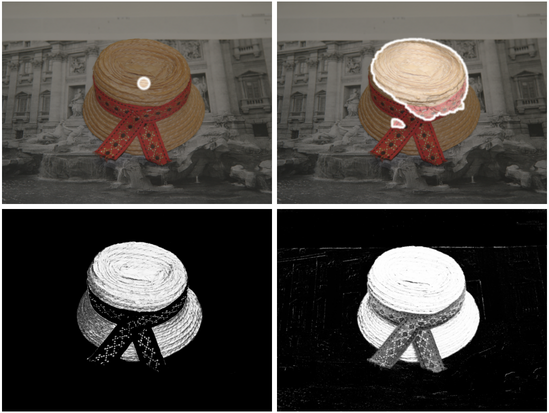


- Single object model, $p(\mathbf{u} \mid \mathrm{object})$
- Per image background models, $p(\mathbf{u} \mid \mathrm{background}, \mathrm{image})$
- Use current iteration silhouettes as masks

Iterative Colour Model Learning

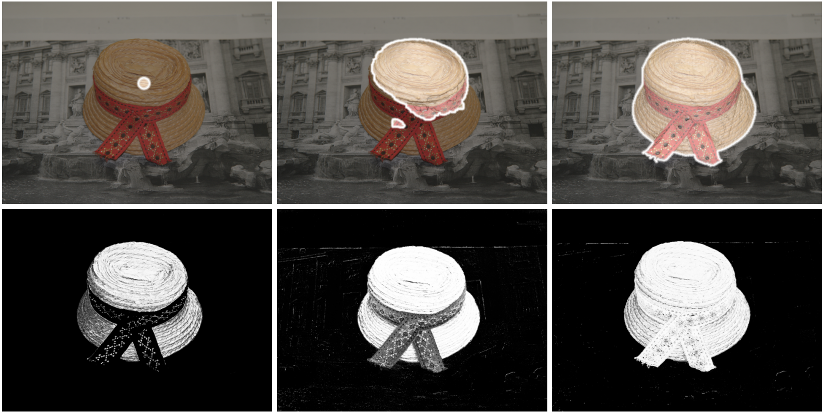
Initialisation

Iterative Colour Model Learning



Initialisation First Iteration

Iterative Colour Model Learning



Initialisation First Iteration Second Iteration

Volumetric Graph-Cut

Volumetric Graph-Cut

Energy cost function comprises two parts

Volumetric Graph-Cut

Energy cost function comprises two parts

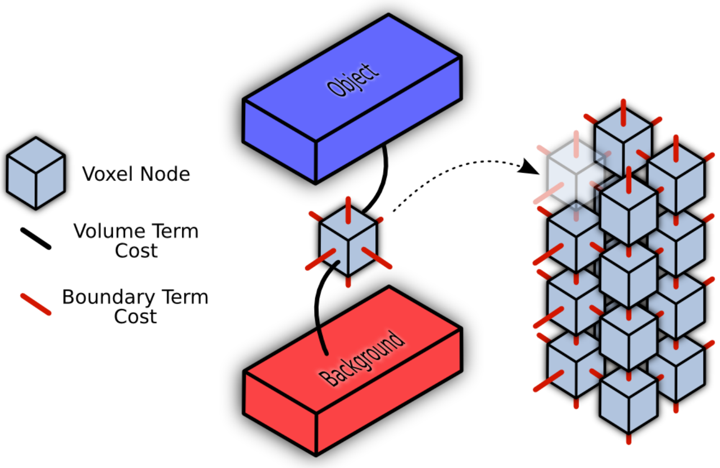
1. Volume term from colour model likelihood

Volumetric Graph-Cut

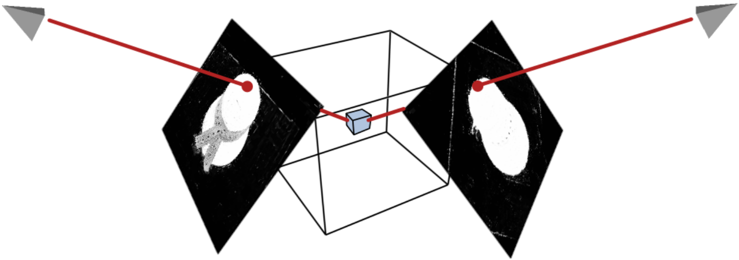
Energy cost function comprises two parts

1. Volume term from colour model likelihood
2. Boundary term from image edges

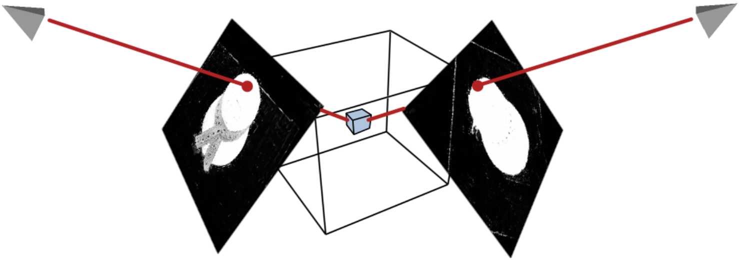
Volumetric Graph-Cut: Voxel Graph



Volumetric Graph-Cut: Volume Term

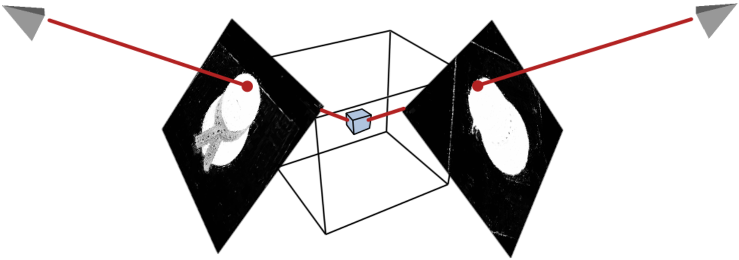


Volumetric Graph-Cut: Volume Term



- Project each node into each image

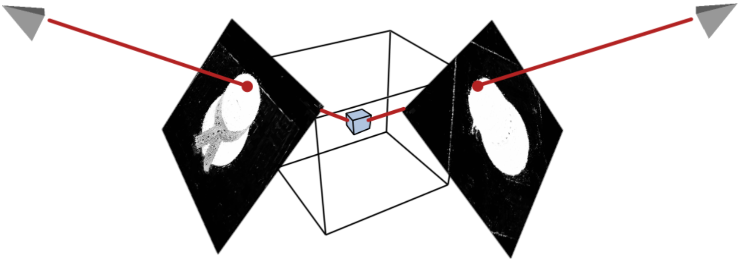
Volumetric Graph-Cut: Volume Term



- Project each node into each image

$$p(\text{object}) = \frac{1}{N} \sum_{\text{images}} \text{Likelihood}(\text{voxel} = \text{object} \mid \text{image})$$

Volumetric Graph-Cut: Volume Term

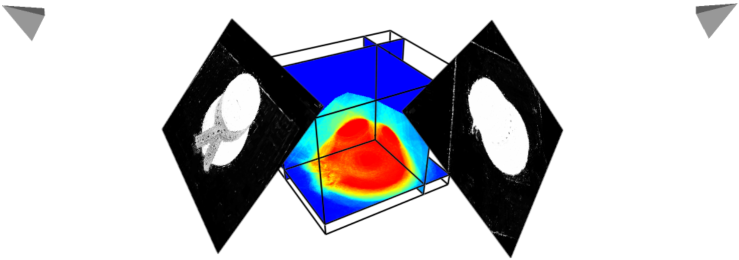


- Project each node into each image

$$p(\text{object}) = \frac{1}{N} \sum_{\text{images}} \text{Likelihood}(\text{voxel} = \text{object} \,|\, \text{image})$$

- Offset against threshold

Volumetric Graph-Cut: Volume Term

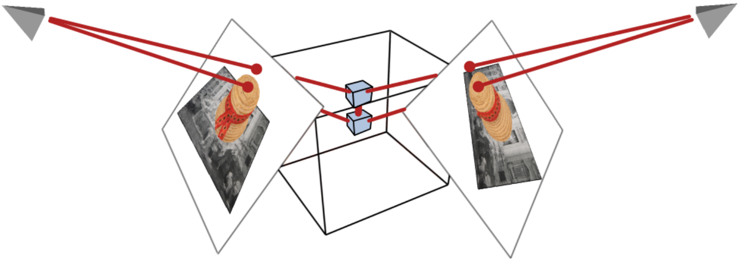


- Project each node into each image

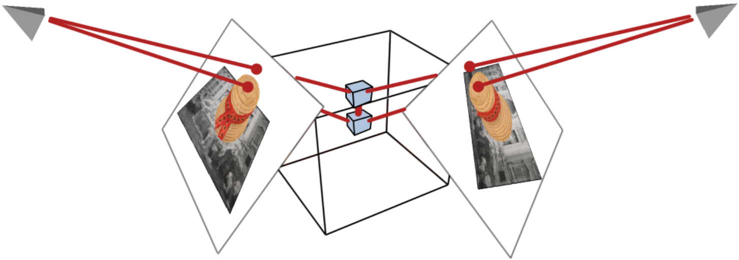
$$p(\text{object}) = \frac{1}{N} \sum_{\text{images}} \text{Likelihood}(\text{voxel} = \text{object} \mid \text{image})$$

- Offset against threshold

Volumetric Graph-Cut: Boundary Term

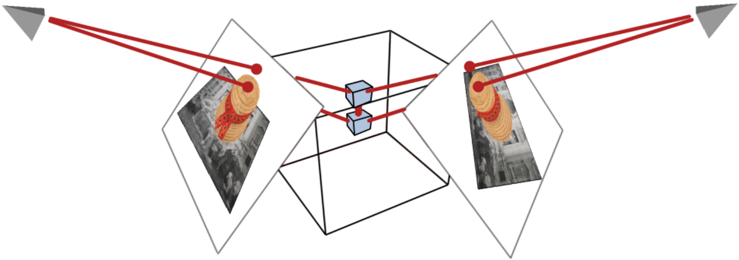


Volumetric Graph-Cut: Boundary Term



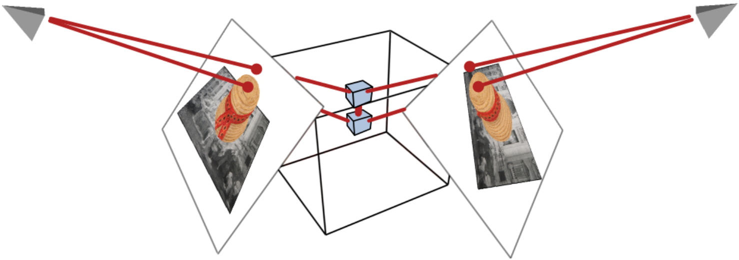
- Project each pair of neighbouring nodes into each image

Volumetric Graph-Cut: Boundary Term



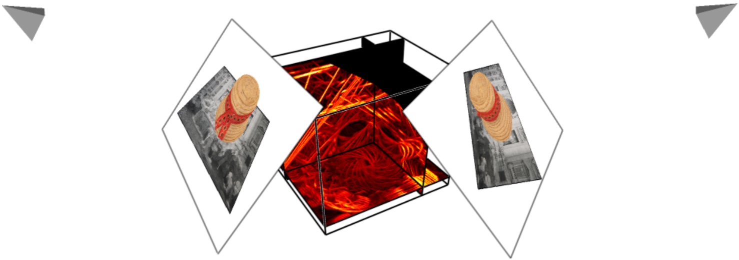
- Project each pair of neighbouring nodes into each image
- Use standard Gibb's model from the image with the maximum colour difference

Volumetric Graph-Cut: Boundary Term



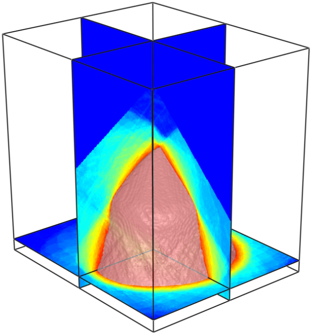
- Project each pair of neighbouring nodes into each image
- Use standard Gibb's model from the image with the maximum colour difference
- Provides cutting planes from image edges

Volumetric Graph-Cut: Boundary Term



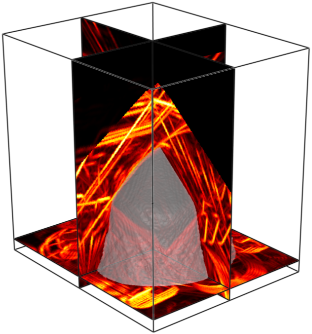
- Project each pair of neighbouring nodes into each image
- Use standard Gibb's model from the image with the maximum colour difference
- Provides cutting planes from image edges

Volumetric Graph-Cut



Final Volume Term

low to high $p(\mathrm{object})$



Final Boundary Term

low to high $p(\mathrm{edge})$

Results

Results: Automatic Calibration

Automation also extended to include automatic calibration of camera intrinsic parameters and pose:

Results: Automatic Calibration

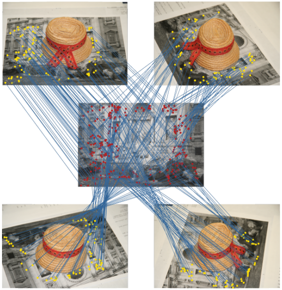
Automation also extended to include automatic calibration of camera intrinsic parameters and pose:

- Can be used with objects located on a textured plane

Results: Automatic Calibration

Automation also extended to include automatic calibration of camera intrinsic parameters and pose:

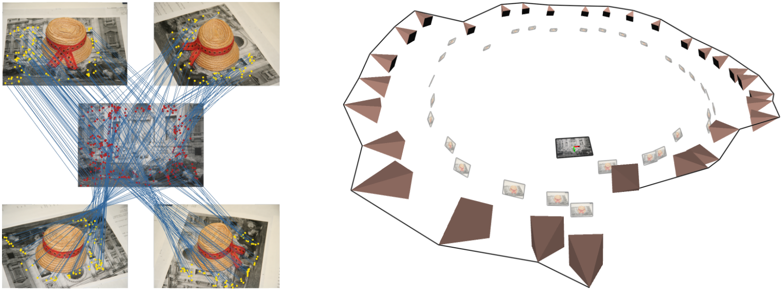
- Can be used with objects located on a textured plane
- Correspondences found and planar homographies estimated



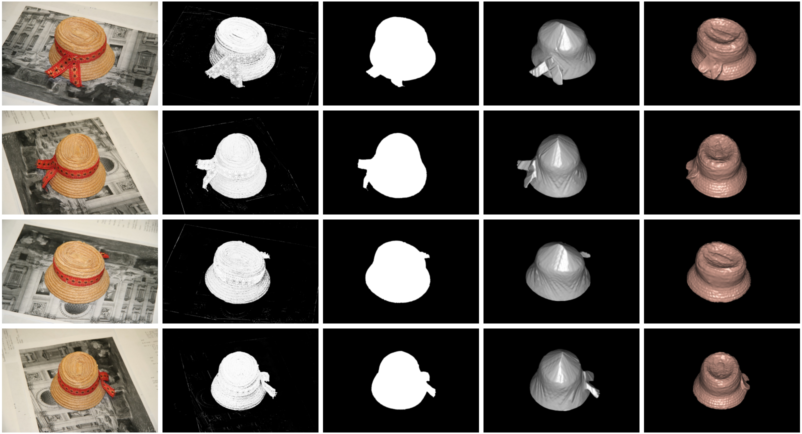
Results: Automatic Calibration

Automation also extended to include automatic calibration of camera intrinsic parameters and pose:

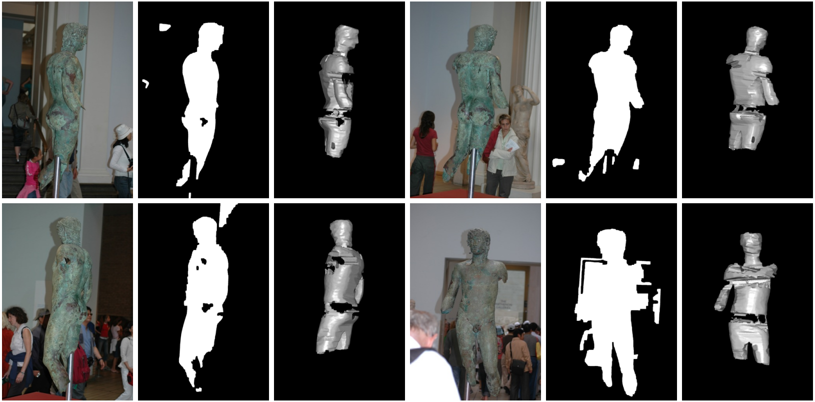
- Can be used with objects located on a textured plane
- Correspondences found and planar homographies estimated
- Non-linear optimisation via bundle adjustment



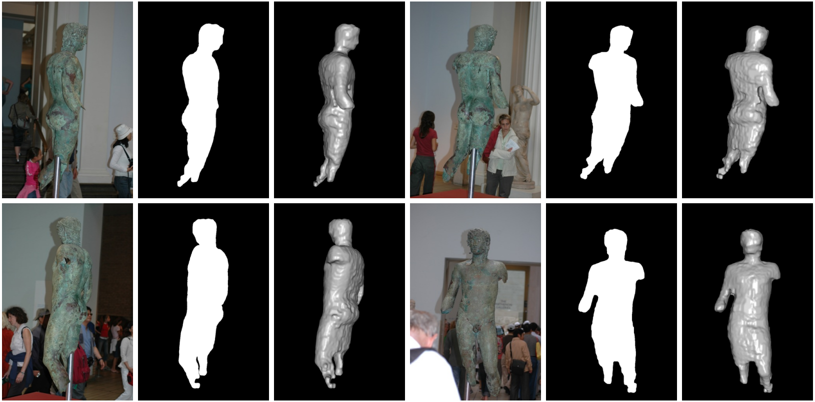
Hat Sequence Results



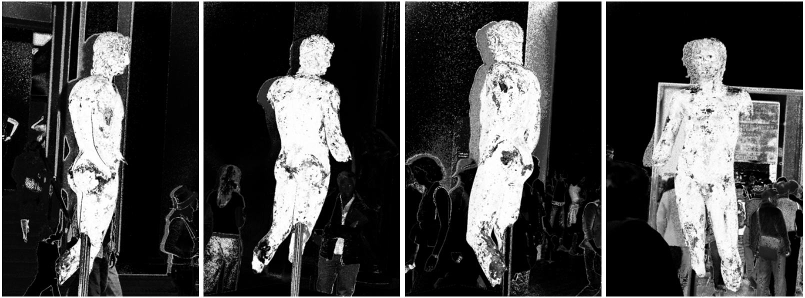
Statue Sequence: Independent 2D Segmentation Results



Statue Sequence: Our Results



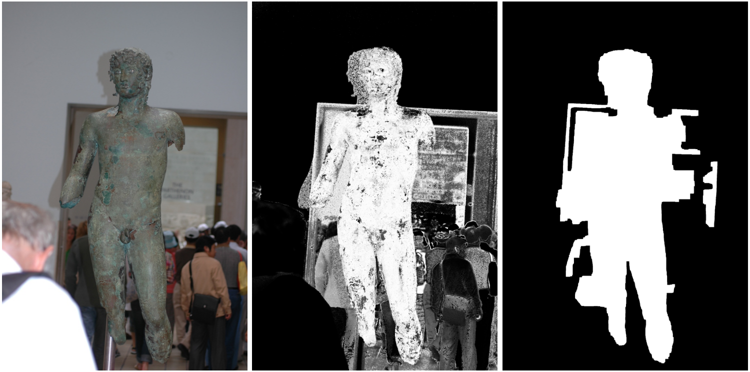
Statue Sequence: Object Likelihoods



Object Likelihoods after convergence:

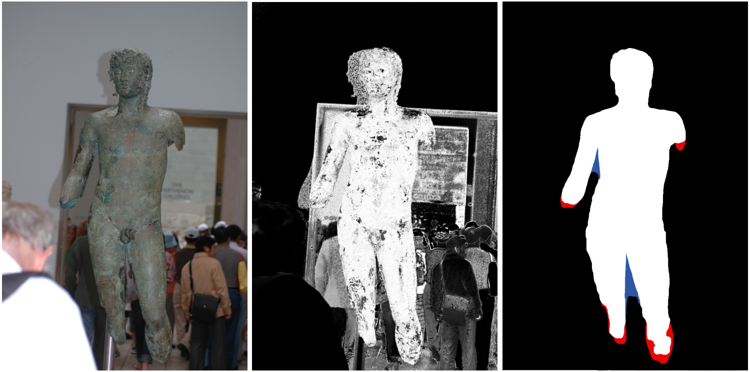
- Many views where the object is not separable in colour-space

Limitations



Best independent 2D result (even when using the converged colour model from the 3D algorithm) is very poor.

Limitations

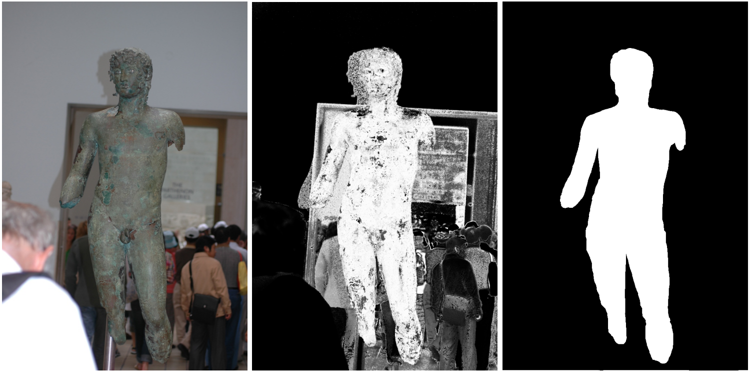


Over estimation due to regularisation and under estimation due to colour models (hollow object).

Limitations

Over estimation due to regularisation and under estimation due to colour models (hollow object)

Limitations



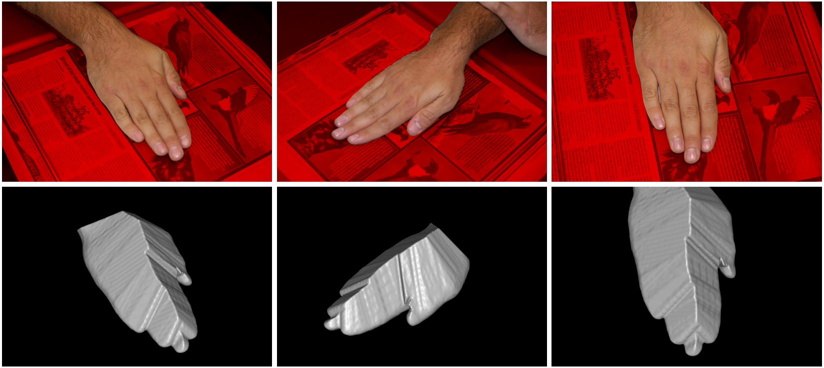
Using the result as a template to perform a boundary 2D graph-cut allows the actual silhouette to be recovered.

Hand Sequence Results



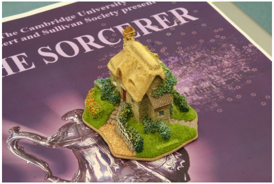
Hand Sequence Results

Hand Sequence Results



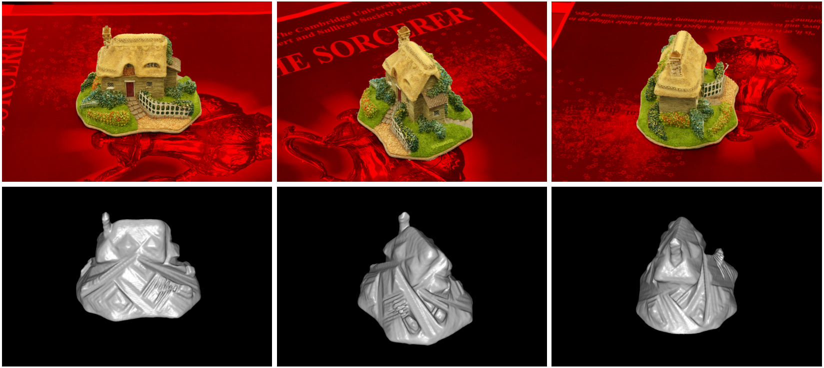
Hand Sequence Results

House Sequence Results

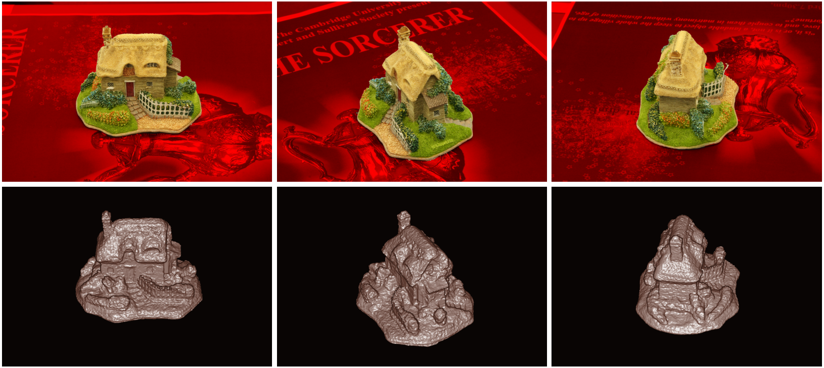


House Sequence Results

House Sequence Results



House Sequence Results



Conclusion: Further Work

There are several avenues for further work to build upon these results, most notably:

Conclusion: Further Work

There are several avenues for further work to build upon these results, most notably:

- Improve the quality and robustness of the object model

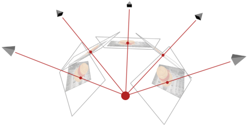
Conclusion: Further Work

There are several avenues for further work to build upon these results, most notably:

- Improve the quality and robustness of the object model
- Optimise algorithm to decrease learning times

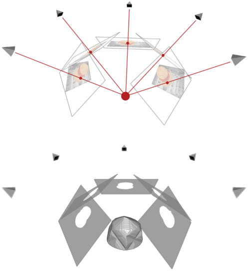
Conclusion: Summary

Conclusion: Summary



A fixation condition may be exploited for automatic initialisation

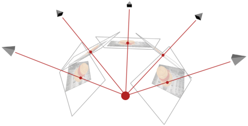
Conclusion: Summary



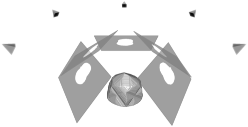
A fixation condition may be exploited for automatic initialisation

Performing segmentation across all images simultaneously in 3D enforces silhouette coherency and improves the results over independent 2D segmentations

Conclusion: Summary



A fixation condition may be exploited for automatic initialisation



Performing segmentation across all images simultaneously in 3D enforces silhouette coherency and improves the results over independent 2D segmentations

These two constraints may be used in an iterative framework to provide automatic segmentation of an object in multiple views.

Acknowledgements

This work is supported by the Schiff Foundation and Toshiba Research Europe.

Questions?

(ndfc2@cam.ac.uk)

Outline