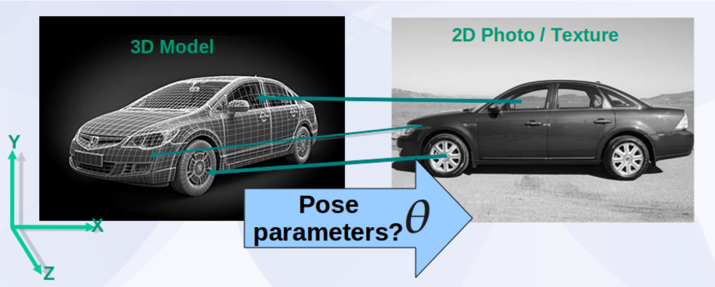
A Novel Illumination-Invariant Loss for Monocular 3D Pose Estimation

Srimal Jayawardena Marcus Hutter Nathan Brewer
Australian National University

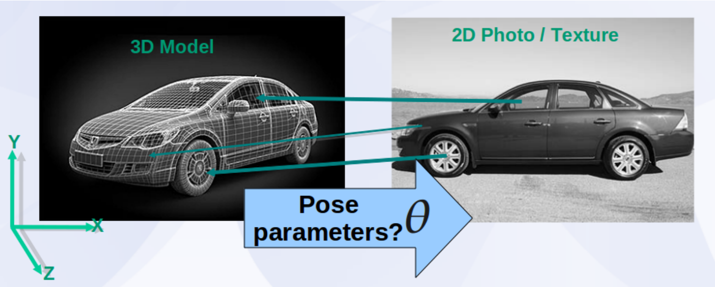
srimal(dot)jayawardena(at)anu(dot)edu(dot)au
http://users.cecs.anu.edu.au/~srimalj

DICTA 2011

The problem

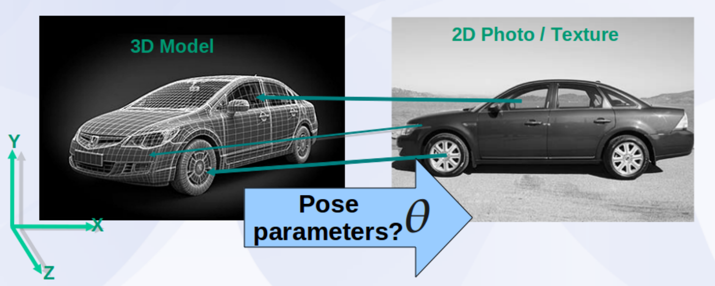


The problem



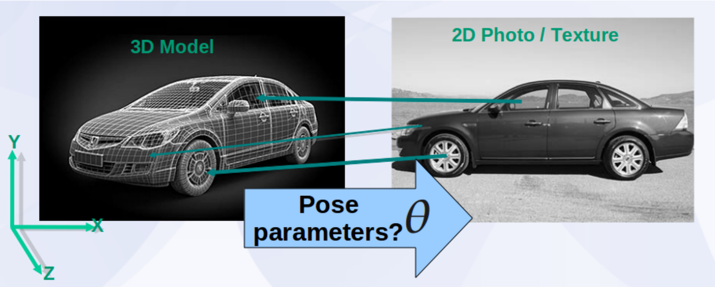
- Input: Photo of a known object and 3D CAD Model

The problem



- Input: Photo of a known object and 3D CAD Model
- Output: Pose parameters $\boldsymbol{\theta}$ that register the model on the photos

The problem



- Input: Photo of a known object and 3D CAD Model
- Output: Pose parameters $\boldsymbol{\theta}$ that register the model on the photos
- Pose - Position/orientation of 3D object w.r.t. camera

Applications

- Use as a ground truth for detailed image analysis

Applications

- Use as a ground truth for detailed image analysis
- Augmented reality applications

Applications

- Use as a ground truth for detailed image analysis
- Augmented reality applications
- Process control work

Applications

- Use as a ground truth for detailed image analysis
- Augmented reality applications
- Process control work
- CV applications needing a non-articulated full monocular 3D pose

Features of our pose estimation method

- Use only a single, static image limited to a single view

Features of our pose estimation method

- Use only a single, static image limited to a single view
- Works in an uncontrolled environment

Features of our pose estimation method

- Use only a single, static image limited to a single view
- Works in an uncontrolled environment
- Work under varying and unknown lighting conditions

Features of our pose estimation method

- Use only a single, static image limited to a single view
- Works in an uncontrolled environment
- Work under varying and unknown lighting conditions
- Avoid user interaction

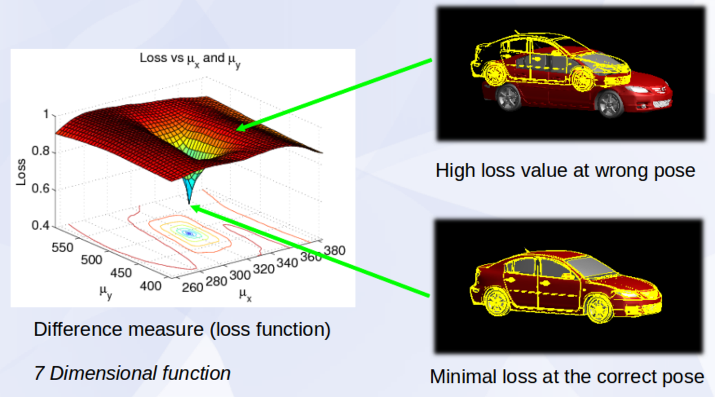
Features of our pose estimation method

- Use only a single, static image limited to a single view
- Works in an uncontrolled environment
- Work under varying and unknown lighting conditions
- Avoid user interaction
- Avoid training/learning [Arie-Nachimson and Basri, 2009, ICCV]

Features of our pose estimation method

- Use only a single, static image limited to a single view
- Works in an uncontrolled environment
- Work under varying and unknown lighting conditions
- Avoid user interaction
- Avoid training/learning [Arie-Nachimson and Basri, 2009, ICCV]
- Estimate the full 3D pose of the object (Not a set of finite Poses [Ozuysal et al., 2009, CVPR] or XY position and angle on ground plane [Sun et al., 2011, 3DIMPVT])

Approach - Minimise a loss function



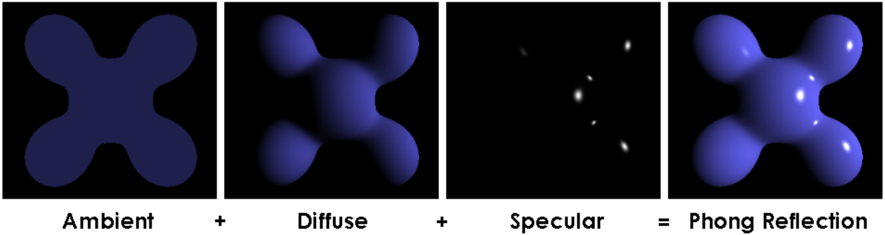
μ_x and μ_y are 2 of the 7 pose parameters estimated (explained later)

Phong reflection model

Based on the Phong reflection model [Foley, 1996]

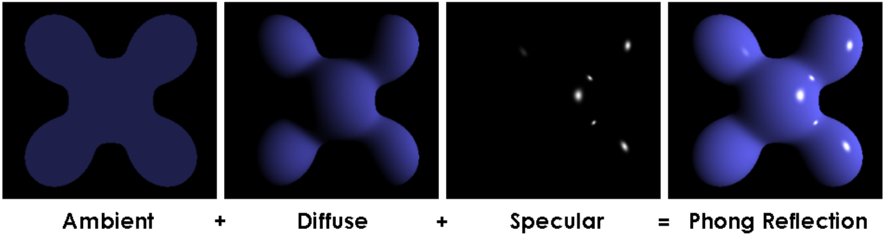
Phong reflection model

Based on the Phong reflection model [Foley, 1996]



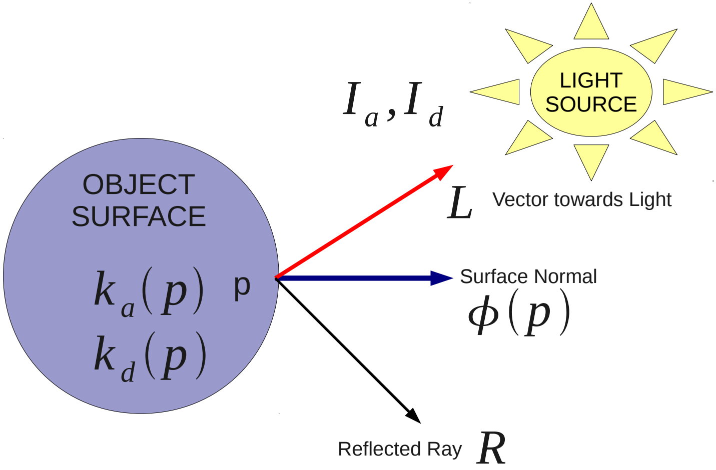
Phong reflection model

Based on the Phong reflection model [Foley, 1996]

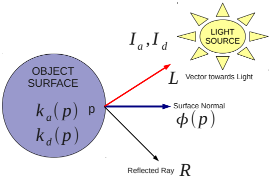


Approximation: Consider only (Ambient) + (Diffuse) terms

Phong reflection model - linear relation



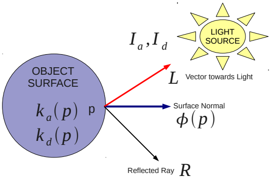
Phong reflection model - linear relation



Intensity at pixel location p (neglecting specular terms)

$$I(p) \equiv \underbrace{\begin{bmatrix} I_a & I_d \mathbf{L} \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} I_a \\ I_d \phi(\mathbf{p}) \end{bmatrix}}_{\mathbf{M}_\theta(p)} + b$$

Phong reflection model - linear relation

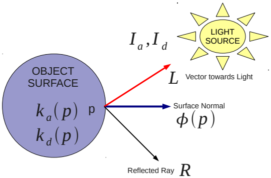


Intensity at pixel location p (neglecting specular terms)

$$I(p) \equiv \underbrace{\begin{bmatrix} I_a & I_d \mathbf{L} \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} I_a \\ I_d \phi(\mathbf{p}) \end{bmatrix}}_{\mathbf{M}_\theta(p)} + b$$

$$I(p) \equiv \mathbf{A} \cdot \mathbf{M}_\theta(p) + b \tag{1}$$

Phong reflection model - linear relation



Intensity at pixel location p (neglecting specular terms)

$$I(p) \equiv \underbrace{\begin{bmatrix} I_a & I_d\mathbf{L} \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} I_a \\ I_d\phi(\mathbf{p}) \end{bmatrix}}_{\mathbf{M}_\theta(p)} + b$$

$$I(p) \equiv \mathbf{A} \cdot \mathbf{M}_\theta(p) + b \tag{1}$$

Not realistic but sufficient for matching purposes.

Loss function

Loss at pose θ

$$L(\theta) := \mathbf{E}[||I(p) - F(p)||^2]$$

Loss function

Loss at pose θ

$$L(\theta) := \mathbf{E}[||I(p) - F(p)||^2] = \mathbf{E}[||A \cdot M_\theta(p) + b - F(p)||^2] \qquad (2)$$

Loss function

Loss at pose θ

$$L(\theta) \;:=\; \mathbf{E}[||I(p) - F(p)||^2] = \mathbf{E}[||A \cdot M_\theta(p) + b - F(p)||^2] \qquad (2)$$

At correct illumination,

$$\text{Loss}(\theta) \;:=\; \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^m} \mathbf{E}[||A \cdot M_\theta + b - F||^2] \qquad (3)$$

Loss function

Loss at pose θ

$$L(\theta) := \mathbf{E}[||I(p) - F(p)||^2] = \mathbf{E}[||A \cdot M_\theta(p) + b - F(p)||^2] \qquad (2)$$

At correct illumination,

$$\text{Loss}(\theta) := \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^m} \mathbf{E}[||A \cdot M_\theta + b - F||^2] \qquad (3)$$

As the expression is quadratic in A and b, A_{min} and b_{min} can be found analytically.

Loss function

Loss at pose θ

$$L(\theta) := \mathbf{E}[||I(p) - F(p)||^2] = \mathbf{E}[||A \cdot M_\theta(p) + b - F(p)||^2] \qquad (2)$$

At correct illumination,

$$\text{Loss}(\theta) := \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^m} \mathbf{E}[||A \cdot M_\theta + b - F||^2] \qquad (3)$$

As the expression is quadratic in A and b, A_{min} and b_{min} can be found analytically.

Details of the derivation are in the paper!

Loss Function - Illumination Invariance

$$\text{Loss}(\theta) \;:=\; \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^{m}} \mathbf{E}[||A \cdot M_\theta + b - F||^2]$$

- Invariant under regular (non-singular) linear transformation of M_θ and F

Loss Function - Illumination Invariance

$$\text{Loss}(\theta) \ := \ \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^m} \mathbf{E}[||A \cdot M_\theta + b - F||^2]$$

- Invariant under regular (non-singular) linear transformation of M_θ and F
- $\text{Loss}(\theta)$ is the same for any $M_\theta \leftarrow A' M_\theta + b'$ for all b' and all non-singular A'

Loss Function - Illumination Invariance

$$\text{Loss}(\theta) := \min_{A \in \boldsymbol{R}^{m \times n}} \min_{b \in \boldsymbol{R}^m} \mathbf{E}[||A \cdot M_\theta + b - F||^2]$$

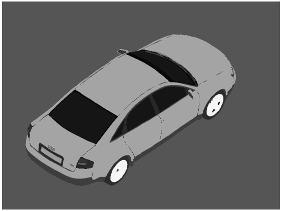
- Invariant under regular (non-singular) linear transformation of M_θ and F
- $\text{Loss}(\theta)$ is the same for any $M_\theta \leftarrow A' M_\theta + b'$ for all b' and all non-singular A'
- Simillarly for linear transformations of F

Loss Function - Illumination Invariance

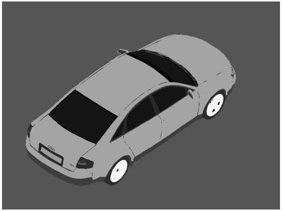
$$\text{Loss}(\theta) \;:=\; \min_{A\in \boldsymbol{R}^{m\times n}} \min_{b\in \boldsymbol{R}^{m}} \mathbf{E}[||A\cdot M_\theta + b - F||^2]$$

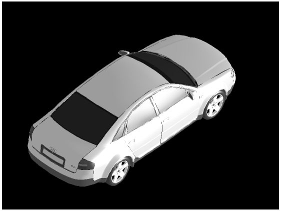
- Invariant under regular (non-singular) linear transformation of M_θ and F
- Loss(θ) is the same for any $M_\theta \leftarrow A' M_\theta + b'$ for all b' and all non-singular A'
- Simillarly for linear transformations of F
- Independent of lighting A

Loss Function - Illumination Invariance

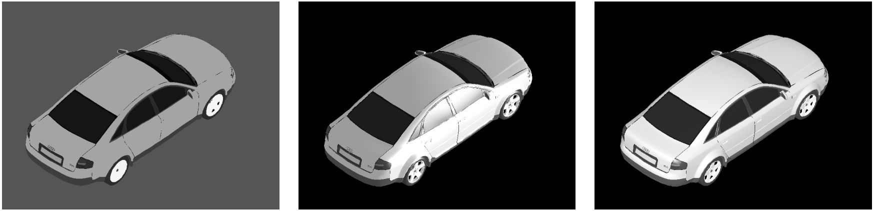


Loss Function - Illumination Invariance





Loss Function - Illumination Invariance



Pose representation

Orthographic projection (6 d.f)

- Rotation (3)

Pose representation

Orthographic projection (6 d.f)

- Rotation (3)
- Shift (2)

Pose representation

Orthographic projection (6 d.f)

- Rotation (3)
- Shift (2)
- Scale (1)

Pose representation

Orthographic projection (6 d.f)

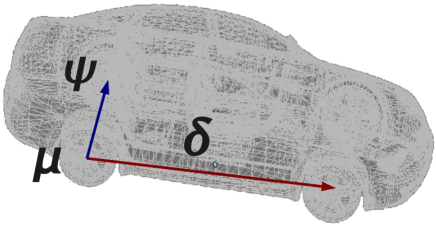
- Rotation (3)
- Shift (2)
- Scale (1)

Pose representation

Orthographic projection (6 d.f)

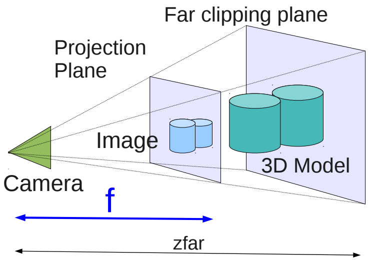
- Rotation (3)
- Shift (2)
- Scale (1)

For vechilce pose:



Pose representation

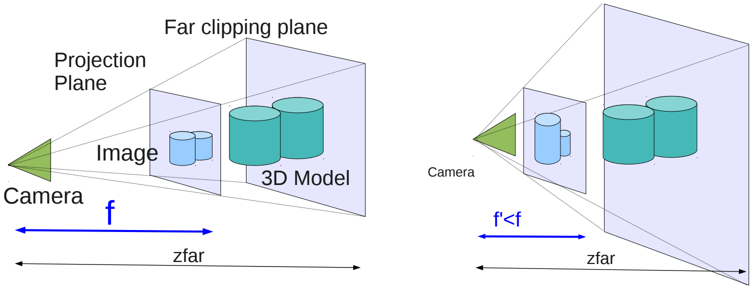
Perspective projection (7 d.f)



(a) Large f

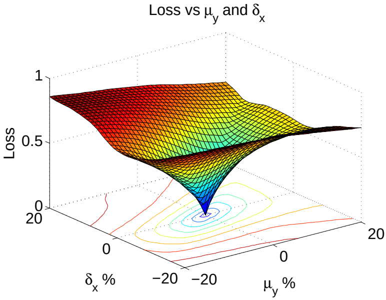
Pose representation

Perspective projection (7 d.f)



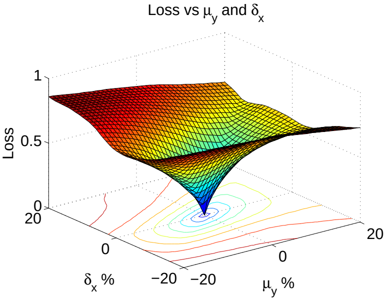
(a) Large f (b) Small f

Loss landscapes

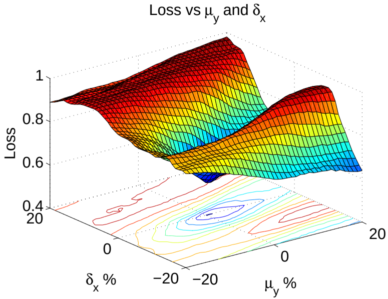


(a) Synthetic photo

Loss landscapes



(a) Synthetic photo



(b) Real photo

Initial rough pose to initialise the optimiser

- Several ways to obtain an initial (rough) pose: [Arie-Nachimson and Basri, 2009, ICCV] [Ozuysal et al., 2009, CVPR] [Sun et al., 2011, 3DIMPVT]

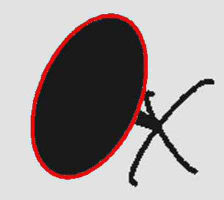
Initial rough pose to initialise the optimiser

- Several ways to obtain an initial (rough) pose: [Arie-Nachimson and Basri, 2009, ICCV] [Ozuysal et al., 2009, CVPR] [Sun et al., 2011, 3DIMPVT]
- We use: *Wheel match method [Hutter and Brewer, 2009, IVCNZ]*

Initial rough pose to initialise the optimiser

- Several ways to obtain an initial (rough) pose: [Arie-Nachimson and Basri, 2009, ICCV] [Ozuysal et al., 2009, CVPR] [Sun et al., 2011, 3DIMPVT]
- We use: *Wheel match method [Hutter and Brewer, 2009, IVCNZ]*

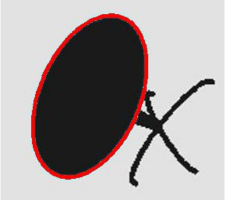
Motivation:



Initial rough pose to initialise the optimiser

- Several ways to obtain an initial (rough) pose: [Arie-Nachimson and Basri, 2009, ICCV] [Ozuysal et al., 2009, CVPR] [Sun et al., 2011, 3DIMPVT]
- We use: *Wheel match method [Hutter and Brewer, 2009, IVCNZ]*

Motivation:



(a) (b) (c) (d) (e) (f) (g)

The optimiser

- Downhill Simplex Method [Nelder and Mead, 1965]

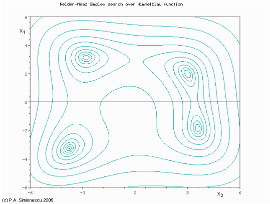
The optimiser

- Downhill Simplex Method [Nelder and Mead, 1965]
- Direct Search Method - Derivative information not required

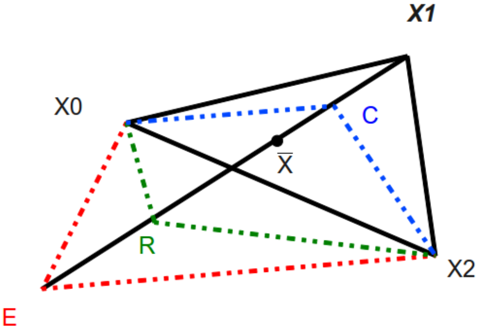
The optimiser

- Downhill Simplex Method [Nelder and Mead, 1965]
- Direct Search Method - Derivative information not required

A 2D example:



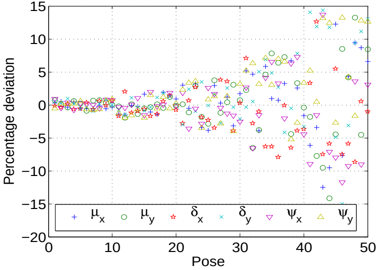
(a) Rosenbrock (2D)



(b) The simplex (3 points)

Reliability tests on loss based pose estimation

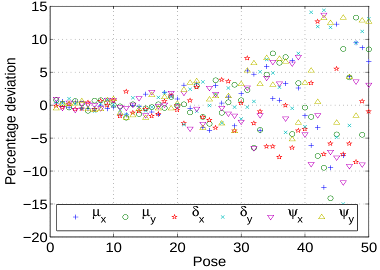
Reliability tests of pose estimation (initial rough pose with increasing deviations)



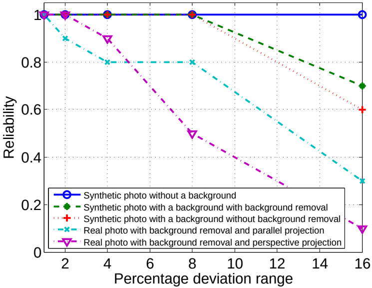
(a) Initial rough pose deviations

Reliability tests on loss based pose estimation

Reliability tests of pose estimation (initial rough pose with increasing deviations)



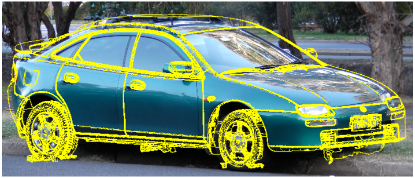
(a) Initial rough pose deviations



(b) Reliability $= \frac{NoCorrectCases}{TotalTestsPerDevnRange}$

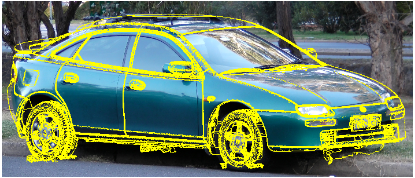
Background removal using GrabCut [Rother et al., 2004, ACM]

Results - Scanned 3D CAD (Mazda Astina)

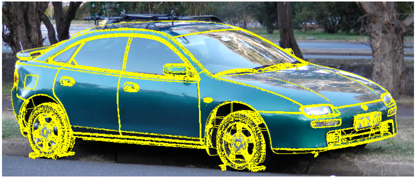


(a) Initial rough pose

Results - Scanned 3D CAD (Mazda Astina)

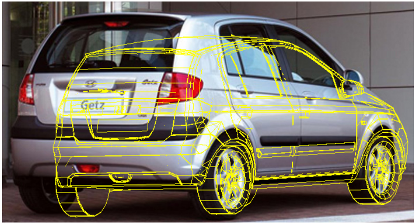


(a) Initial rough pose

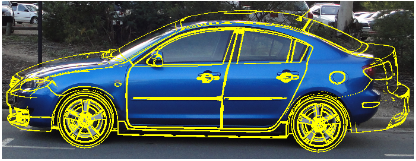


(b) Final pose

Results - Internet 3D CAD models

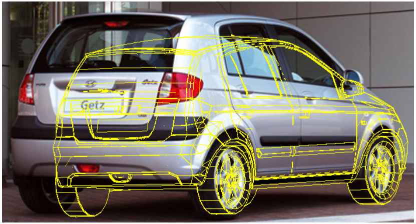


(a) Initial

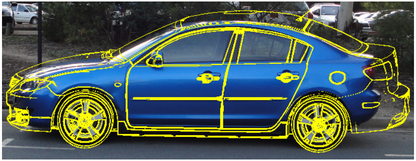


(b) Initial

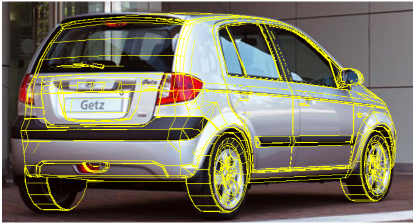
Results - Internet 3D CAD models



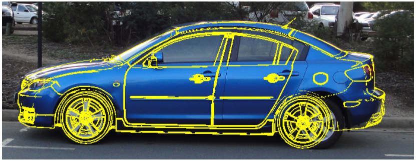
(a) Initial



(b) Initial



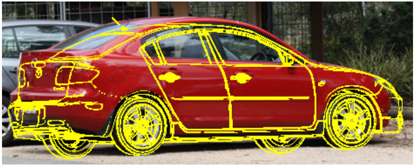
(c) Final



(d) Final

Results - Internet 3D CAD models

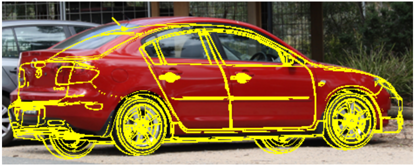
(a) Initial



(b) Initial

Results - Internet 3D CAD models

(a) Initial



(b) Initial

(c) Final

(d) Final

Computation times

Table: Rendering and loss calculation times.

Approach	**Loss calc.**	**Render**
MATLAB	0.16 s	2.28 s
C/OpenGL	0.04 s	0.17 s

Approx 2 minutes to optimise 800x600 image

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos
- Downhill-simplex optimiser is effective with simplex re-initialisations

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos
- Downhill-simplex optimiser is effective with simplex re-initialisations

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos
- Downhill-simplex optimiser is effective with simplex re-initialisations

Outlook:

- A planned application - automatic damage detection in vehicles

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos
- Downhill-simplex optimiser is effective with simplex re-initialisations

Outlook:

- A planned application - automatic damage detection in vehicles

Conclusion and outlook

Conclusion:

- The loss function works successfully on real photos
- Downhill-simplex optimiser is effective with simplex re-initialisations

Outlook:

- A planned application - automatic damage detection in vehicles

Thank you!

References I

Arie-Nachimson, M. and Basri, R. (2009).
Constructing implicit 3d shape models for pose estimation.
In *ICCV*.

Foley, J. (1996).
Computer graphics: principles and practice.
Addison-Wesley Professional.

Hutter, M. and Brewer, N. (2009).
Matching 2-D Ellipses to 3-D Circles with Application to Vehicle Pose Identification.
In *Image and Vision Computing New Zealand, 2009. IVCNZ'09. 24th International Conference*, pages 153–158.

References II

Nelder, J. and Mead, R. (1965).
A simplex method for function minimization.
The computer journal, 7(4):308.

Ozuysal, M., Lepetit, V., and P.Fua (2009).
Pose estimation for category specific multiview object localization.
In *Conference on Computer Vision and Pattern Recognition*, Miami, FL.

Rother, C., Kolmogorov, V., and Blake, A. (2004).
Grabcut: Interactive foreground extraction using iterated graph cuts.
ACM Transactions on Graphics (TOG), 23(3):309–314.

Sun, M., Kumar, S. S., Bradski, G., and Savarese, S. (2011).
Toward automatic 3d generic object modeling from one single image.
In *3DIMPVT*, Hangzhou, China.