Image Based Rendering using Image-based Priors

Andrew Fitzgibbon, Yonatan Wexler,
Andrew Zisserman

Presented by Gary Gettig.

Outline

- Summarize the method
- Photoconsistency
- Texture Priors
- Results
- Implementation

New Image Synthesis

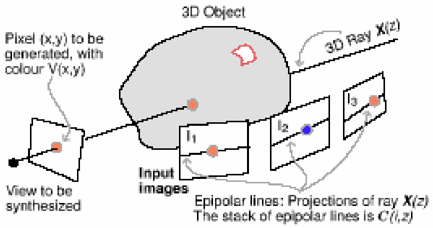


Figure 2: **Geometric configuration**. The supplied information is a set of 2D images $I_{1..n}$ and their camera positions $P_{1..n}$. At each pixel in the view to be synthesized, we wish to discover the colour which is most likely to be a reprojection of a 3D object point, based on the implied projection into the source images.

Infer most likely view V that maximizes the posterior:

$$p(V \mid I_1, ... I_n)$$

Based on color derived from depth

Bayesian Framework

- Bayes Rule

$$p(V \mid I_1 ... I_n) = \frac{p(I_1 ... I_n \mid V)\, p(V)}{p(I_1 ... I_n)}$$

Ignore

- Maximize

$$q(V) = p(I_1 ... I_n \mid V)\, p(V)$$

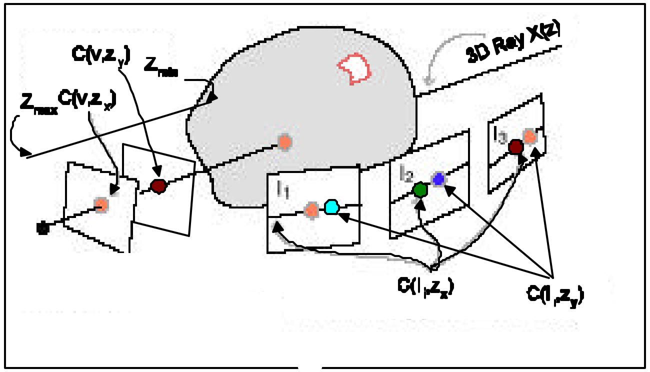
Texture Prior

Photoconsistency likelihood

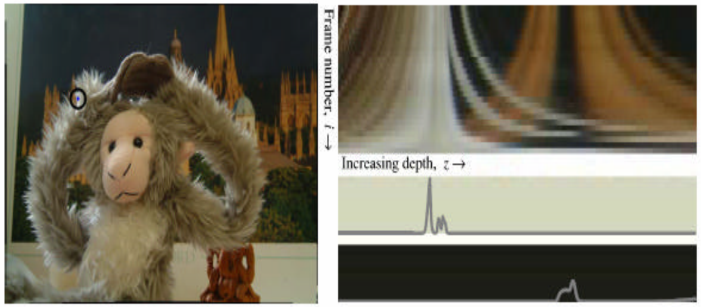
Photoconsistency Constraint

- $p(I_1 \ldots I_n \mid V)$
- Assumption is that objects are diffuse and opaque
- Any deviation from assumption is noise

Photoconsistency



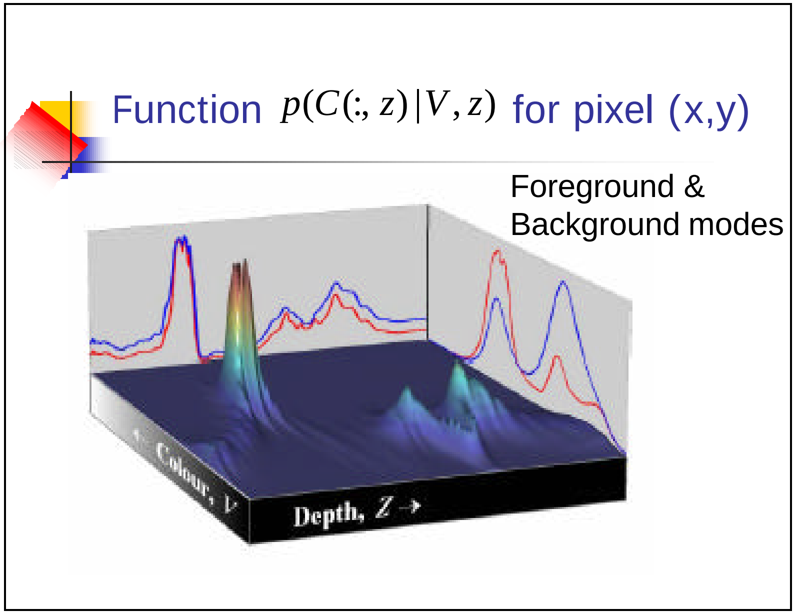
Photoconsistency at point (x,y)



Formulation of $p(I_1 \ldots I_n \mid V)$

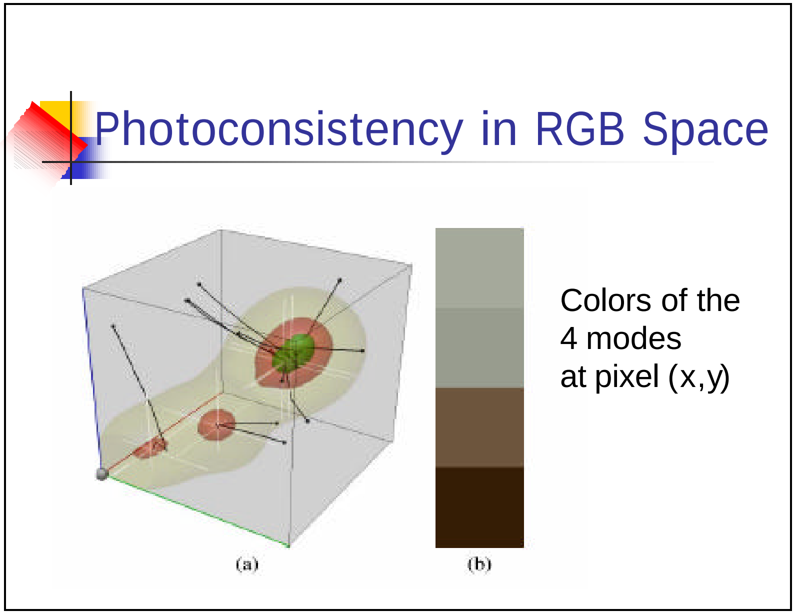
- = p(C | V) where $C=\{C(i,z) \mid 1 \le i \le n, z_{\min} < z < z_{\max}\}|$
- = $\int p(C(:,z) \mid V,z)dz$
- = $\prod_{i=1}^{n} \exp(-br(\|V - C(i,z)\|)$ where $r(x) = |x|, x^2$
- $\approx \max_{z} p(C(:,z) \mid V,z)$ Approx. marginal by the maximum
- Take the - log or Energy Deviation at pixel (x,y)

$$E_{photo}(V) = \min_{z_{\min}<z<z_{\max}} \sum_{i=1}^{n} r(\|V - C(i,z)\|))$$



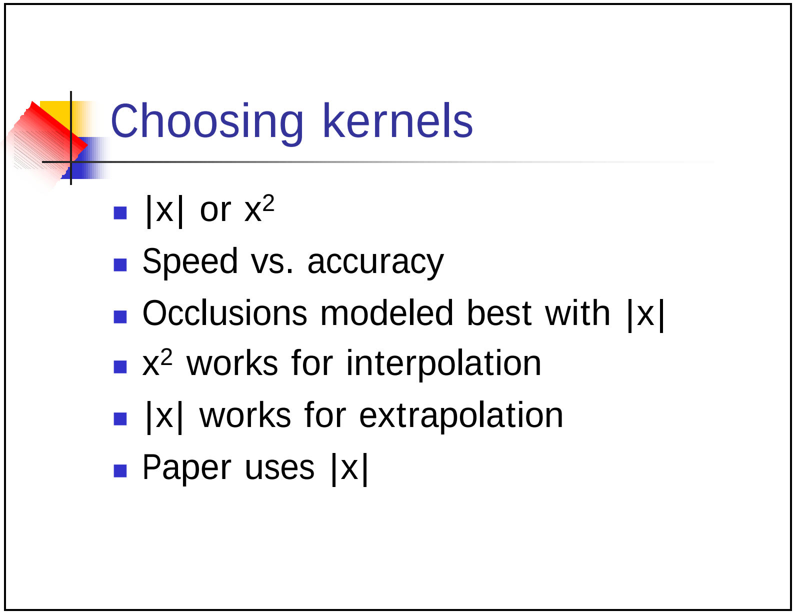
Function $p(C(:, z)|V, z)$ for pixel (x,y)

Foreground &
Background modes



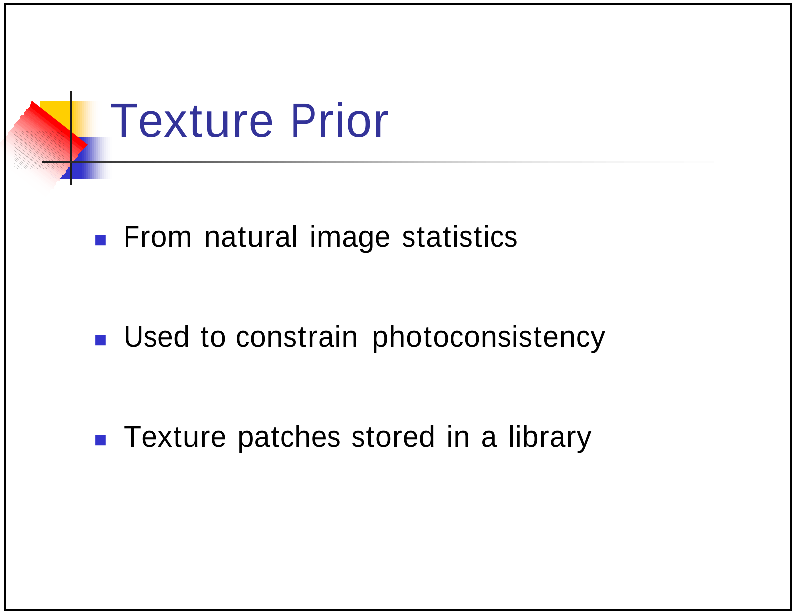
Photoconsistency in RGB Space

Colors of the
4 modes
at pixel (x,y)



Choosing kernels

- $|x|$ or x^2
- Speed vs. accuracy
- Occlusions modeled best with $|x|$
- x^2 works for interpolation
- $|x|$ works for extrapolation
- Paper uses $|x|$



Texture Prior

- From natural image statistics
- Used to constrain photoconsistency
- Texture patches stored in a library

Texture Prior $p_{tx}(V)$

$$p_{tx}(V) = \prod_{x,y} p_{tx}(N(x,y))$$

$$N(x,y) = \{V(x+i, y+j) \mid -2 \le i, j \le 2\}$$ 5x5 Texture Patch

$$p_{tx}(N(x,y)) = \exp\left(-l \min_{T \in \mathrm{T}} \left\| T - N(x,y) \right\|^2\right)$$

Taking the –log or Energy

$$E_{tx}(N(x,y)) = -l \min_{T \in \mathrm{T}} \left\| T - N(x,y) \right\|^2$$

Prior likelihood of texture patch surrounding (x,y)

Need for Texture Priors

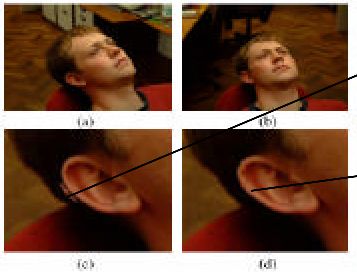


Figure 1: **View synthesis.** (a,b): Two from a set of 39 images taken by a hand-held camera. (c): Detail from a new view generated using state-of-the-art view synthesis. The new view is about 20° displaced from the closest view in the original sequence. Note the spurious echo of the ear. (d): The same detail, but constrained to only generate views which have similar local statistics to the input images.

Without priors

With priors

Leave-one-out Example

(a) (b) (c) (d)

Figure 7: **Leave-one-out test.** Using 26 views to render a missing view allows comparison to be made between the rendered view and ground truth. (a) Maximum-likelihood view, in which each pixel is coloured according to the highest mode of the photoconsistency function. High-frequency artifacts are visible throughout the scene. (b) View synthesized using texture prior. The artifacts are significantly reduced. (c) Ground-truth view. (d) Difference image between (b) and (c).

Leave-one-out sequence

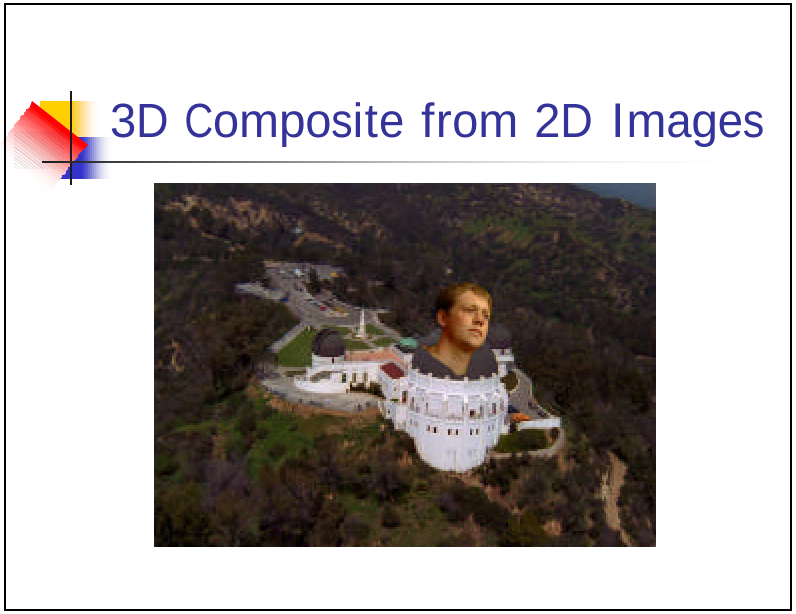
- 27 frames
- Each frame is derived from the other 26
- Artifacts occur under monkeys arm

Steadicam Example

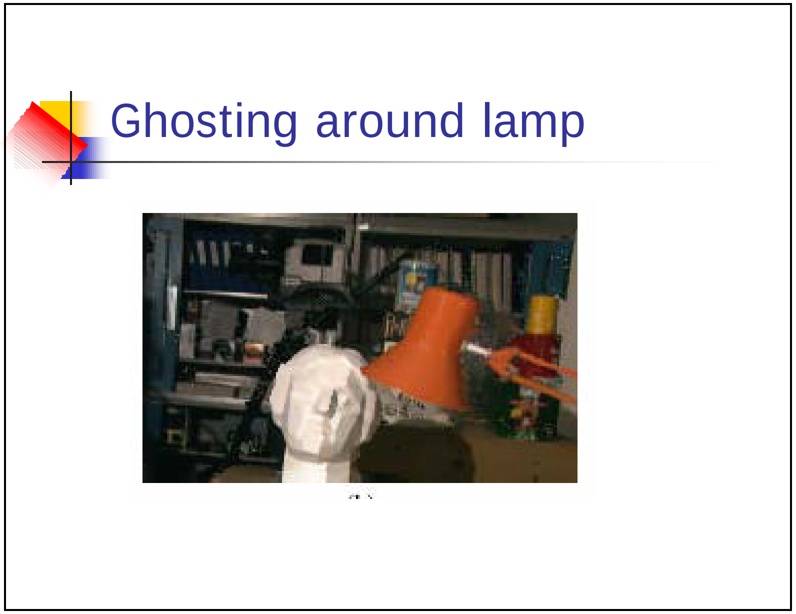
Figure 8: **Steadicam test**. Three novel views of the monkey scene from viewpoints not in the original sequence. The complete sequence may be found at `http://www.robots.ox.ac.uk/~awf/ibr`.

Steadicam Sequence

- Scene rendered from an interpolation of the first and last frame



3D Composite from 2D Images



Ghosting around lamp

Implementation

- Compute set of modes (colors) for $E_{photo}(V)$ and limit solution to this set
- Use $E_{tx}(V)$ to select solution from this set

Enumerating $E_{photo}(V)$

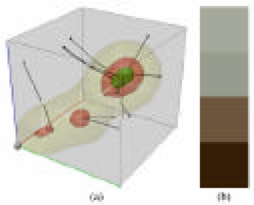


Figure 5: **Minima of E_{photo}.** (a) Isosurfaces in RGB space of the photoconsistency function $E_{photo}(V)$ at the pixel studied in figure 3. Minima are computed by gradient descent from random starting positions, of which twelve are shown (black circles), with the gradient descent trajectories plotted in black. Four modes were retained after clustering; their locations are marked by white 3D "axes" lines in (a), and their RGB colours are shown in (b).

- Gradient Descent
- 12 iterations
- 20 random starting points
- Finds 4-6 modes

Algorithm

- Preprocessing - For each pixel
  - Compute d(x,y) – *from virtual camera calibration*
  - Compute Color matrix - C_{ij}
  - Compute E_{photo} minima {K}
  - Sort $E_{photo}(V_k)$
  - Set initial estimate of new view: $V^0(x,y) = V_1(x,y)$
- Iterate over (r) – For each pixel
  - Extract N – texture window
  - Find best matching texture patch
  - Calculate new V^r
  - Set $V^r(x,y)$ to the closest {K}

$$V^r = \frac{V^{r-1} + l T}{1 + l}$$

Conclusion

- Importance of Texture Priors
- Good algorithm with excellent insight into the implementation
- Occlusions and object delineations require more image samples
  - Or large texture library

Algorithm Overview

- Each pixel has a pre-computed $E_{photo}(V)$
- Iterated conditional modes (ICM) algorithm
  - Alternates minimizing $E_{photo}(V)$ and $E_{tx}(N)$
- Minimize $<E_{photo}(V) + E_{tx}(V)>$ for each pixel over a 5x5 window by approximation:

$$E_{photo} \approx \left\| V - V^{r-1} \right\|^2$$

Where V^{r-1} = color at previous iteration

Let $$V^r = \frac{V^{r-1} + l\,T}{1 + l}$$

Where T is the color of the center pixel of the best $E_{tx}(V)$

From Probability to Energy

$$q(V) = \prod_{x,y} p_{photo}(V(x,y))\, p_{tx}(N(x,y)$$

Find the max

Find the min

$$E(V) = \sum_{x,y} E_{photo}(V(x,y)) + \sum_{x,y} E_{tx}(N(x,y))$$

Prior likelihood of texture
Patch surroundin (x,y)

Deviation from photoconsistency
at pixel (x,y)

What is the best way to solve this equation?