Applications

Presented by
Sherin Aly

Most slides Courtesy to the papers’ authors
Photo Tourism: Exploring Photo Collections in 3D

Noah Snavely
Steven M. Seitz
University of Washington

Richard Szeliski
Microsoft Research
The goal of Photo Tourism is to use location information to build a better visualization and photo exploration tool for collections of photos of the same scene.
Photo Tourism

• A demo
• The system takes the set of photos and automatically determines the relative positions and orientations from which each photo was taken.
• We can then load the photos into the immersive 3D browser where the user can visualize and explore the photos using spatial relationships.
Photo Tourism overview

Input photographs
- Internet
- Personal collection

Scene geometry reconstruction
- Relative camera positions and orientations
- Point cloud for scene geometry
- Sparse correspondence

Photo Explorer
Related work

- Image-based modeling (recover scene geometry)

Debevec, et al.  
SIGGRAPH 1996

Schaffalitzky and Zisserman  
ECCV 2002

Brown and Lowe  
3DIM 2005

automatic structure from motion on unordered sets of images

However, this modeling system is the first to be successfully applied to hundreds of images taken from the Internet.
Related work

- Image-based rendering (depict transitions between images)

Photorealistic IBR:
Levoy and Hanrahan, SIGGRAPH 1996
Gortler, et al, SIGGRAPH 1996
Seitz and Dyer, SIGGRAPH 1996
Aliaga, et al, SIGGRAPH 2001
and many others

Aspen Movie Map
Lippman, et al., 1978

They created an interactive virtual tour of the city of Aspen,

Same exploration features, but which required extensive manual effort to create.
Related work

• Image browsing and image retrieval

WWMX
Toyama, *et al*.
Int. Conf. Multimedia, 2003
Use location information to organize photos,
No immersive browsing experience

The Realityflythrough project
McCurdy and Griswold
Mobisys 2003
Use GPS to locate images in space
PhotoTourism system
don’t use GPS

Video Google
Sivic and Zisserman
ICCV 2003
It allow users to find objects in videos by selecting them.
PT:“3D photo browser” context
Photo Tourism overview

Input photographs → Scene reconstruction → Photo Explorer
Scene reconstruction

- Automatically estimate
  - position, orientation, and focal length of cameras (i.e. zoom)
  - 3D positions of feature points
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

invariant to
-scale
-Rotation
- affine changes in image intensity.

squares are scaled and rotated to reflect the scale and orientation of the features.

The Trevi Fountain
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]
Feature matching

Match features between each pair of images
Feature matching

Refine matching using RANSAC [Fischler & Bolles 1987] to estimate fundamental matrices between pairs and then keep only matches consistent with that fundamental matrix.
Structure from motion

\[
\text{minimize } f(R, T, P)
\]

minimize sum of squared reprojection error between the projected and observed 2D points

\[
\minimize f(R, T, P)
\]

between the projected and observed 2D points
Structure from motion (Cont.)

• This is a non-linear least squares problem and can be solved with algorithms such as Levenberg-Marquart.

• However, because the problem is non-linear, it can be sensitive to local minima.

• Therefore, it’s important to initialize the parameters of the system carefully.

• In addition, we need to be able to deal with erroneous correspondences.
Incremental structure from motion

reconstruct the scene incrementally
Incremental structure from motion
Incremental structure from motion
Demo 1
repeat until no more photos match any points in the scene
Reconstruction performance

- For photo sets from the Internet, 20% to 75% of the photos were registered.
- Most unregistered photos belonged to different connected components (e.g. when searching for the Notre Dame cathedral, you get back photos of both the interior and exterior).
- Some failure cases (noisy, dark, too low resolution, too different angel than others).
- Running time: < 1 hour for 80 photos
  > 1 week for 2600 photo
Photo Tourism overview

Input photographs → Scene reconstruction → Photo Explorer
Photo Tourism overview

Input photographs → Scene reconstruction → Photo Explorer

- Navigation
- Rendering
- Annotations
Navigation controls

- Free-flight navigation
- Object-based browsing
- Relation-based browsing
- Overhead map
Object-based browsing
Object-based browsing

- Visibility
- Resolution
- Head-on view or oblique
Relation-based browsing

Find all details

Find all similar images

Find all zoom outs

Zoom in

Move left

Move right

Zoom out
Relation-based browsing

These relations are inferred based on the relative positions of corresponding feature points between photographs.
Relation-based browsing
Relation-based browsing

Image A

to the right of

Image B

Image C
Relation-based browsing

Image A

Image C

to the right of

Image B
Relation-based browsing

Image A

to the right of

is detail of

Image B

Image C

Image D
Relation-based browsing

Image A

Image B

Image C

Image D
Relation-based browsing

Image A

is zoom-out of

is detail of

to the left of

to the right of

Image C

is zoom-out of

is detail of

Image D

is zoom-out of

is detail of

Image B

is zoom-out of

is detail of
Photo Tourism overview

Input photographs → Scene reconstruction → Photo Explorer

- Navigation
- Rendering
- Annotations
Rendering(Scene)

3D line segments extracted automatically from the photo collection and these washed-out looking colors
Rendering
Rendering transitions (photographs)
Image-based technique
Photo Tourism overview

Input photographs \rightarrow \text{Scene reconstruction} \rightarrow \text{Photo Explorer}

- Navigation
- Rendering
- Annotations
Annotations

Notre Dame de Paris

Reproduced with permission of Yahoo! Inc. © 2005 by Yahoo! Inc.
YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.
Contributions

- Automated system for registering photo collections in 3D for interactive exploration
- Structure from motion algorithm demonstrated on hundreds of photos from the Internet
- Photo exploration system combining new image-based rendering and photo navigation techniques
Limitations / Future work

- Not all photos can be reliably matched

- Structure from motion scalability
  - More efficient (sparse) algorithms

- Plane-based transitions lack parallax
Conclusion

Indexing everyone’s photos provides a new way to share and experience our world

To find out more:

– http://phototour.cs.washington.edu
– http://research.microsoft.com/IVM/PhotoTourism
– http://labs.live.com/photosynth
– Exhibition booth #2619

Saint Basil's Cathedral  Trafalgar Square  Rockefeller Center  Mount Rushmore
FaceTracer: A Search Engine for Large Collections of Images with Faces

Neeraj Kumar, Peter Belhumeur, Shree Nayar

Columbia University
How Can We Describe This Face?

Woman
Young
Asian
Brunette
Smiling
...
How Can We Describe The Image?

Indoors
Flash
In Focus

Frontal
Alone

…
We Need a Search Engine Based on Facial and Image Appearance
Some Numbers

• Billions of Images
• Hundreds of Attributes
• Thousands of Manual Labels

We need to do this automatically
• OKAO face detector
  • Pose angles
  • 6 Facial keypoints location
• Alignment by linear least squares on detected facial points and corresponding points on a template face.
Detect and Align
1. Face Regions
# Feature Types

<table>
<thead>
<tr>
<th>2. Pixel Value Type</th>
<th>3. Normalizations</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB ((r))</td>
<td>None ((n))</td>
<td>None ((n))</td>
</tr>
<tr>
<td>HSV ((h))</td>
<td>Mean-Norm ((m))</td>
<td>Histogram ((h))</td>
</tr>
<tr>
<td>Image Intensity ((i))</td>
<td>Energy-Norm ((e))</td>
<td>Statistics ((s))</td>
</tr>
<tr>
<td>Edge Magnitude ((m))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge Orientation ((o))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Feature Types

<table>
<thead>
<tr>
<th>Pixel Value Type</th>
<th>Normalizations</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB (r)</td>
<td>None (n)</td>
<td>None (n)</td>
</tr>
<tr>
<td>HSV (h)</td>
<td>Mean-Norm (m)</td>
<td>Histogram (h)</td>
</tr>
<tr>
<td>Image Intensity (i)</td>
<td>Energy-Norm (e)</td>
<td>Statistics (s)</td>
</tr>
<tr>
<td>Edge Magnitude (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge Orientation (o)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Region:pixel type.normalization.aggregation."

RGB, Mean Norm., No Aggreg. (r.m.n)
## Feature Types

<table>
<thead>
<tr>
<th>Pixel Value Type</th>
<th>Normalizations</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB (r)</td>
<td>None (n)</td>
<td>None (n)</td>
</tr>
<tr>
<td>HSV (h)</td>
<td>Mean-Norm (m)</td>
<td>Histogram (h)</td>
</tr>
<tr>
<td>Image Intensity (i)</td>
<td>Energy-Norm (e)</td>
<td>Statistics (s)</td>
</tr>
<tr>
<td>Edge Magnitude (m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge Orientation (o)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Region:pixel type.normalization.aggregation.”

Edge Orientations, No Norm, Histogram (o.n.h)
Train Classifiers

Pool of Classifiers - one per region/feature type

Mouth
Raw RGB
Train Classifiers

Eyes
Mean-Normalized RGB

Pool of Classifiers- one per region/feature type
Train Classifiers

Pool of Classifiers - one per region/feature type

Whole Face
Raw Intensity
Train Classifiers

Pool of Classifiers - one per region/feature type

Whole Face
Gradient Directions
Select Classifiers

Pool of Classifiers

Selected Classifiers

Error Rate

Iteration

© 2006 Noah Snavely
Feature Selection: Smiling

1. Mouth: RGB, Mean Norm., No Aggreg. (M:r.m.n)

2. Mouth: RGB, No Norm., No Aggreg. (M:r.n.n)

3. Mouth: RGB, Energy Norm., No Aggreg. (M:r.e.n)

4. Whole Face: Intensity, No Norm., No Aggreg. (W:i.n.n)

5. ...
Selected Features

Smiling
Selected Features

Gender
Selected Features

Indoor/Outdoor
Selected Features

Hair Color
# Classification Accuracy

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>8.62%</td>
</tr>
<tr>
<td>Age</td>
<td>16.65%</td>
</tr>
<tr>
<td>Race</td>
<td>6.49%</td>
</tr>
<tr>
<td>Hair Color</td>
<td>5.54%</td>
</tr>
<tr>
<td>Eye Wear</td>
<td>5.14%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mustache</td>
<td>4.61%</td>
</tr>
<tr>
<td>Smiling</td>
<td>4.60%</td>
</tr>
<tr>
<td>Blurry</td>
<td>3.41%</td>
</tr>
<tr>
<td>Lighting</td>
<td>1.61%</td>
</tr>
<tr>
<td>Environment</td>
<td>12.15%</td>
</tr>
</tbody>
</table>

© 2006 Noah Snavely
## Comparison to State-of-the-Art

<table>
<thead>
<tr>
<th>Method</th>
<th>Gender Error Rate</th>
<th>Smiling Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>8.62%</td>
<td>4.60%</td>
</tr>
<tr>
<td>Baluja &amp; Rowley, IJCV 2007</td>
<td>13.13%</td>
<td>7.41%</td>
</tr>
<tr>
<td>Shakhnarovich et al., ICAFGR 2002</td>
<td>12.88%</td>
<td>6.40%</td>
</tr>
<tr>
<td>Moghaddam &amp; Yang, TPAMI 2002</td>
<td>9.52%</td>
<td>13.54%</td>
</tr>
</tbody>
</table>
Results
“Asian Babies”
“Adults Outside”
“Middle-Aged White Men”
“Old Men With Mustaches”
"People Wearing Sunglasses Outside"
“Kids Indoors Not Smiling”
“Men With Dark Hair”
Personal FaceTracer Search

"Children outside"
A Computer Vision System for Automatic Plant Species Identification

Neeraj Kumar  University of Washington
Peter N. Belhumeur  Columbia University
Arijit Biswas  University of Maryland
David W. Jacobs  University of Maryland
W. John Kress  Smithsonian Institution
Ida C. Lopez  Smithsonian Institution
João V.B. Soares  University of Maryland

© 2006 Noah Snavely
What Plant Species is this?
Let's Use a Field Guide

What Tree Is That?
A guide to the more common trees found in North America

© 2006 Noah Snavely
Like a normal field guide...
Like a normal field guide…

- that you can search and sort
Like a normal field guide…
- that you can search and sort
- and with visual recognition
Large Intra-Class Variation

Images of Paper Mulberry (*Broussonettia papyrifera*)
What do iPhone Images Look Like?

Blurry, Not flat
What do iPhone Images Look Like?

Blurry, Not flat, Varying color, Shadows
What do iPhone Images Look Like?

Blurry, Not flat, Varying color, Shadows, No venation, Thin Structures
What do iPhone Images Look Like?

Blurry, Not flat, Varying color, Shadows, No venation, Thin Structures, White balance, Color splotches, …
Classification

• Classifying whether the image is of a valid leaf
  – Of a single leaf
  – Placed on light
  – Un-textured background with no other clutter

• They employ a binary RBF SVM classifier applied to gist features (by LEAR).
Segmentation

Leaf Shape is Distinctive

- 5-lobed
- Serrated edges
- Smooth edges
- Single-lobed

Use curvatures at many scales to distinguish leaves!
Segmentation in HSV Colorspace

We do this by estimating foreground and background color distributions in the saturation-value space of the SV colorspace.
In Saturation-Value Space

Hue is not useful because the background often has a greenish tinge due to reflections from the leaf or surrounding foliage.
Segmentation in HSV Colorspace

Some difficulty with pine leaves→ they employed pixel weighting

Expectation-Maximization (6-7 iterations)

Use downsampled version during EM
Segmentation Results

Original  Initial Result (60 ms)
Segmentation Results

Original | Initial Result (60 ms) | Rem. False Positives (6 ms)
Segmentation Results

Original | Initial Result (60 ms) | Rem. False Positives (6 ms) | Remove Stem (36 ms)

Morphological operations
Computing Curvature

- Rotations?
- Segmentation Errors?
- Scale Changes?
- Complex Boundaries?
- Axis Alignment?

Differential measures: not robust on discrete grids
And amplify noise
Curvature Using Integral Measures

Curvature = (white pixels in circle)/area = 0.5 (straight)

[Connolly 1986] [Manay et al. 2006] [Pottmann et al. 2009]
Curvature Using Integral Measures

Curvature = (white pixels in circle)/area = 0.2 (convex)
Curvature Using Integral Measures

Curvature = (white pixels in circle)/area = 0.8 (concave)
Curvature Using Integral Measures

Curvature = (white pixels in circle)/area = 0.5
Build Histogram of Curvatures (HoCS)

compute histograms of the curvature values at each scale
Histograms of Curvature over Scale

25 Scales

Total time: 0.11s
- Upload image
- Perform recognition (5.4s)
  - Nearest neighbor classifier
- Get ranked results
Which species is it?
Accuracy on the Trees of the Northeast

Within top 5 matches 93% of time

1st match is right 72% of time
Which species is it?
Which species is it?

- Explore the images in the app
Which species is it?

- Explore the images in the app
  - Including the bark and stem
Which species is it?

- Explore the images in the app
  - Including the bark and stem
- Read the text descriptions

Japanese Maple
Acer palmatum

Native to Asia, this maple is commonly grown in the United States for its attractive leaf shape and bright colors. Palmately-lobed leaves (4-12 cm long and wide) turn vibrant shades of red in the fall. This small understory tree has a distinctive dome-like crown that provides light shade in gardens.

Habitat: Planted as an ornamental.

Growth Habit: Deciduous shrub or small tree, growing 4.6-6 m tall.

Bloom Time: Mid to late spring.
Which species is it?

- Explore the images in the app
  - Including the bark and stem
- Read the text descriptions
- Once confident, label it!
Look back through your collection…
Look back through your collection…
- Including the location!
- Nearly 1 million downloads
  - 40k new users per month
  - 100k active users
- 1.7 million images taken
  - 100k new images/month
  - 100k users with > 5 images
- Users from all over the world
- Botanists, educators, kids, hobbyists, photographers, …
－ very fast, suitable for use in an interactive application
－ adapt to major sources of color variability such as lighting changes and natural variations in leaf color
- Nearly 1 million downloads
  - 40k new users per month
  - 100k active users
- 1.7 million images taken
  - 100k new images/month
  - 100k users with > 5 images
- Users from all over the world
- Botanists, educators, kids, hobbyists, photographers, …
Lots of Future Directions

Education and Outreach

Tracking Biodiversity

Discovering New Species

© 2006 Noah Snavely
- Apps
- Dataset
- Code
available at leafsnap.com

- Plant images in apps taken by FindingSpecies.org

volunteers

...and users!
Thank You