Unwrap Mosaics: A new representation for video editing

Alex Rav-Acha, Pushmeet Kohli, Carsten Rother, Andrew Fitzgibbon
Microsoft Research
SIGGRAPH 2008
Abstract

• New representation for common editing tasks
• Representation has the power of full 3D surface reconstruction

• This representation is “Unwrap Mosaics”
• Editing can be performed and re-composited into the original sequence
What is a Mosaic?

A group of pictures used to represent a broader/wider image.

What is a Mosaic?

4 or 5 images used to form this mosaic

Mosaic Example

http://www.infotech.oulu.fi/Annual/2005/pics/mvg_01.jpg
Mosaic Example

http://www.vision.deis.unibo.it/MosPerf/MosPerf-Files/MosaicRM.jpg
Goal

To recover the object’s texture map, rather than its 3D shape

THE CONCEPT
Video represented as several unwrap mosaics
  • One per object
  • One per background
Goal

It is possible in principle to recover the 3D shape.

For many editing tasks this is not necessary: edits can be performed on the mosaic itself and re-rendered without ever converting to a 3D representation.
Unwrap Mosaics

A method to recover and edit dense models of deforming 3D surfaces from a video.

Accompanying the recovered texture map will be a 2D-to-2D mapping describing the texture map’s projection to the images, and a sequence of binary masks modeling occlusion.

The combination of texture map, 2D–2D mapping, and occlusion masks is called the unwrap mosaic.
Unwrap Mosaics

3D models – But not 3D!

Uses
- Wide range of editing operations on the input video
- Model easier to recover from video and helps us work with self-occlusion and deforming surfaces

The main contribution of this paper is the algorithm to recover the unwrap mosaic from images. This algorithm is an energy minimization procedure.
High level steps of Algorithm

Steps that correspond to minimization of the model energy via coordinate descent

1. Segmentation
2. Tracking
3. Embedding
4. Mosaic stitching
5. Track refinement
6. Iterate
Segmentation

Steps

• Segmentation into independent moving objects
• Variant used here is “video cut and paste” – allows segmentation maps to be obtained with relatively little effort
• Automatic segmentation maps computation is also possible.
Tracking

Aim
To recover the texture map of a deforming 3D object from a sequence of 2D images.

Although the model is changing its shape, the texture map may be assumed to be constant, as on skin and cloth. Consider a point on the object. As this point moves through the video, it generates a 2D trajectory or track. Conversely, standard computer vision algorithms for point detection and tracking can be used to compute such tracks from the input sequence.
Embedding

- Computational cornerstone of the method
- The tracks that were generated are viewed as high dimensional projection of 2D surfaces
- The point \((u; v)\) in parameter space generates a vector of image positions \((x_1; y_1; x_2; y_2; \ldots; x_T; y_T)\)
- We can recover the surface’s \((u; v)\) parameter space by computing an embedding of the point tracks into 2D, directly yielding \((u; v)\) coordinates for each track.
Mosaic stitching

- The embedding defines a map from the tracked points in each image to the \((u; v)\) parameter space.
- Interpolating this mapping allows each image to be warped into the common frame.
- A variation of the standard mosaic stitching technique [Agarwala et al. 2004] emerges naturally from the energy formulation.
- In simple terms, creating a mosaic from different images.
Track Refinement

- The mosaic is already available in this step (prev 3 steps build the mosaic)
- The generated mosaic is not accurate, but good enough to create a reference template to match the original frames
- Reduces drift that may have been present after the tracking phase
- Track info propagates to occluded areas giving complete description of the object motion
Iterate

- Repeat the above steps again to fine tune precision
- Hope to compute a better embedding and mosaic by iterating
Two Input Video Frames
Edited Mosaic
Two Output Video Frames
Limitations

1. It requires quite well textured objects in order to work
   - Textured surfaces are required for point tracking. Low texture, such as on a face, requires good focus and lighting. One dimensional textures (stripes, rather than spots) suffer from the aperture problem.

2. Motion blur reduces efficacy of point tracking.

3. Object with protrusions do not yield useful mosaics

4. Strong shadows will disrupt tracking. Smooth lighting is assumed

5. Limited to disc-topology objects. Eg: A rotating cylinder will be reconstructed as a long tapestry rather than a mosaic with cylindrical topology.
Disc topology limitation

Unwrap Mosaic
User interaction & Hinting

- Algorithm can work completely automatically
- Situations where lack of texture, repeated structures, or motion blur could cause the recovered mosaic not to cover all of the object.
- New brushes (brush strokes) are introduced to specify points that were not covered by automatic segmentation.
User interaction & Hinting

a. Automatic Computation
User interaction & Hinting

b. User brush strokes
c. New mosaic computed with only the brush strokes
Results

• For video editing, the primary criterion is obviously the range and quality of effects that are enabled by the model.
• Computation times for all of these sequences (show in the video) are of the order of a few hours.

Check out this video !!
Thank you

Any Questions ?!??!