Dynamic Shape Tracking via Region Matching

Ganesh Sundaramoorthi
Asst. Professor of EE and AMCS
KAUST

(Joint work with Yanchao Yang)
The Problem: “Shape Tracking”

**Given:** exact object segmentation in frame 1

**Determine:** exact object segmentation in subsequent frames
Why? Many Applications

- **Entertainment Applications**
  - Video editing
  - 3D reconstruction (3D video, etc...)

- **Medical Applications**
  - e.g., Accurate shape of ventricles helpful in analyzing structure / function of heart
Why? Fundamental to Vision

• **Bounding box tracking and drift**

  Accurate shape of object could help!

• **Shares some same difficulties as other vision problems (e.g., recognition)**
  - viewpoint / illumination / quantization / noise
Outline

- Very Brief Summary of Existing Techniques for Shape Tracking
- New Method for Shape Tracking
- Summary / Future Directions
Large Body of Work in Shape Tracking
Image Segmentation Based on Global Region Statistics
e.g., Active Contours (Kass et al., IJCV 1988), Graph Cuts (Kolmogorov et al., 2001), etc...

Example:
minimize for $R_{in}$ (object segmentation)

$$E(R_{in}) = -D(p_{in}, p_{out})^2,$$

($D$ distance between pdf’s )

$p_{in} =$ color histogram inside $R_{in}$

$p_{out} =$ color histogram inside $R_{out}$

Main Idea: Determine shape by dividing image into regions with maximally different global image statistics

Many developments on this idea to specialize for tracking:
- Better energies for segmentation
- Better optimization method for energies (convex relaxations, e.g., Bresson et al., 2009)
- Localize segmentation near segmentation from previous frame (Lankton/Tannenbaum TIP 2010)
- Predict motion of object to better localize segmentation
  - particle filters (Blake/Isard ECCV1998, Rathi et al. TPAMI 2007)
- Predict motion/deformation to do even better localization (Sundaramoorthi et al., SIAM 2011)
- etc ...
Large Body of Work in Shape Tracking

Main Problem: Object/background in images from complex scenes not easily separable by global statistics (intensity statistics shared by object/background)
Shape Tracking by Matching

**Main Observation:** Region statistics of object are roughly stationary across frames in video

**Main Idea:** Determine shape in next frame by deforming region to match statistics of region in next frame.

Frame # t

(Scene image)

Frame # t+1

(Scene image)

Transfer segmentation by matching region statistics

Obtain segmentation

(Segmentation given)

**No need to model background**
Main Challenges: Nuisances in Image Formation

- viewpoint / 3D deformation
- illumination
- occlusions
- quantization
- noise

“What makes [vision] hard is not that any of the above transformations are that hard to detect and decode in isolation, but rather that all of them tend to coexist, and then the decoding becomes hard.”

- Mumford, “Pattern Theory: A Unifying Perspective,” 1994
Our Model: Incorporating Occlusions

**Occlusion**: part of object that goes out of view between frames

**Dis-occlusion**: part of object that comes into view

**Dynamic Model:**

\[
R_{t+1} = w_t(R_t \setminus O_t) \cup D_{t+1}
\]

\[
a_{t+1}(x) = \begin{cases} 
  a_t(w_t^{-1}(x)) + \eta_t(x) & x \in w_t(R_t \setminus O_t) \\
  a_{t+1}^d(x) + \eta_t(x) & x \in D_{t+1}
\end{cases}
\]

Besides occlusion/dis-occlusion, this model is **Brightness-Constancy Plus Noise**

**Goal**: Determine segmentation in next frame, \( w(R \setminus O) \cup D \)
Overview: Frame-wise Processing

Given at each time $t$:
- estimate of radiance and object region (from previous frame)
- next frame

Step 1:
Joint estimation of
- occlusion
- deformation
- warped un-occluded region

Step 2:
- dis-occlusion detection
- final estimated object region in next frame
Joint Estimation of Occlusion and Deformation

- $w$ is defined on un-occluded region $R \setminus O$, but $O$ is unknown
- $O$ is subset of $R$ where $w$ is not defined, but $w$ is unknown

**w and O are coupled**: should be estimated *jointly*

### Optimization Problem:

\[
E(O, w, |I, a, R) = \int_{w(R \setminus O)} |I(x) - a(w^{-1}(x))|^2 dx + \alpha \text{Reg}(w) + \beta_o \text{Reg}(O)
\]

- **1st term**: penalizes $w$ that doesn’t transform radiance between frames
- **2nd term**: regularization on $w$ to resolve aperture ambiguity
- **3rd term**: regularization on $O$ corresponds to a prior on $O$
Optimization: Small Deformation First

As in optical flow problems, linearize the data term:

\[ w(x) = x + v(x), \text{ where } v(x) \text{ is small} \]

\( (w \text{ and } v \text{ defined on } R\backslash O) \)

As in optical flow problems, linearize the data term:

\[
E(O, v) = \int_{R\backslash O} |I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 \, dx + \alpha \int_{R\backslash O} |\nabla v(x)|^2 \, dx + \beta_0 \int_O dx
\]

Analogous to Mumford/Shah Problem (Mumford/Shah 1989)

- known solutions are extremely expensive and sensitive to local optima (Tsai et al., 2001, Vese and Chan 2001)
- fortunately with an adjustment of energy, can be optimized efficiently ...
Optimization: Small Deformation First

Let’s illustrate the simple idea: extend \( v \) to all of \( \mathbb{R} \):

\[
E(O, v) = \int_{\mathbb{R}\setminus O} |I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 dx + \alpha \int_{\mathbb{R}} |\nabla v(x)|^2 dx + \beta_o \int_{O} dx
\]

- Extended deformation need not fit data in \( O \) (desired)
- Modification makes optimization simple and efficient
- Empirically, achieves better optimizer (and faster convergence) than standard optimization of Mumford/Shah

Optimization: Small Deformation First

Given \( v \), optimizing in \( O \) is easy:

\[
E(O|v) = \int_{R\setminus O} |I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 dx + \alpha \int_R |\nabla v(x)|^2 dx + \beta_o \int_O dx
\]

Pixel \( x \) adds to the energy

\[
\begin{cases} 
|I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 & \text{if } x \text{ assigned to } R\setminus O \\
\beta_o & \text{if } x \text{ assigned to } O 
\end{cases}
\]

Since \( v \) is defined on all of \( R \), the global optimum for \( O \) given \( v \) is

\[
O = \{ x \in R : \beta_o < |I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 \}
\]

Optimizing for \( v \) given \( O \) is standard variational problem:

\[
E(v|O) = \int_{R\setminus O} |I(x) - a(x) - \nabla a(x) \cdot v(x)|^2 dx + \alpha \int_R |\nabla v(x)|^2 dx + \beta_o \int_O dx
\]
Optimization: Small Deformation First

Simple alternating minimization scheme:

1. **Initialize** \( O \) as empty set
2. **Optimize in \( v \) given \( O \):** solve linear system (fast using conjugate gradient):

\[
-\alpha \Delta v(x) = \begin{cases} 
(I(x) - a(x) - \nabla a(x) \cdot v(x)) & x \in \mathbb{R} \setminus O \\
0 & x \in O
\end{cases}
\]

3. **Smooth/threshold** to solve for occlusion \( O \) given \( v \)

\[
O = \{ x \in \mathbb{R} : (G_{\sigma} \ast F)(x) > \beta_o \}
\]

\[
F(x) = (I(x) - a(x) - \nabla a(x) \cdot v(x))^2
\]

4. **Repeat** until pixels of \( O \) do not change

At end, infinitesimal deformation \( v \) and occlusion \( O \) are determined
Larger Deformations: Iterate

1. Set backward map to be identity
   \[ \phi_0^{-1}(x) = x \]

2. Compute infinitesimal deformation / current estimate of occlusion
   \[ v_\tau, O_\tau = \arg\min_{O, v} E(O, v|a_\tau, I, R_\tau) \]

3. Deform region by deformation
   \[ \partial_\tau \Phi_\tau = \nabla \Phi_\tau \cdot v_\tau, \quad R_\tau = \{ \Phi_\tau < 0 \} \]
   \[ \partial_\tau \phi_\tau^{-1} = -\nabla \phi_\tau^{-1} \cdot v_\tau, \text{ in } R_\tau \]

4. Update backward map
   \[ a_\tau = a \circ \phi_\tau^{-1} \]

5. Update region radiance

6. Repeat 2-5 until deformation is zero

For convenience, deforming region R is represented implicitly via level sets
(Osher & Sethian 1988)
Simulation of Joint Deformation/Occlusion Estimation

Radiance, $a$

Region $R$ in Image to Match

Infinitesimal Deformation, $\nu$

Occlusion Likelihood Map
Result of Occlusion/Deformation Estimation

Region and Radiance in Frame $t$

Warped Region in Frame $t+1$

Warped Occlusion Determined (White)

Warped Occlusion Removed
Un-occluded Region Determined

Dis-occlusion must be determined
Determining the Dis-occlusion

Prior assumptions on the object appearance and/or shape must be made

**Assumptions:**
- self-similarity of radiance function
- dis-occlusion is close to warped un-occluded region
- area of dis-occlusion not too large wrt object region (last two true for high camera frame-rate)

**Optimization problem to determine dis-occlusion** $D$:

$$E_d(D \mid I, a', R') = - \int_D p(x)dx + \beta_d \text{Area}(D) , \ D \subset \{x : 0 < d_{R'}(x) \leq \varepsilon\}$$

**Solution:**

$$D = \{x : 0 < d_{R'}(x) \leq \varepsilon , G_\sigma * p(x) > \beta_d\}$$

Likelihood pixel belongs to dis-occlusion:

$$p(x) \propto \exp\left(-\frac{d_{R'}(x)^2}{2\sigma_\varepsilon^2} + p_f(I(x)) - p_b(I(x)) \right)$$

large distances from un-occluded region unlikely

large if pixel appearance is more similar to un-occluded region than background
Sample Dis-occlusion Detected

warped un-occluded region (green) likelihood of dis-occlusion detected dis-occlusion

final region (dis-occlusion union warped un-occluded region) in frame $t+1$
Experiment: Occlusion/Dis-occlusion

Modeling is Crucial

Occlusion / dis-occlusion not modeled

Occlusions modeled, but not dis-occlusions

Dis-occlusions detected, but occlusions not modeled

Occlusions and dis-occlusions modeled
Comparisons to State-of-the-Art

Comparison to
- Adobe After Effects 2012, some components based on
Comparisons to State-of-the-Art
Comparisons to State-of-the-Art
Comparisons to State-of-the-Art
Summary

• **Shape tracking can be treated as a matching problem** rather than segmentation of intensity statistics: beneficial for complex appearance / clutter

• **Occlusions / dis-occlusions posed a challenge to tracking by matching**

• **Occlusions, warp treated as a joint problem** (Mumford-Shah)

• **Dis-occlusions required discrimination of intensity statistics**
Future Directions

• Tracking under larger displacements/deformations (smaller frame-rate)

Overlay of Frame 1 on 2  Frame 1  Frame 2

• Dis-occlusions by segmentation of more descriptive statistics?
Thank-you; Questions?

• References for this presentation:

• Y. Yang and G. Sundaramoorthi, “Modeling Self-Occlusions in Dynamic Shape and Appearance Tracking,” ICCV 2013
  • http://vision.ucla.edu/~ganeshs/articulated_object_tracking_html/ObjectTrackingSelfOcclusions.html


• Website: https://sites.google.com/site/ganeshsun/