Image Segmentation and 3-D Reconstruction via Low-Rank Texture

Allen Y. Yang
Department of EECS, UCB
yang@eecs.berkeley.edu

Lawrence Berkeley Laboratory, October 4, 2011
Distributed Sensing and Perception

Centralized Perception

Up: powerful processors
Up: unlimited memory
Up: unlimited bandwidth
Down: single modality

Distributed Perception

Down: mobile processors
Down: limited onboard memory
Down: band-limited communications
Up: distributed, multi-modality

When the sensing resources are limited or scarce:

1. What is the optimal strategy to deploy mobile sensors?
2. How to effectively take local measurements of the events?
3. How to properly tally the local observations to reach a global consensus?
When the sensing resources are limited or scarce:

1. What is the optimal strategy to deploy mobile sensors?
2. How to effectively take local measurements of the events?
3. How to properly tally the local observations to reach a global consensus?
Some Examples

1. Robust face recognition

2. 3-D motion segmentation with multiple rigid bodies and outliers

3. Accelerated parallel implementation of sparsity-minimization problems ($\ell_1$-min)
Outline: How to effectively represent texture in natural images?

1 Natural Image Segmentation via Texture and Boundary Compression

- Optimal segmentation achieved via the shortest coding length of texture and boundary.

\[ L^S_{w, \varepsilon}(\mathcal{R}) \triangleq \sum_{i=1}^{k} L_{w, \varepsilon}(R_i) + \frac{1}{2} B(R_i). \]  

- A pairwise steepest descent optimization.

animation

http://www.eecs.berkeley.edu/~yang
Low-Rank Texture for **Holistic 3-D Reconstruction**

- Extract low-rank texture regions via **Robust PCA**

\[
A^* = \arg \min_{A, E, \tau} \|A\|_* + \lambda \|E\|_1 \quad \text{subj. to} \quad I \circ \tau = A + E.
\]

- From low-rank texture to 3-D model.
Image Segmentation

Image segmentation literature

- Based on color: For example, using EM or K-Means.

- Based on texture:

- Based on semantic or domain-specific knowledge:
Texture-Based Image Segmentation

- Challenges:
  1. How to define the problem mathematically?
  2. How to determine data distribution and the number of models?
Texture-Based Image Segmentation

- Challenges:
  1. How to define the problem mathematically?
  2. How to determine data distribution and the number of models?

- Technical Highlights:
  1. Our model is a mixture Gaussian and (degenerate) mixture subspaces.
  2. Optimal segmentation achieved via the shortest coding length to compress.
  3. Segmentation applied to natural images is consistent with human vision.
Construct Texture Measurements

Texture Data in Vector Form:

1. Response to 2-D texture filter banks.

2. Appearance model from cut-off local windows.

Reference:
Coding a single texture region

- Given a set of vectors $X = (x_1, \cdots, x_N) \in \mathbb{R}^{D \times N}$, a **lossy coding scheme** maps the vectors to a sequence of binary bits, such that the original vectors can be recovered up to a distortion

$$
\mathbb{E}[\|x_i - \hat{x}_i\|^2] \leq \epsilon^2.
$$
Coding a single texture region

- Given a set of vectors \( X = (x_1, \cdots, x_N) \in \mathbb{R}^{D \times N} \), a **lossy coding scheme** maps the vectors to a sequence of binary bits, such that the original vectors can be recovered up to a distortion

\[
\mathbb{E}[\| x_i - \hat{x}_i \|_2^2] \leq \epsilon^2.
\]

- The **coding length of a Gaussian model** is denoted as \( L(V) : \mathbb{R}^{D \times N} \to \mathbb{Z}_+ \)

[Cover-Thomas1991, Tse-Viswanath2005].

\[
L_\epsilon(\hat{X}) = \frac{D}{2} \log_2 \det(I + \frac{D}{\epsilon^2} \Sigma) + \frac{N}{2} \log_2 \det(I + \frac{D}{\epsilon^2} \Sigma) + \frac{D}{2} \log_2(1 + \frac{\| \mu \|^2}{\epsilon^2}),
\]

(2)
Coding the boundary of a texture region

- In the presence of **multiple texture regions**, the membership of each pixel needs to be coded.
Coding the boundary of a texture region

- In the presence of multiple texture regions, the membership of each pixel needs to be coded.
- Freeman chain code:
  \[
  B(R) = 3 \sum_{i=0}^{7} \#(o_t = i). 
  \]  
  (3)
- Difference chain code leads to smoother boundary
  \[
  B(R) = - \sum_{i=0}^{7} \#(\Delta o_t = i) \log_2(P[\Delta o = i]).
  \]  
  (4)
Minimization of the Total Coding Length Function

- Given a hypothetical segmentation \( \mathcal{R} = \{R_1, \ldots, R_k\} \),

\[
L_{w,\varepsilon}^S(\mathcal{R}) = \sum_{i=1}^{k} L_{w,\varepsilon}(R_i) + \frac{1}{2} B(R_i). \tag{5}
\]

Clearly, minimizing \( L_{w,\varepsilon} \) over all possible segmentation is combinatorial. Approximate by agglomerative region merging:
Minimization of the Total Coding Length Function

- Given a hypothetical segmentation $\mathcal{R} = \{R_1, \ldots, R_k\}$,

$$L^S_{w, \varepsilon}(\mathcal{R}) = \sum_{i=1}^{k} L_{w, \varepsilon}(R_i) + \frac{1}{2} B(R_i).$$ (5)

- Clearly, minimizing $L^S$ over all possible segmentation is combinatorial.

- Approximate by agglomerative region merging:

$$(R^*_i, R^*_j) = \arg\max_{R_i, R_j \in \mathcal{R}} \Delta L_{w, \varepsilon}(R_i, R_j)$$

animation

http://www.eecs.berkeley.edu/~yang
Simulations

(g) one plane two lines  
(h) two planes one line  
(i) three planes

Image Segmentation Results on Berkeley Segmentation Dataset


http://www.eecs.berkeley.edu/~yang
Reconstruct Large-Scale 3D Models using Low-Rank Texture

- Recover a motion model from a pair of images
  Structure-from-Motion

\[ x_2^T F x_1 = 0 \]

Small-baseline matching generates dense point clouds
Reconstruct Large-Scale 3D Models using Low-Rank Texture

- Recover a motion model from a pair of images
  Structure-from-Motion
  \[ x_2^T F x_1 = 0 \]

Small-baseline matching generates dense point clouds

- Large-baseline matching is prone to ambiguities
Introduce a New Class of Robust Features for 3-D Modeling

Ambiguities of Image Features

Challenges in large-baseline matching stem from the fact that traditional features (points, lines, patches, etc) do NOT possess any 3-D geometric information alone.

- Introduce a new class of features that contain richer 3-D information.
Introduce a New Class of Robust Features for 3-D Modeling

Ambiguities of Image Features

Challenges in large-baseline matching stem from the fact that traditional features (points, lines, patches, etc) do NOT possess any 3-D geometric information alone.

- Introduce a new class of features that contain richer 3-D information.

Questions:

1. How to properly define large image regions with richer 3-D information?
2. How to match these regions in two camera views?
3. Is the process robust to image corruption and occlusion?

Answer: Image texture sparsity in rank provides a principled solution.
Many object images are low-rank

(e) Output ($r = 14$)  (f) Output ($r = 8$)  (g) Output ($r = 19$)  (h) Output ($r = 6$)
Many object images are low-rank

Low-rank texture (TILT) features possess rich 3-D information

TILT: Transform Invariant Low-rank Texture [Zhang et al. '10]
Estimating Low-Rank Texture under Transformation and Corruption

- Objective function [Zhang et al. ’10]

\[
\min_{A, E, \tau} \text{rank}(A) + \lambda \|E\|_0 \quad \text{subj. to} \quad I \circ \tau = A + E,
\]

where \( A \) is low-rank and \( E \) is sparse, \( \tau \) parametrizes an image transformation.

http://www.eecs.berkeley.edu/~yang
Estimating Low-Rank Texture under Transformation and Corruption

- Objective function [Zhang et al. '10]

$$\min_{A, E, \tau} \text{rank}(A) + \lambda \|E\|_0 \quad \text{subj. to} \quad I \circ \tau = A + E,$$

where $A$ is low-rank and $E$ is sparse, $\tau$ parametrizes an image transformation.

- An iterative solution using Robust PCA [Candès et al. '10]

$$\min_{A, E, \Delta \tau} \|A\|_* + \lambda \|E\|_1 \quad \text{subj. to} \quad I \circ \tau_k + \nabla I \Delta \tau = A + E,$$

which also has a corresponding ALM formulation [Lin et al. '10]

$$\|A\|_* + \lambda \|E\|_1 + \langle Y, I \circ \tau_k + \nabla I \Delta \tau - A - E \rangle + \frac{\mu}{2} \|I \circ \tau_k + \nabla I \Delta \tau - A - E\|_F^2$$

Its cost is roughly equal to a small constant times the cost of SVD.
More TILT Examples

Corner point

Edge

Characters

Frieze Pattern

Face

Objects

http://www.eecs.berkeley.edu/~yang
Large-Baseline Matching I: Two Intersecting Facades in One Image

Detect intersection using TILT features:

\[
\begin{align*}
\min & \quad \|[A_1, A_2]\|_1 + \lambda\|[E_1, E_2]\|_1 \\
\text{subj. to} & \quad [l_1 \circ \tau_1, l_2 \circ \tau_2] = [A_1, A_2] + [E_1, E_2]
\end{align*}
\]

More Examples

http://www.eecs.berkeley.edu/~yang
Large-Baseline Matching II: One Common Feature in Two Images

(a) View 1  (b) View 2  (c) Alignment result

- **Pixel-wise alignment in orthographic views** up to subpixel accuracy

\[
A_2^* = \arg \max_{\phi=(x,y,u,v)} \frac{\mathbf{vec}(A_1)^T \mathbf{vec}(A_2 \circ \phi)}{\| \mathbf{vec}(A_1) \| \cdot \| \mathbf{vec}(A_2 \circ \phi) \|}
\]
Complete 3-D Reconstruction Pipeline

- Partial 3-D reconstruction from a single image without extracting any points and lines.

- Full 3-D reconstruction from multiple images.


http://www.eecs.berkeley.edu/~yang

Image Segmentation and 3-D Reconstruction via Low-Rank Texture
Acknowledgments

- UC Berkeley
  Dr S. Sastry

- Univ. Illinois
  H. Mobahi, Dr. S. Rao, Z. Zhou

- MSR Asia
  Dr Y. Ma

Publications


Funding Support

- ARO MURI: Heterogeneous Sensor Networks in Urban Terrains