Figure-Ground Segmentation

Bastian Leibe

Mobile Multimedia Processing
Computer Science 8 - Computer Graphics, Vision, and Multimedia
RWTH Aachen University

ICCV’11 Tutorial “Looking at People”, Barcelona, 07.11.2011
Figure-Ground Segmentation - Pixel based

- Background subtraction

  New image - Background model (learned over several frames) = Pixels not explained by background model

- Main issues
  - Choice of background model (parametric, non-parametric)
  - Updating strategy to cope with temporal changes
  - Shadow removal
  - Dealing with camera motion

  ⇒ Good overview over current techniques in Chapter 3.

B. Leibe

Image source: Tobias Jäggli
Figure-Ground Segmentation - Object based

- Necessary for scenarios with moving cameras

Main ideas

- Apply an object detector to find objects of a certain class
- Based on the detections, obtain figure-ground segmentations (useful for more detailed articulated body pose analysis)
- Propagate this information over time

⇒ Overview given in Chapter 4, and topic of this presentation
Outline

• Object detection
  - Popular detector representations
  - Incorporating geometric constraints
  - Use for tracking-by-detection

• Figure-ground segmentation
  - Class-agnostic segmentation
  - Class-specific top-down segmentation
  - Use for articulated multi-person tracking

• Object propagation
  - Propagating detections
  - Propagating object-specific classifiers
  - Propagating segmentations
Sliding-Window Object Detection

- Brute-force approach with many local decisions
  - Easy to combine with powerful machine learning techniques
  - Very effective in practice, dominant detection paradigm

B. Leibe
Representations for Human Detection

- Holistic representations
  - *Histograms of Oriented Gradients (HOG)*
  - Single feature vector encoding entire image window content
  - Efficient evaluation with linear SVM

- Part-based representations
  - *Deformable Part-based Model (DPM)*
  - Root filter (similar to HOG)
  - 5-6 additional part filters at higher resolution
  - Pictorial Structure deformation model
  - Part appearance and location learned by Latent SVM
  - State-of-the-art detector for many classes

[Dalal & Triggs, CVPR 2005]

[Felzenszwalb et al., PAMI 2010]
Representations for Human Detection (2)

- Feature-based representations
  - *Implicit Shape Model (ISM)*
  - Model objects as an assembly of 1000s of local features with learned spatial probability distributions relative to the object center
  - Find object locations by Hough Voting
  - *Hough Forests (HF)*
  - Densely sampled features
  - Efficient feature classification with Randomized Forest, leaves store spatial probability distributions
  - Find object locations by Hough Voting
  - Discriminative training

[Leibe et al., IJCV 2008]

[Gall & Lempitsky, CVPR’09]
Incorporating Ground Plane Constraints

• Efficient integration into detector design (*groundHOG*)
  ➢ Idea: only evaluate geometrically valid detection windows
  ➢ Derivation: Region of interest lies between two parabolas...
  ➢ ...that can in most cases be approximated by straight lines.
  ⇒ Only touch pixels inside the ROI for all computations.
  ⇒ Factor 2-4 speed improvement on top of all other optimizations

[P. Sudowe, B. Leibe, ICVS’11]
You Can Try This At Home...

- Detector code is publicly available
  - **HOG:** Dalal’s original implementation: [http://www.navneetdalal.com/software/](http://www.navneetdalal.com/software/)
    - Our CUDA-optimized *groundHOG* code (>80 fps on GTX 580) [http://www.mmp.rwth-aachen.de/projects/groundhog](http://www.mmp.rwth-aachen.de/projects/groundhog)
  - **DPM:** Felzenswalb’s original implementation: [http://www.cs.uchicago.edu/~pff/latent](http://www.cs.uchicago.edu/~pff/latent)
  - **ISM:** My original implementation: [http://www.vision.ee.ethz.ch/~bleibe/code/ism.html](http://www.vision.ee.ethz.ch/~bleibe/code/ism.html)
  - **HF:** Gall’s original implementation: [http://www.vision.ee.ethz.ch/~gallju/index.html#software](http://www.vision.ee.ethz.ch/~gallju/index.html#software)
  - **HF+ISM:** Planning to release code soon…
Application: Tracking-by-Detection

Object detections

Simple f/g model: Elliptical region in detection box

Joint optimization
(e.g., using Model Selection)

Spacetime trajectories

3D Camera path estimation

[Leibe, Cornelis, Cornelis, Van Gool, CVPR’07]
[Leibe, Schindler, Van Gool, ICCV’07]
Mobile Tracking Through Crowds

(Slightly simplified implementation of this approach runs at 8-10 fps)

[Ess, Leibe, Schindler, Van Gool, CVPR’08; ICRA’09]
Outline

• Object detection
  - Popular detector representations
  - Incorporating geometric constraints
  - Use for tracking-by-detection

• Figure-ground segmentation
  - Class-agnostic segmentation
  - Class-specific top-down segmentation
  - Use for articulated multi-person tracking

• Object propagation
  - Propagating detections
  - Propagating object-specific classifiers
  - Propagating segmentations
Class Agnostic Segmentation

- Idea
  - Initialize object mask from high-confidence object detections
  - Use general-purpose Grabcut segmentation by [Rother04] to find object region
  - Used successfully in [Ferrari, Marin, Zisserman, CVPR’08]
Class Agnostic Segmentation (2)

- This can be generalized to multi-view detection
  - E.g., using a multi-view DPM

- Initialize Grabcut mask by recognized object view.

- Foreground confidence map by averaging over several detections
  - Input for further pose parsing

Figure-14

B. Leibe

[Wang, Leibe, RSS’11]
Class-Specific Top-Down Segmentation

• During initial Hough Voting
  ➢ When we first observe a feature, we do not know its context.
  ➢ Different figure-ground labels may be consistent with the appearance.
  ⇒ Strategy: we cast votes for many locations...

• After voting
  ➢ Voting groups features that are consistent with the same object.
  ➢ We can now consider each feature conditioned on the selected object location hypothesis.
  ➢ This allows us to backproject a local figure-ground label from selected votes.
ISM Top-Down Segmentation

- Interpretation of $p(\text{figure})$ map
  - Per-pixel confidence in object hypothesis
  - Use for hypothesis verification

Bastian Leibe
Top-Down Segmentation with Hough Forests

- Extend HFs with top-down segmentation mechanism
- Better results than for ISM due to dense sampling

[K. Rematas, B. Leibe, CORP’11]
HF-ISM: Qualitative Results

- Observations
  - Improved detection performance compared to original HF (competitive with HOG + HIKSVM on pedestrians).
  - Better segmentations than original ISM due to dense sampling.

(no ground plane constraints used)
Application: Articulated Multi-Person Tracking

- **Multi-Person tracking**
  - Recovers trajectories and solves data association

- **Articulated Tracking**
  - Estimates detailed body pose for each tracked person

[Gammeter, Ess, Jaeggli, Schindler, Leibe, Van Gool, ECCV’08]
Guided Adaptive Segmentation

- Guided segmentation for each frame
  - No reliance on background modeling
  - Approach applicable to scenarios with moving camera
  - Feedback from body pose estimate to improve segmentation
Guided Segmentation Results

No detections available here
Articulated Tracking under Egomotion

[Gammeter, Ess, Jaeggli, Schindler, Leibe, Van Gool, ECCV’08]
Outline

- **Object detection**
  - Popular detector representations
  - Incorporating geometric constraints
  - Use for tracking-by-detection

- **Figure-ground segmentation**
  - Class-agnostic segmentation
  - Class-specific top-down segmentation
  - Use for articulated multi-person tracking

- **Object propagation**
  - Propagating detections
  - Propagating object-specific classifiers
  - Propagating segmentations
Propagating Detections

- Detector output is often not perfect
  - Missing detections and false positives
  - But continuous detector confidence still contains useful cues.

- Idea employed here:
  - Use detector confidence to track persons over time.
Detector Confidence Particle Filter

• Main idea
  ➢ Initialize particle cloud on strong object detections.
  ➢ Propagate particles using continuous detector confidence as observation model.

• Disambiguate between different persons
  ➢ Train a person-specific classifier with online boosting.
  ➢ Use classifier output to distinguish between nearby persons.

[Breitenstein, Reichlin, Leibe et al., ICCV’09; PAMI’11]
Results

[Breitenstein, Reichlin, Leibe et al., ICCV’09; PAMI’11]
Propagating Object-Specific Classifiers

- Tracking by online classification
  - Train an online classifier to distinguish a specific object from its background
  - Main issue: robust classifier update to prevent drift
  - Many approaches available, e.g., [Avidan’05, Grabner’06, Kalal’10]

- Related problem
  - Train an online classifier to disambiguate between several tracked objects
  - Helps data association for robust tracking in crowded situations
  - E.g., [Kuo, Huang, Nevatia, CVPR’10]
Propagating Segmentations

- **Hybrid tracking framework**
  - Combine detection with fast low-level tracking.
  - Initialize tracks with object detections.
  - Follow each persons using cheap segmentation-based tracker.
  - (Alternative: using person-specific online classifier)

[D. Mitzel, E. Horbert, A. Ess, B. Leibe, ECCV’10]
Detector-based Reinitialization

- Tracked contour degrades due to lighting changes...
- Reinitialization is triggered.
- Tracking continues...

- Important to integrate failure detection
  - LS tracker automatically detects when its contour degrades.
  - Request a new detection and reinitalize.

⇒ Object detector only needs to be activated every $k$ frames!

[D. Mitzel, E. Horbert, A. Ess, B. Leibe, ECCV’10]
Results with High-Level Tracker

[D. Mitzel, E. Horbert, A. Ess, B. Leibe, ECCV’10]
Extension: Improved Appearance Models

(Results without high-level tracker)

[E. Horbert, K. Rematas, B. Leibe, ICCV’11]
Conclusions

- **Object based methods**
  - Powerful object detection approaches available
  - Applicable for known object categories
  - Enables tracking-by-detection with a moving camera

- **Figure-ground segmentation as a result of detection**
  - Segmentation as a refinement after detection
  - Both generic and class-specific approaches available
  - Enables more detailed body pose analysis

- **Propagating object information over time**
  - Tracking using continuous detector confidence
  - Tracking by online classification
  - Hybrid low-level/high-level tracking
  - Reduce dependence on generic object detector.
Thank you very much!

Questions?