Human Detection and Tracking in Crowded Scenes Using a Fast Mean Shift Procedure

C. Beleznai\(^1\), T. Schloegl\(^1\), B. Fruehstuck\(^2\), H. Bischof\(^3\)

\(^1\)Advanced Computer Vision GmbH. – ACV, Vienna, Austria
\(^2\)Siemens AG Österreich, Programm- und Systementwicklung, Graz, Austria
\(^3\)ICG, Graz University of Technology, Austria
Contents

1. Introduction

2. Human detection and tracking in crowded scenarios
   - state-of-the art
   - human detection using mean shift
   - fast mean shift computation
   - results and evaluation

3. Mean shift mode tracking

4. Extensions
   - arbitrary camera geometry
   - fast data-driven clustering

5. Concluding remarks
Introduction - the company

Kplus Research Areas and Industrial Applications:

A  Visual surveillance and tracking

B  3-Dimensional Segmentation, Modelling and Reconstruction

C  Statistical methods and learning

D  Matching
Human detection and tracking in crowded scenarios

Introduction:

Application: Video surveillance

Objectives:

Robust detection of moving humans by a static camera.
- Detection in occluded situations
- Tracking (short duration of occlusions)
Human detection and tracking in crowded scenarios

Desired output:
- how many objects,
- approximate location
- consistent motion path

original  "blob" image  detected "blob" objects  proposed method
State-of-the-art approaches

• Silhouette analysis
  (Kuno 1996, Haritaoglu 2001)

• Color-based segmentation / tracking
  (Elgammal 2001, Comaniciu 2000)

• Particle filter
  (Kuno 1996, Isard 2001)

• EM clustering
  (Pece 2000)

• Appearance models
  (Senior 2001)

• Stochastic segmentation
  (Zhao 2003)
Real-time mean shift-based human localization

- change detection – difference between frame and a reference
- no thresholding – clustering
- fast mean shift algorithm
- model-based validation
Mean shift offset:

\[ \Delta_x = \frac{\sum_a K(a - x) w(a) a}{\sum_a K(a - x) w(a)} - x \]

\( K \) is a kernel function

\( w(a) \) is the weight (intensity) at data point (pixel) \( a \)

Concept introduced by:

Fukunaga and Hostetler (1975)
Cheng (1995)
Comaniciu and Meer (1998)
**Mean shift clustering**

**Assumption:** Difference image: high intensity ~ high probability of motion

1. locating initial points (sample set),
2. mean shift procedure until convergence,
3. mode grouping

**Output:** mode, basin of attraction, attraction path, points along attraction path
**Fast Mean Shift Computation**

Boxlets - P. Simard et al. (1999)

A fundamental property of convolution operation:

\[(f \ast g)^n = f^n \ast g = f \ast g^n\]

Thus:

\[(f'') \ast (\int \int g) = f \ast g.\]

Integral image

Convolution can be significantly accelerated, if the 2nd derivative of \(f\) is sparse.

\[\Delta_x = \frac{\sum_a K(a - x) w(a) \ a}{\sum_a K(a - x) w(a)} - x\]
Fast Mean Shift Computation using Integral Images

Using:

\[ K \ast (w(a) a) = K'' \ast \int \int w(a) a \]

Mean shift offset:

\[ \Delta x = \sum_a \frac{K''(a - x) \sum_{i < a} w(i) i}{\sum_a [K''(a - x) \sum_{i < a} w(i)]} - x \]

Speedup factors (Matlab):

- 6 window size: 30-by-50 pixels
- 30 window size: 90-by-90 pixels
Fast Mean Shift Computation using Integral Images

3 integral images (SAT) are computed:

- Fast computation of integral images
  \[
  I_{\text{int}}(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')
  \]
  \[
  I_{\text{int}}(x, y) = I_{\text{int}}(x, y - 1) + I_{\text{int}}(x - 1, y) \\
  + I(x, y) - I_{\text{int}}(x - 1, y - 1)
  \]

  Sum of intensities within a region:
  \[
  S_{\text{area}} = I_A + I_D - I_B - I_C
  \]

- Simple human model consisting of rectangular regions
- Fast Hypothesize-and-Test steps are possible
Model-based validation

- **Penalized likelihood criterion**

1. Model insertion at detected mode

2. Cost computation

   \[ C(\theta_z) = (1 - P(I|\theta_z))e^{\beta Z} \]

   using

   \[ P(I|\theta) = \exp\left( -a \left[ 1 - \frac{1}{A_{\theta}} \sum_{x,y \in R_{\theta}} I(x,y) \right] - b \left[ \frac{1}{A_{\theta}} \sum_{x,y \in R_{\theta}} I(x,y) \right] \right) \]

3. Model insertion at most probable location.

4. Insertion stopped upon cost increase
Evaluating detection results

Ground truth
(manual annotation)

Moving humans with more than 50% visibility

One-to-one mapping between ground truth and detection results
Detection experiments (independent processing of each frame)

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of annotated frames</td>
<td>187</td>
</tr>
<tr>
<td>Valid humans</td>
<td>4749</td>
</tr>
<tr>
<td>Detection rate</td>
<td>81.2%</td>
</tr>
<tr>
<td>False alarm rate</td>
<td>12.1%</td>
</tr>
</tbody>
</table>
Detection experiments (independent processing of each frame)
# Evaluation of detection performance

## Two evaluation sequences:

<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of annotated frames</td>
<td>470</td>
<td>879</td>
<td></td>
</tr>
<tr>
<td>Valid humans</td>
<td>6147</td>
<td>5380</td>
<td></td>
</tr>
</tbody>
</table>

**blob-based method**

<table>
<thead>
<tr>
<th></th>
<th>Correct detections</th>
<th>3096</th>
<th>2762</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit rate</td>
<td>50.3%</td>
<td>51.3%</td>
<td></td>
</tr>
<tr>
<td>False alarm rate</td>
<td>23%</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

**proposed approach**

<table>
<thead>
<tr>
<th></th>
<th>Correct detections</th>
<th>5400</th>
<th>4533</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit rate</td>
<td>87.9%</td>
<td>84.3%</td>
<td></td>
</tr>
<tr>
<td>False alarm rate</td>
<td>29%</td>
<td>19%</td>
<td></td>
</tr>
</tbody>
</table>
Combining mean shift-based detection with tracking

Detected humans generate hypotheses

Hypothesis-object association on frame-to-frame basis.

Occlusions:
Priors on the number of objects constituting the group are available.

Model-based validation using priors
Tracking experiment: crowded indoor scene

- Blob tracking
- Mean shift-based detection and tracking

Evaluation of tracking performance:
- Number of annotated frames: 1013
- Number of unmatched tracks: 292 (83.2%)
- Track integrity: 1.4
Tracking experiment
Tracking experiment
Mean Shift Mode Tracking
Mean shift procedure using oriented kernels
### Mean shift procedure using oriented kernels

#### Operations for Area Computation

<table>
<thead>
<tr>
<th>Angle</th>
<th>Operations for Creation</th>
<th>Approx.</th>
<th>Operations for Area Computation</th>
</tr>
</thead>
<tbody>
<tr>
<td>45°</td>
<td>4/pixel</td>
<td>Exact</td>
<td></td>
</tr>
<tr>
<td>27°, 63°</td>
<td>7/pixel</td>
<td>Even side lengths</td>
<td>3</td>
</tr>
</tbody>
</table>
Mean shift procedure using oriented kernels
Fast mean shift-based clustering

**Unconstrained clustering:** kernel size – unknown parameter

**Distribution:** multi-modal, multi-scale patterns

**Known:** range of scales at which structures appear
Fast mean shift-based clustering

1. Generating sample set with \((x_0, \sigma_0)\)

2. Mean shift mode seeking

3. Estimating local covariance

4. Orienting kernel (discrete angles)

5. Stop upon convergence
Fast mean shift-based clustering
Fast mean shift-based clustering

Scale selection problem: *most-stable-over-scales* criterion
Fast mean shift-based clustering - Experiment
Fast mean shift-based clustering

Applied to:

• Clustering regions within cornerness measure map
• Texture similarity measure
• Clustering the output of boosted cascade classifier
• Finding significant modes in 2D color histograms
• Intensity template correlation
Summary

- Fast mean shift-based clustering:
  Relying on unfiltered data, number of clusters is unknown

- Efficient combination of low-level information and independent high-level knowledge

- Promising (real-time) performance on challenging data

- Data-driven model selection - detecting arbitrary objects
Future Work

- Mean shift-based feature point tracking
- Data association using spatio-temporal reasoning
- Inferring model from clustering