Data clustering for image segmentation

Dipartimento di Informatica e Sistemistica

Antonio Ruberti

Sapienza
Università di Roma

Domenico Bloisi
bloisi@dis.uniroma1.it
Summary

• ARGOS Project overview
• Image Segmentation and Optical Flow (OF)
• K-means based OF data clustering
• Rek-means vs k-means
• Examples and experimental evaluation
ARGOS project overview

Automatic Remote Grand Canal Observation System

The ARGOS system controls a waterway of about 6 km length, 80 to 150 meters width, through 14 observation posts (Survey Cells).
Rai Tg1 video
ARGOS objectives

• management and evaluation of navigation rules
• traffic statistics and analysis
• security
• preservation of historical heritage (reduction of wave motion)
Motivation

Human Surveillance

Until 2006 the Municipal Administration of Venice paid people for boats counting:

- Error prone
- Una tantum

Automatic Surveillance

From November 2007 Venice has an automatic video surveillance system

- 24 hours/day  7 days/week
- recorded traffic violations
- Certified error
ARGOS functions

• optical detection and tracking of moving targets
• computing position, speed and heading of targets
• event detection (speed limits, access control, …)
• recording 24/7 video and track information (post-analysis)
• rectifying camera frames and stitching them into a composite view
• automatic PTZ tracking
• on line traffic information http://www.argos.venezia.it
Examples

- Detect boats docking in the highlighted area
- Speed limit control
ARGOS features

- Dynamic background (water)
- Multi object tracking (up to 10 boats in the same view)
- Multi camera for large areas (56 cameras x 6 km length)
- Third party extensive evaluation (2.2\cdot10^7 frames analyzed)
General architecture

![Diagram showing the general architecture of the system with modules for camera input, segmentation, tracking, and event handling.]

- **Camera i-1**
  - Segmentation Module
  - Tracking Module

- **Camera i**
  - Background Estimate
  - Foreground Extraction
  - Association Estimate
  - Tracking Module

- **Camera i+1**
  - Segmentation Module
  - Tracking Module

- Event Handling Module

- Multiple camera data fusion
Survey cell

3 high resolution network cameras, a PTZ camera for zoom and tracking of the selected target, and 2 computers running the image processing and tracking software.

The survey cells are installed on the top of several buildings leaning over the Grand Canal.
Survey cells
Segmentation module

Rek-means clustering algorithm
Segmentation
(detecting objects of interest)

- Background Formation
- Foreground Computation
- Background Model

Current Frame $I$

List of Detected Objects

continuously updated

$B$
Background formation

Problems:
- gradual illumination changes and sudden ones (clouds)
- motion changes (camera oscillations)
- high frequency noise (waves in our case)
- changes in the background geometry (parked boats).

Approach:
- computation of color distribution of a set of frames
- highest component form the background
Background modeling

Set $S$ of $n$ images from a camera

Natural images

Artificial image

Background Image computed from $S$ (the image display only the higher gaussian values)
Foreground computation

(background subtraction technique)

current frame

THRESHOLD $T$

(based on illumination conditions)

$>$

blobs (Binary Large Objects)

foreground image

background image
Tracking module

Single-hypothesis Tracking
We use a set of Kalman Filters (one for each tracked boat).

Data Association: Nearest Neighbor rule
Track formation: unassociated observations
Track deletion: high covariance in the filter

Multi-hypotheses Tracking
Track splitting: in ambiguous cases (data association has multiple solutions)
Track merging: high correlation between tracks
Multi hypothesis tracking

3 tracks (240, 247, 285)
only 1 actual observation (285)
Rectification
Unified Views
DENSITA’ MEDIE E MASSIME DEL TRAFFICO

02/11/2006 ore 11,30
Numero Totale Imbarcazioni in Canal Grande: 121

<table>
<thead>
<tr>
<th>Tratto</th>
<th>Da</th>
<th>A</th>
<th>Densità media</th>
<th>Densità max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ponte Libertà</td>
<td>Scomensera</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Scomensera</td>
<td>Ponte Calatrava</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>Ponte Calatrava</td>
<td>Ferrovia</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Ferrovia</td>
<td>Cannaregio</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>Cannaregio</td>
<td>Santa Fosca</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>Santa Fosca</td>
<td>Ca D’oro</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Ca D’oro</td>
<td>Rialto</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>Rialto</td>
<td>San Silvestro</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>S.Silvestro</td>
<td>San Tomà</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>San Tomà</td>
<td>Ca’ Rezzonico</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>11</td>
<td>Ca’ Rezzonico</td>
<td>Accademia</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>Accademia</td>
<td>Salute</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>13</td>
<td>Salute</td>
<td>Bacino S.Maro</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>
Example
Background subtraction problems

Background subtraction is a fast and effective technique, but it presents a series of problems:

- How to compute a correct background?
- How to manage gradual and sudden illumination changes?
- How to manage high-frequencies background objects (such as tree branches, sea waves, and similar)?
- How to manage partially occluded objects (for example, two cars crossing in a street)?
How to manage partially occluded objects?

UNDER SEGMENTATION

1 blob 2 boats

OVER SEGMENTATION

3 blobs 2 boats
Proposed solution

Background Subtraction
+
Optical Flow
+
Rek-means

[Bloisi and Iocchi 2008]
Optical flow

We use a sparse iterative version of Lucas-Kanade optical flow in pyramids [Bouget 2000]. It calculates coordinates of the feature points on the current video frame given their coordinates on the previous frame. The function finds the coordinates with sub-pixel accuracy. Every feature point is classified into one of four principal directions.
Under segm. solution using OF

Solution (non trivial): clustering the OF sparse map

$k=?$

noise

outliers
Over segmentation example.
The system detects 3 boats while the real number of boats is 2.

Solution.
Discard blobs associated with zero (or a negligible number) Optical Flow points.
K-means clustering

- true centroids
- k-means centroids

random initialization
K-means limitations

- the user may know the exact number of clusters ($k$) beforehand

- k-means is not guaranteed to return a global optimum: The quality of the final solution depends largely on the initial set of centroids, and may, in practice, be much poorer than the global optimum.

- If we choose the $k$ initial clusters at random, k-means can converge to the wrong answer (in the sense that a different and optimal solution to the minimization function exists).
Rek-means features

1. it provides better results in clustering data coming from different Gaussian distributions;
2. it does not require to specify $k$ beforehand;
3. it maintains real-time performance.
Improved clustering using Rek-means
Rek-means algorithm

1. *(over-clustering step)* Compute k-means with $k = n/4$ where $n$ is the number of points to cluster.

2. *(cutting step)* Discard every centroid $c_i$ having $\dim(c_i) \leq 2$.

3. *(discretizing step)* For each of the remaining centroids $c_j$, consider a rectangle $\text{RECT}(c_j)$ containing all the points belonging to $c_j$.

4. *(associating step)* If $\text{dist}(\text{RECT}(c_i), \text{RECT}(c_j)) \leq d$ then merge clusters $c_i$ and $c_j$.

5. *(validating step)* When the associating step is terminated, apply a validating test for each found cluster.
Rek-means steps

a) Initial data set
b) Over-clustering
c) Rectangles merging
d) Final centroids
Rectangle distance
Validating step

Let $S$ be the set of centroids found by the associating algorithm. For every $c_i \in S$:

1. Find $c_i^1$ and $c_i^2$ (applying k-means with $k = 2$), project the points belonging to $c_i$ onto the line connecting $c_i^1$ and $c_i^2$, translate, normalize and do the AD test.
2. If the test is successful then $c_i$ is a true centroid, otherwise add $c_i^1$ and $c_i^2$ to $S$.
3. Discard $c_i$ from $S$.
Repeat until $S$ is empty.
Anderson Darling test

The test we use is based on the Anderson-Darling statistic. This one-dimensional test has been shown empirically to be the most powerful normality test that is based on the empirical cumulative distribution function (ECDF). Given a list of values \( x_i \) that have been converted to mean 0 and variance 1, let \( x_{(i)} \) be the \( i \)th ordered value. Let \( z_i = F(x_{(i)}) \), where \( F \) is the \( N(0, 1) \) cumulative distribution function. Then the statistic is

\[
A^2(Z) = -\frac{1}{n} \sum_{i=1}^{n} (2i - 1) \left[ \log(z_i) + \log(1 - z_{n+1-i}) \right] - n
\]

if \( A^2 > CV \) the test is negative
else if \( A^2 < CV \) the test is positive
where \( CV \) is the critical value
Validating step example

The associating step finds 1 cluster

The validating step corrects the error and finds 2 clusters

C1 C2 C

K1 K K2
Rek-means real data example

Current frame and foreground image

Rek-means output

$A^2 = 0.409376$
$CV = 0.706069$

$A^2 = 4.702802$
$CV = 0.747689$

CENTROID REJECTED
Rek-means evaluation

100 images (with resolution 2000x2000)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Percentage of Correct Output</th>
<th>CPU Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rek-means</td>
<td>100</td>
<td>29</td>
</tr>
<tr>
<td>k-means</td>
<td>59</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

Time complexity \( O(n \log n) \)

1. over clustering step \( O(kn) \)
2. cutting step \( O(1) \)
3. discretizing step \( O(1) \)
4. associating step \( O(n \log n) \)
5. validating step \( O(n \log n) \)
ARGOS experimental evaluation

<table>
<thead>
<tr>
<th>Test</th>
<th>Modality</th>
<th>Ground Truth</th>
<th>Time (hours)</th>
<th>Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detection &amp; Tracking</td>
<td>On-Line</td>
<td>Human</td>
<td>28.5</td>
<td>2.2 \times 10^7</td>
</tr>
<tr>
<td>Count</td>
<td>Rec. On-Line</td>
<td>Human</td>
<td>2</td>
<td>1.3 \times 10^5</td>
</tr>
<tr>
<td>Position &amp; Velocity</td>
<td>Off-Line</td>
<td>GPS</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*on-line*, evaluation is performed during the actual operation of the system;
*recorded on-line* evaluation is performed on a video recording the output of the system running on-line;
*off-line* evaluation is performed on the system running off-line on recorded input videos.
### Online evaluation

<table>
<thead>
<tr>
<th>Day</th>
<th>Duration (min.)</th>
<th>Meteo</th>
<th>FN</th>
<th>FP-R</th>
<th>FP-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>07/01/2008</td>
<td>130 Cloud/Fog</td>
<td>0.062</td>
<td>0.215</td>
<td>0.531</td>
</tr>
<tr>
<td>2</td>
<td>08/01/2008</td>
<td>130 Sun/Cloud</td>
<td>0.038</td>
<td>0.192</td>
<td>0.431</td>
</tr>
<tr>
<td>3</td>
<td>15/01/2008</td>
<td>130 Sun/Cloud</td>
<td>0.031</td>
<td>0.154</td>
<td>0.323</td>
</tr>
<tr>
<td>4</td>
<td>31/01/2008</td>
<td>120 Cloud</td>
<td>0.075</td>
<td>0.158</td>
<td>0.400</td>
</tr>
<tr>
<td>5</td>
<td>01/02/2008</td>
<td>120 Cloud/Fog</td>
<td>0.000</td>
<td>0.150</td>
<td>0.392</td>
</tr>
<tr>
<td>6</td>
<td>04/02/2008</td>
<td>120 Cloud/Rain</td>
<td>0.000</td>
<td>0.200</td>
<td>0.342</td>
</tr>
<tr>
<td>7</td>
<td>05/02/2008</td>
<td>120 Sun/Cloud</td>
<td>0.000</td>
<td>0.225</td>
<td>0.392</td>
</tr>
<tr>
<td>8</td>
<td>06/02/2008</td>
<td>120 Sun/Cloud</td>
<td>0.017</td>
<td>0.200</td>
<td>0.333</td>
</tr>
<tr>
<td>9</td>
<td>07/02/2008</td>
<td>120 Sun</td>
<td>0.033</td>
<td>0.167</td>
<td>0.442</td>
</tr>
<tr>
<td>10</td>
<td>11/02/2008</td>
<td>120 Sun</td>
<td>0.017</td>
<td>0.292</td>
<td>0.375</td>
</tr>
<tr>
<td>11</td>
<td>12/02/2008</td>
<td>120 Sun</td>
<td>0.025</td>
<td>0.158</td>
<td>0.383</td>
</tr>
<tr>
<td>12</td>
<td>13/02/2008</td>
<td>120 Sun</td>
<td>0.033</td>
<td>0.267</td>
<td>0.367</td>
</tr>
<tr>
<td>13</td>
<td>14/02/2008</td>
<td>120 Sun</td>
<td>0.067</td>
<td>0.108</td>
<td>0.300</td>
</tr>
<tr>
<td>14</td>
<td>15/02/2008</td>
<td>120 Sun</td>
<td>0.000</td>
<td>0.150</td>
<td>0.250</td>
</tr>
<tr>
<td>Avg.</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.028</td>
<td>0.188</td>
</tr>
</tbody>
</table>

**FN:** False negatives, i.e. boats not tracked  
**FP-R:** False positives due to reflections (wrong track with a random direction)  
**FP-W:** False positives due to wakes (wrong track following the correct one)
COUNTING EVALUATION TEST

A virtual line has been put across the Canal in the field of view of a survey cell, the number of boats passing this line has been counted automatically by the system $n_{Sys}$, and the same value is manually calculated by visually inspection $n$, the average percentage error is then computed as

$$\varepsilon = \left| \frac{n_{Sys} - n}{n} \right|$$

An additional error measure is calculated by considering the probability of making an error in counting a single boat passing the line

$$P(e) = \frac{1}{n} \sum_{t=0}^{n} \delta(\hat{f}_t - f_t)$$

where $\delta(\cdot)$ is 0 when the argument is 0 and 1 otherwise.
Counting evaluation (2)

<table>
<thead>
<tr>
<th>Video</th>
<th>n boats</th>
<th>FN</th>
<th>FP</th>
<th>count accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>20070928_1335_c09</td>
<td>47</td>
<td>0.11</td>
<td>0.04</td>
<td>93.6</td>
</tr>
<tr>
<td>20071030_1015_c07</td>
<td>37</td>
<td>0.05</td>
<td>0.03</td>
<td>97.3</td>
</tr>
<tr>
<td>20070928_1335_c10</td>
<td>36</td>
<td>0.11</td>
<td>0.06</td>
<td>94.4</td>
</tr>
<tr>
<td>20071031_1000_c03</td>
<td>35</td>
<td>0.17</td>
<td>0.03</td>
<td>85.7</td>
</tr>
<tr>
<td>20071030_1035_c04</td>
<td>35</td>
<td>0.06</td>
<td>0.06</td>
<td>100.0</td>
</tr>
<tr>
<td>20071030_1025_c05</td>
<td>33</td>
<td>0.03</td>
<td>0.00</td>
<td>97.0</td>
</tr>
<tr>
<td>20071214_0939_c08</td>
<td>31</td>
<td>0.10</td>
<td>0.00</td>
<td>90.3</td>
</tr>
<tr>
<td>20071030_1355_c12</td>
<td>29</td>
<td>0.03</td>
<td>0.03</td>
<td>100.0</td>
</tr>
<tr>
<td>20071210_1300_c06</td>
<td>17</td>
<td>0.12</td>
<td>0.00</td>
<td>88.2</td>
</tr>
<tr>
<td>20071213_1130_c03</td>
<td>17</td>
<td>0.00</td>
<td>0.06</td>
<td>94.1</td>
</tr>
<tr>
<td>20071030_1335_c10</td>
<td>14</td>
<td>0.07</td>
<td>0.07</td>
<td>100.0</td>
</tr>
<tr>
<td>20071210_1145_c01</td>
<td>9</td>
<td>0.11</td>
<td>0.00</td>
<td>88.9</td>
</tr>
<tr>
<td>Avg.</td>
<td>28.3</td>
<td>0.08</td>
<td>0.03</td>
<td>94.1</td>
</tr>
</tbody>
</table>
New features

• It is crucial to know who does what
  - Behavior analysis based on the identity of boats observed is very difficult through Computer Vision only (plate recognition impossible with occlusion)
  - Add sensors such as GPS or Radio Frequency Identifiers (RFID)
References


Homework

Generate a data set from a distribution formed by a mixture of $k$ Gaussians with the same circular variance (i.e., $\Sigma = \text{diag} \{ \sigma, \sigma \}$) and with $|\mu_i - \mu_j| > 3\sigma$ for each pair $i,j$

Apply $k$-means on this data (with the correct value of $k$) several times starting from different initial values and compute the percentage of correct clustering.
Data clustering for image segmentation

Dipartimento di Informatica e Sistemistica
Antonio Ruberti

Sapienza
Università di Roma

Domenico Bloisi
bloisi@dis.uniroma1.it