An Object-Based Comparative Methodology for Motion Detection based on the F-measure

Neda Lazarevic-McManus
J.R. Renno, D. Makris, G. Jones

Digital Imaging Research Centre
Kingston University
Efforts to establish appropriate evaluation methods, datasets, groundtruth.

Metrics chaos – difficult to establish an accepted method for ranking the competing algorithms, particularly for end-users, technology integrators and governmental agencies.

What is required is a methodology capable of generating a meaningful scalar measure of performance and supporting the optimisation of algorithm parameters.
**Metric Chaos**

- **Metrics for object detection evaluation:**
  
  Villegas *WIAMIS '99*, Correia *CIP '00*, Di Stefano *ICIAP '01*, Erdem *ICIP '01*, Cavallaro *ICIP '02*, Jaynes *PETS '02*, Mriano *ICIP '02*, Black *PETS '03*, Villegas *IEEE IP '04*, Nascimento *PETS '04*, Gelasca *CVPR '04*, Schlogl *ICPR '04*, Aguilera *PETS '05*, Hall *PETS '05* …

- **Frameworks and initiatives - datasets, groundtruth, tools, metrics…**

  ETISEO, PETS, CAVIAR, ViPER, CLEAR, VIVID, AVITRACK, VITAL, VAREAAE, CANTATA, REASON…

- **Approaches:**

  pixel vs. object, temporal integration, local vs. global …
Overview

Aim to highlight the complexity and range of the issues which underpin the design of good motion detection evaluation methodology.

- The distinction and relative merits of pixel-based vs. object-based metrics
- Making explicit evaluation parameters, their effect on evaluation
- The impact of defining the end-user application
- Selecting appropriate algorithm and evaluation parameters values using F-measure
- Comparing the performance of different algorithms on two datasets
Typical Evaluation System

Metrics
- Pixel based - fails to provide an evaluation of individual objects
- Object based - subjective interpretation of vectors of different metrics

COMPARE

System parameters

OPTIMISE

Algorithm Parameters

Evaluation Parameters

Evaluation Method

Ground Truth

input

segmentation

metrics
Methodology for Interpretation of Metrics

ROC curves
- Objective and easy to interpret
- Cost-ratio defines different end-user applications
- non-existence of true negatives in object based evaluation

Ideal properties of an alternative methodology:
- a ‘ROC-like’ scalar to optimise parameters
- exploits the idea of detection evaluation in the context of the end-user application
- avoids the true negative objects
In the previous pixel-based ROC methodology, we defined two application scenarios with different misclassification cost ratios:

- Evaluate motion detection not as a stand-alone module but in the context of the surveillance system.
  \[ \frac{C_{FP}}{C_{FN}} = 0.1 \]
- A terrorist bomber scenario
Evaluation Parameters

Object-based approach requires the correspondence of GT and detected blobs, and hence selection of evaluation parameters.
Varying Evaluation Parameters

Ticket Fraud Application Scenario

Find appropriate values for evaluation parameters for each end-user application scenario!
1. Select the evaluation parameters for each application scenario (ideally over the space of all algorithms, all combinations of algorithm parameters, and all combinations of evaluation parameters).

2. Optimise each algorithm for best performance.

3. Compare the performance of the optimised algorithms for a given application scenario.
F-Measure Methodology

F-measure method

- Information Retrieval (C.J. van Rijsbergen, 1979)
- Object Boundary Detection (Martin et al, 2004)
- Not used in Visual Surveillance

**Recall** – the fraction of objects that is detected

\[
R = \frac{N_{TP}}{N_{TP} + N_{FN}}
\]

**Precision** – the fraction of detected blobs that is true positive

\[
P = \frac{N_{TP}}{N_{TP} + N_{FP}}
\]
Definition of the F-Measure

F-measure – the weighted harmonic mean of R and P

\[ F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha) \frac{1}{R}} \]

- Parameter \( \alpha \) controls the relative importance of P and R for a given scenario.
- Maximise F for the optimal performance for given \( \alpha \).
Optimal Point Selection

The graph shows precision and recall curves for different recall values.

- **F1, F2, F3, F4** represent different performance metrics.
- **α_{high}, F_{max}** indicate optimal point selection.
- The curve F_{curve} highlights the optimal point for highest F_max.
Parameterising the Application Scenarios

\[
F = \frac{1}{\alpha \frac{1}{P} + (1-\alpha) \frac{1}{R}}
\]

\(\alpha\) reflects the relative importance of precision and recall for a specific surveillance task – *application scenario*

\[
R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad P = \frac{N_{TP}}{N_{TP} + N_{FP}}
\]

**Ticket Fraud Scenario:**

- P essential \(\rightarrow\) \(\alpha_{TF}\) high

**Evidence Gathering Scenario:**

- R essential \(\rightarrow\) \(\alpha_{EG}\) low
Optimal Evaluation Parameters

PETS-01-cam1

Kingston Carpark

\( \alpha_{TF} = 0.8 \)

\( \alpha_{EG} = 0.04 \)
Optimal Algorithm Parameters

Kingston Carpark dataset

PETS-01-cam1 dataset
### Comparison using F-measure

<table>
<thead>
<tr>
<th></th>
<th>PETS-2001-camera1</th>
<th>Kingston Carpark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evidence Gathering</td>
<td>Ticket Fraud</td>
</tr>
<tr>
<td>Stauffer and Grimson</td>
<td>0.875</td>
<td>0.94</td>
</tr>
<tr>
<td>Motion distillation</td>
<td>0.867</td>
<td>0.937</td>
</tr>
<tr>
<td>Renno et al</td>
<td>0.85</td>
<td>0.807</td>
</tr>
</tbody>
</table>

**NOTE:** The relationship between achieved levels of the F-measure for the standardised application scenarios can indicate the level of difficulty of a dataset!
Where is it best to evaluate motion detection?
Thank You