Story comparison for estimating field of view overlap in a video collection

ICPR 2020 - Oral Presentation

January 11th 2021

Institut de Recherche en Informatique de Toulouse (IRIT)

Thierry Malon (speaker)
Sylvie Chambon
Alain Crouzil
Vincent Charvillat
Overview of the proposed approach

**Goal:** automatically finding videos of a collection that have overlapping fields of view

**Hypothesis:** static cameras and temporally synchronised videos, no other metadata
Detection step

**Detection:** using 3 different existing CNN approaches (SSD [Liu, 2016], Mask-RCNN [He, 2017], YOLOv3 [Redmon, 2018])

**Descriptors:** using 3 different existing appearance descriptors (HOG [Dalal, 2005], ColorNames [Yang, 2014], Latent representation of Resnet18 [He, 2016])
Definition of a region story

**Story of a region:** list of objects (category + appearance descriptor) detected at regular time steps in the region
Comparison between stories

**Distance between stories:** proportion of objects from each story which have a correspondent object in the spatiotemporal neighborhood of the other story.
Multiresolution

To avoid comparing all possible region pairs, we compare regions at different scales and keep on comparing at a thinner scale if the distance if neither in the accepted link interval (in green) nor in the rejected link interval (in red)

\[ d(S_R, S_{R'}) = 0.809 \]

(a) \( s = 1 \)

\[ d(S_R, S_{R'}) = 0.992 \]

(b) \( s = 2 \)

\[ d(S_R, S_{R'}) = 0.573 \]

(c) \( s = 2 \)

\[ d(S_R, S_{R'}) = 0.183 \]

(d) \( s = 3 \)
Experiments

Datasets: multiview datasets (ToCaDa, EPFL, MEVA, Youtube videos)
Experiments

**Evaluation:** F1 score of the overlap links between videos

True Positive (TP)          False Negative (FN)          True Negative (TN)

Precision = TP / (TP + TN)  Recall = TP / (TP + FN)

F1 score = 2 × Precision × Recall / (Precision + Recall)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SSD</th>
<th>Mask-RCNN</th>
<th>YOLOv3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>base</td>
<td>ctg</td>
<td>ctgcen</td>
</tr>
<tr>
<td>Live Cameras</td>
<td>19</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>MEVA</td>
<td>28</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>EPFL</td>
<td>56</td>
<td>56</td>
<td>64</td>
</tr>
<tr>
<td>ToCaDa</td>
<td>25</td>
<td>29</td>
<td>34</td>
</tr>
<tr>
<td>All real videos</td>
<td>16</td>
<td>19</td>
<td>21</td>
</tr>
</tbody>
</table>
Experiments

**Overlap graph:** overlap links found on a subset of the dataset
Questions

\[ S^0_R \rightarrow^3 = \]

\[ d(S_R, S_{R'}) = 0.809 \quad \text{(a)} \quad s = 1 \]
\[ d(S_R, S_{R'}) = 0.992 \quad \text{(b)} \quad s = 2 \]
\[ d(S_R, S_{R'}) = 0.573 \quad \text{(c)} \quad s = 2 \]
\[ d(S_R, S_{R'}) = 0.183 \quad \text{(d)} \quad s = 3 \]
In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)
Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)
Context
In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)
Context

In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources)
In the case of a particular event such as a competition, a demonstration, or an attack, a huge amount of videos is available (from multiple sources).
Problem statement

From a collection of videos, how to automatically