SVM-based Shot Type Classification of Movie Content

I. Tsingalis, N. Vretos, N. Nikolaidis, I. Pitas

Department of Informatics
Aristotle University of Thessaloniki, Greece
A method to classify movie shots with respect to their type is proposed in this paper.

- Problem definition
- Method description
- Experimental results
What is a Shot?

- Shot: a continuous sequence of frames that originates from a single camera take.
Classify a video shot into one of the shot types defined in filmmaking:

- eXtreme Close Up (XCU)
- Close Up (CU)
- Medium Close Up (MCU)
- Medium Shot (MS)
- Medium Long Shot (MLS)
- Long Shot (LS)
- eXtreme Long Shot (XLS)

Defined on the basis of how much of the subject (actor) and its surrounding area is visible.
The subject is small in the overall image. Usually used as an establishing shot and occurs in outdoor scenes.

Almost the entire body of the actor is visible (full body shot).
Shot Types

The actor’s body is usually framed from the knees and up.

The human is visible down to the waist level.
Shot Types

The human body is framed from the elbow and above.

Head and shoulders are visible.

A part of the actor's face is visible. The frame contains no information for the background.
Shot Types in Sports

**In-Field Long Shot**

**In-Field Close Up**

**Audience (Out-of-field) Shot:** Only the audience is visible.

**Long shot:** The global view of the field and sometimes a part of the stadium is displayed.
- Automatic analysis of film structure
- Content annotation & management
- Support for advanced queries
Shot Classification Approaches

Actor or object

Feature extraction

Color distribution, saliency maps, motion, head geometry, e.t.c.

Classification

PCA or LDA if needed
Method Outline

Actor -> Feature extraction -> Face dimensions

SVM Classification
Feature Selection

- Face dimensions (width, height of the facial bounding box, BB) in relation to frame dimensions.

- 2D feature vector: \( \mathbf{v} = (\frac{h}{H}, \frac{w}{W}) \)
  
  - \( \frac{h}{H} \) = face to frame height ratio
  - \( \frac{w}{W} \) = face to frame width ratio
Values of Used Features

CU, MCU, LS, ... shot types

h/H vs. w/W

- CU
- MCU
- XLS
- MS
- MLS
- LS
- XCU
Support Vector Machines (SVM) were used to classify each frame, based on the computed features.

- Number of SVM classes: 7
- Kernel type: RBF
- SVM parameterization
  - Tuning and kernel parameters C and γ were selected, using grid search along with cross validation.
Step 1: Manually mark the first face BB or use a face detector.

Step 2: Evaluate the face BB in all frames by applying a tracking algorithm and calculate the feature vectors.

Step 3: Classify each frame using the trained SVM classifier.
• **Step 1**: Frame based classification (as previous).
• **Step 2**: Apply majority voting: the entire shot is assigned the label assigned to the majority of its frames.
Database and Experimental Results

- **Database**
  - 173 shots, variety of genres
  - 12178 frames
  - Manual labeling to create the ground truth.

MS

CU

LS
5-fold cross validation:
- 1/5 of the data are used as a test set
- The SVM is trained on the remaining 4/5.
- The procedure is repeated five times.

Performance metric: \( A = \frac{N_{CC}}{N_{GT}} \)
- \( N_{CC} \): frames/shots correctly classified
- \( N_{GT} \): total number of the frames/shots
Database and Experimental Results

- Per-frame shot type classification accuracy
  - 98.44%
- Per-shot shot type classification accuracy
  - 100%
# Confusion Matrix of Frame Based Results

<table>
<thead>
<tr>
<th></th>
<th>XCU</th>
<th>CU</th>
<th>MCU</th>
<th>MS</th>
<th>MLS</th>
<th>LS</th>
<th>XLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>XCU</td>
<td>0.97</td>
<td>0.03</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU</td>
<td>0.004</td>
<td>0.99</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCU</td>
<td></td>
<td>0.003</td>
<td>0.98</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td></td>
<td></td>
<td>0.019</td>
<td>0.976</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLS</td>
<td></td>
<td></td>
<td></td>
<td>0.023</td>
<td>0.97</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.004</td>
<td>0.996</td>
<td></td>
</tr>
<tr>
<td>XLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Comparison between SVM and rule based method [1]

<table>
<thead>
<tr>
<th></th>
<th>Rule based</th>
<th>SVM based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame based Ac.(%)</td>
<td>90.3%</td>
<td>98.44%</td>
</tr>
<tr>
<td>Shot based Ac.(%)</td>
<td>91.4%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Conclusions

- A method to classify movie shots according to their type was proposed.
- The method utilizes features related to face dimensions and SVMs.
- Despite its simplicity, the method achieves good results.
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n 287674 (3DTVS).

http://www.3dtvs-project.eu/