Singlets
Multi-resolution Motion Singularities for Soccer Video Abstraction

Katy Blanc, Diane Lingrand, Frederic Precioso
I3S Laboratory, Sophia Antipolis, France
Overview

VIDEO & MOTION

SINGULARITIES & SINGLETS

SOCCER SALIENT MOMENTS
Video Analysis

- Burst of video content production
- New sources of videos
  - Big databases: Youtube 8M [1]
- Much analyzed types: meeting/conferences, movies, news and sports

- Diverse applications:
  - Browsing in database, automatic video surveillance, driverless car, ...

- Exponential amount of information:
  - A match of soccer of HDTV → 324000 images of 1920×1080 pixels each

- Collaboration with Wildmoka, themselves in collaboration with l’INA and BeIN.
Related works: Video description

Modelling human motion: [2]

Handcrafted features: Stip [3], iDT [4]

Deep learning representations [5]
Related works: sport abstraction

Clue Detection to detect highlights: ground color, jersey color, shot segmentation and view classification


- Goals, attacks and other events using logo and score appearances and goal mouth position [7] Zawbaa et al.

- Face and skin detection, whistle detector and user specifications [8] Raventos et al.
Overview

VIDEO & MOTION

SINGULARITIES & SINGLETS

SOCCER SALIENT MOMENTS
Inspiration: fluid movement

Inspired from the work of Druon et al. [9] and the further work of Kihl et al. [10]
Optical Flow Approximation

Optical Flow = discrete bivariavle vector field

\[ F : \Omega \rightarrow \mathbb{R}^2 \]
\[ (x_1, x_2) \rightarrow (U(x_1, x_2), V(x_1, x_2)) \]
with \( \Omega = [-1, 1]^2 \)

Polynomial subspace and the Legendre Basis

\[ P_{K,L}(x_1, x_2) = \sum_{k=0}^{K} \sum_{l=0}^{L} \gamma_{k,l} x_1^k x_2^l \]
with \( K + L < D \)

Projection on the Legendre Basis

\[ \begin{cases} U = u_{0,0} P_{0,0} + u_{0,1} P_{0,1} + u_{1,0} P_{1,0} \\ V = v_{0,0} P_{0,0} + v_{0,1} P_{0,1} + v_{1,0} P_{1,0} \end{cases} \]
Polynomial projection

Original Flow $F$

Flow $U$

Flow $V$

Projection

\[
\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1.38 \\ -0.01 \end{pmatrix} + \begin{pmatrix} 0 \\ 4.53 \end{pmatrix} + 0.007 + 0.01 + 0.023 + 0 + 0 + 0 + 0 - 4.47 - 7E^{-5}
\]

\[
\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 0 & 0.02 \\ 3.91 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0.69 \\ -0.006 \end{pmatrix}
\]
Approximation and coefficient analysis

From a simple analysis example to production, scenarization, event sementization

\[
(U) = \begin{pmatrix} 1.38 \\ 0 \end{pmatrix}
\]

Coefficient value

Frame number

counterattack

- $u_{0,0}$ horizontal global displacement
- $v_{0,0}$ vertical global displacement
- $\int u_{0,0}$ horizontal global position
- $\int v_{0,0}$ vertical global position
First projection on the Legendre basis

\[
\begin{align*}
U &= u_{0,0}P_{0,0} + u_{0,1}P_{0,1} + u_{1,0}P_{1,0} \\
V &= v_{0,0}P_{0,0} + v_{0,1}P_{0,1} + v_{1,0}P_{1,0}
\end{align*}
\]

Then on the canonical basis

\[
\begin{pmatrix} U \\ V \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b \quad \text{with} \quad A \in M_{2,2} \quad \text{and} \quad b \in \mathbb{R}^2
\]

6 types of singularities

\[
\Delta(A) = \text{tr}(A)^2 - 4.\det(A) ,
\]

\[
\lambda_1 \text{ and } \lambda_2 \text{ the Eigenvalues of } A
\]
Singularities extraction

From a multi-resolution analysis of the optical flow ...

\[
\begin{bmatrix}
U \\
V
\end{bmatrix} = A \begin{bmatrix}
\chi_1 \\
\chi_2
\end{bmatrix} + b
\]
Singlets: match singularities during time

From singularity on optical flow frame to tracks of singularities: Singlets

\[ V(d(A, A'), \|x - x'\|_2) = \|A - A'\|_F + \lambda \|x - x'\|_2 \]
Overview

VIDEO & MOTION

SINGULARITIES & SINGLETS

SOCCER SALIENT MOMENTS
Application on soccer abstraction

- Our database
- Zoom
- Slow Motion
- Global Excitement
- Soccer Salient Moment
Soccer Summarization: Our database

Lack of standard benchmark for comparison sake...

Germany vs Portugal: 4-0
Nigeria vs Argentina: 2-3
France vs Honduras: 3-0
Switzerland vs France: 2-5

© http://www.fifa.com/worldcup/matches/
Soccer Summarization: Our database

<table>
<thead>
<tr>
<th>Germany (GER)</th>
<th>Statistics</th>
<th>Portugal (POR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shots/Shots on goal</td>
<td>14/9</td>
</tr>
<tr>
<td>13/9</td>
<td>Fouls</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>Corner kicks</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Direct free kicks leading to a goal</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>Indirect free kicks leading to a goal</td>
<td>0</td>
</tr>
<tr>
<td>1/1</td>
<td>Penalty kicks/Converted</td>
<td>0/0</td>
</tr>
<tr>
<td>2</td>
<td>Offsides</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>Own goals</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>Cautions</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>Expulsion due to second caution</td>
<td>0</td>
</tr>
</tbody>
</table>

32' HUMMELS (Germany) scores!!

32' KROOS (Germany) swings in the corner.

<table>
<thead>
<tr>
<th>Match</th>
<th>FIFA ground-truth</th>
<th>Extended ground-truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany vs Portugal</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>Nigeria vs Argentine</td>
<td>51</td>
<td>35</td>
</tr>
<tr>
<td>France vs Honduras</td>
<td>54</td>
<td>32</td>
</tr>
<tr>
<td>Switzerland vs France</td>
<td>40</td>
<td>26</td>
</tr>
</tbody>
</table>
Zoom detection

Zoom motion is a pure star node singularity. Zoom combined with a translation is an improper node.

Conditions:
  - there is a singularity
  - it is a star or improper node: $\Delta(A) = 0$
  - time consistent: last for one second

Positives eigenvalues
  - $\rightarrow$ zoom out

Negatives eigenvalues
  - $\rightarrow$ zoom in

Determine the zoom direction center
Zoom detection

- \( \Delta(A) = 0 \iff |\Delta(A)| \leq 0.2 \)
- Threshold on an average \( |\Delta(A)| \) on a second set to 0.2
- Comparison

- Global motion estimation [6,11]

- Duan et al. method [12]:

Two histograms, one on magnitude one on angles to detect diagonal pattern (DL)

\[
\tau_k = \left( -1 \right)^{k+1} \alpha + \left[ \frac{k}{2} \right] \cdot 90^\circ, (-1)^k \alpha + \left[ \frac{k+1}{2} \right] \cdot 90^\circ \quad (6)
\]

\( k = 0 \ldots 7, \quad \alpha = 15^\circ. \)

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GME</td>
<td>3.68 %</td>
<td>68.4 %</td>
<td>19.79 %</td>
</tr>
<tr>
<td>Duan</td>
<td>8.92 %</td>
<td>50.62 %</td>
<td>75.06 %</td>
</tr>
<tr>
<td>ours</td>
<td>19.45 %</td>
<td>63.47 %</td>
<td>86.82 %</td>
</tr>
</tbody>
</table>

Table 2. Precision, recall and accuracy for zoom detection.
Slow Motion Detection

How to differentiate a fast motion that has been artificially slowed down from a slow motion?
Slow Motion Detection

Fast Motion

Slow Motion

Singlets

Singlets Length Histograms

SVM training

Slow Motion Classification

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>97.06 %</td>
<td>80.49 %</td>
<td>89.41 %</td>
</tr>
<tr>
<td>Test</td>
<td>76.32 %</td>
<td>87.88 %</td>
<td>79.36 %</td>
</tr>
</tbody>
</table>
Global excitement

Spatial histogram of 3x3 per frame
Sum spatial histograms on 10 frames
Threshold set on 1500 to select the most agitated moments
Soccer Saliant Moment

Within 30 seconds:
- at least two zoom direction changes
- an activity peaks higher than 1500 in the farthest view
- a slow motion replay in a close up view
Example of Summarization

- Zooms
- Agitation
- Slow Motion

<table>
<thead>
<tr>
<th>Match</th>
<th>FIFA ground-truth</th>
<th>Extended ground-truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany vs Portugal</td>
<td>80 %</td>
<td>88.9 %</td>
</tr>
<tr>
<td>Nigeria vs Argentine</td>
<td>53 %</td>
<td>77.2 %</td>
</tr>
<tr>
<td>France vs Honduras</td>
<td>53.7 %</td>
<td>90.7 %</td>
</tr>
<tr>
<td>Switzerland vs France</td>
<td>62.5 %</td>
<td>96.6 %</td>
</tr>
<tr>
<td>Mean</td>
<td>62.3 %</td>
<td>88.2 %</td>
</tr>
</tbody>
</table>

3rd International Workshop on Computer Vision in Sports (CVsports) at CVPR 2017
Extension to other sports

Set and Trained on Soccer and Tested on Handball

All hyperparameters, parameters and SVM were trained and set on soccer videos and we test our framework on 5 minutes of Qatar Handball World Championship final without any adjustment or retraining.
Optical Flow Approximation
Possibility to use:
- different basis
- different degree
- other space than the polynomial one

Singularity extraction
Vanishing point
6 types for the affine approximation
Motion description
Global distribution

Singlets
Track singularities
Length histograms
Compute a description like for IDT

Sport Video Abstraction
Zoom detection
Slow Motion Detection
Global excitement
Because there is not only soccer in life

Abstraction of concert

Facial Emotion

Action recognition

Writing On Board  Yo Yo  Baby Crawling  Blowing Candles  Body Weight Squats  Handstand Pushups
Pull ups  Push ups  Rock Climbing Indoor  Rope Climbing  Swing  Tai Chi
Band Marching  Haircut  Head Massage  Military Parade  Salsa Spin  Drumming
Playing Flute  Playing Guitar  Playing Piano  Playing Sitar  Playing Tabla  Playing Violin
THANK YOU
References


Optical Flow Extraction

Gunnar Farneback Method

Pixel Neighborhood Approximation

Approximation translation

\[ f(x) = x^T A x + b^T x + c \]

\[ f_2(x) = f_1(x - v) = (x - d)^T A_1 (x - d) + b_1^T (x - d) + c_1 = x^T A_2 x + b_2^T x + c_2 \]

Relations

\[ A_2 = A_1 \]
\[ b_2 = b_1 - 2A_1 v \]
\[ c_2 = v^T A_1 v - b_1^T v + c_1 \]

Speed vector estimation

\[ v = \frac{-1}{2} A_1^{-1} (b_2 - b_1) \]
Polynomial Space Projection

Scalar product

Legendre basis with \( w(x_1, x_2) = 1 \) and \( \Omega = [-1,1]^2 \)

\[
\langle F_1(x_1, x_2) \mid F_2(x_1, x_2) \rangle = \int_{\Omega} F_1(x_1, x_2) F_2(x_1, x_2) w(x_1, x_2) \, dx_1 dx_2
\]

- Polynomial degree \( D \Rightarrow nd = \frac{(D+1)(D+2)}{2} \) polynomials in the basis and so \( nd \) coefficients
- Flot optique blurring (decreasing with the degree)
Multi-Resolution Singularity

One sliding windows -> possibly one singularity

The same singularity can be at different size and approximately the same position: We keep the one with less angular deviation from the original flow.

\[ dev = \sum_{\omega \in \Omega} \frac{1}{2} |\sin(\theta(\omega) - \theta'(\omega))| \]

With \( \theta(\omega) \) the angle of the motion vector at the pixel position \( \omega \) for the original flow and \( \theta'(\omega) \) for the polynomial approximation.