



MULTI-SENSORED VISION FOR AUTONOMOUS PRODUCTION OF PERSONALIZED VIDEO SUMMARIES

www.apidis.org, New Technology Campus - IBC2010





UCL
Université
catholique
de Louvain



Centre
d'Innovació

22 Barcelona
Media


Queen Mary
University of London

Acic


ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

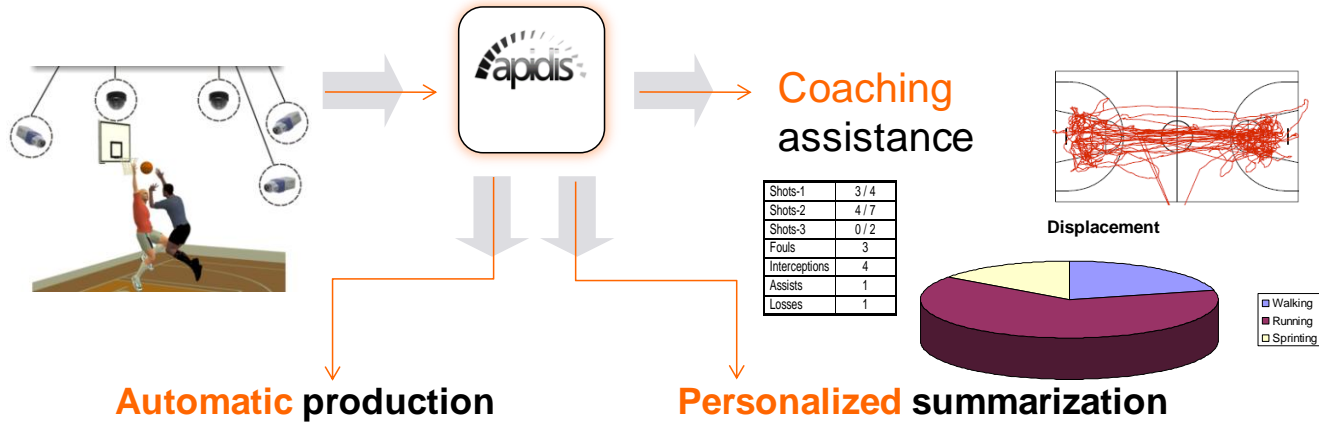

MEDIAPRO



APIDIS is co-funded by
the European Union



The big picture



Automatic production



Personalized summarization





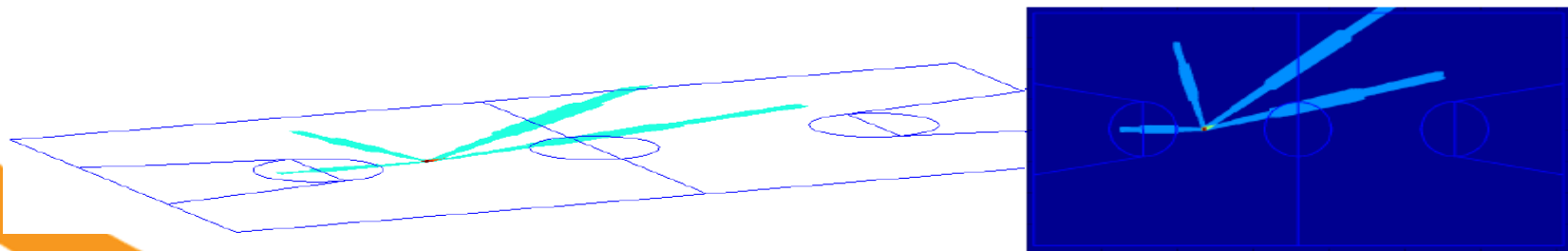
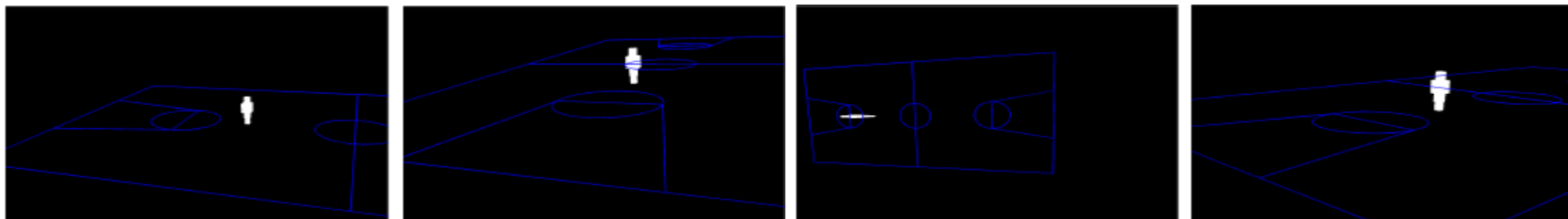
Outline

- Understanding the scene
 - Player detection&tracking
 - Event recognition
- Automatic production of personalized summaries
 - 1st step: production of edited video content through viewpoint selection;
 - 2nd step: personalized summarization through time segment selection and local story organization



Detection: foreground masks projections

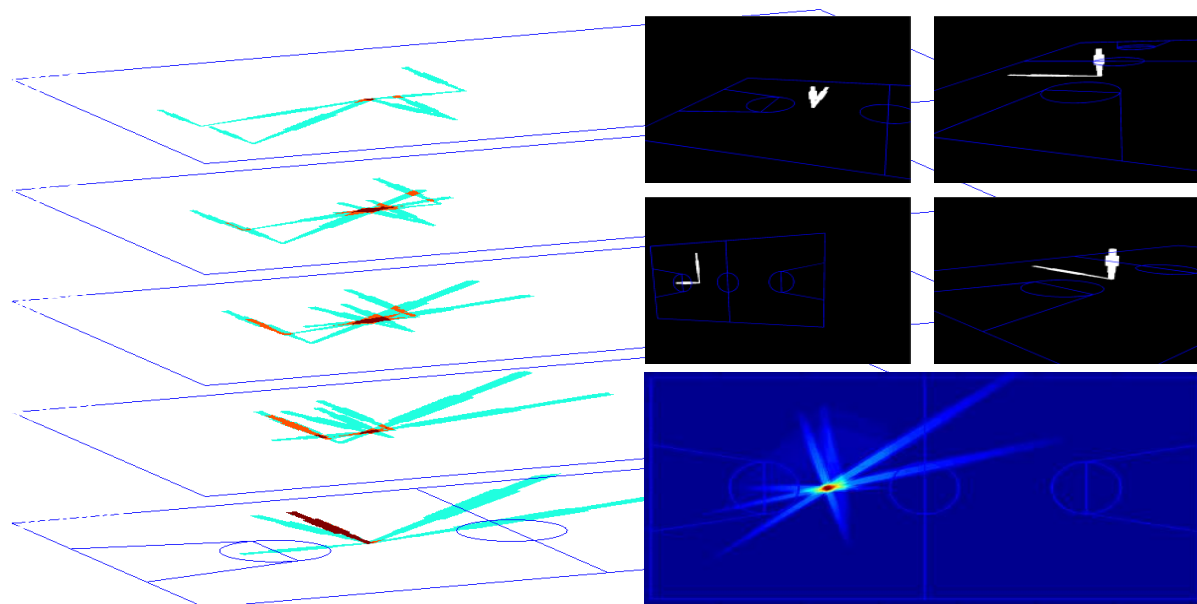
Main idea: projections of silhouettes from all views are consistent in player location.





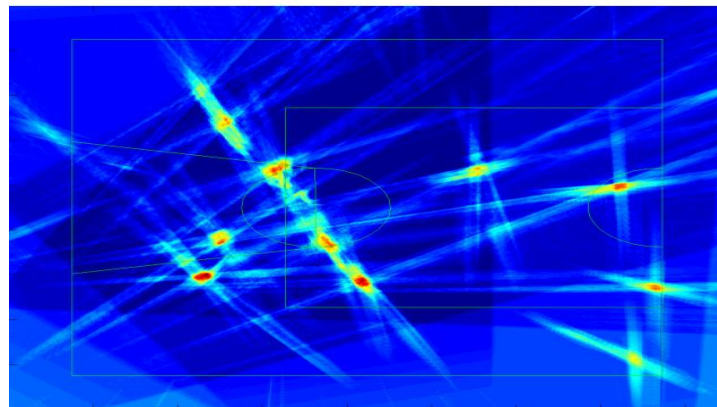
Aggregated occupancy mask

Main idea: projection consistency is maintained on planes parallel to ground plane.



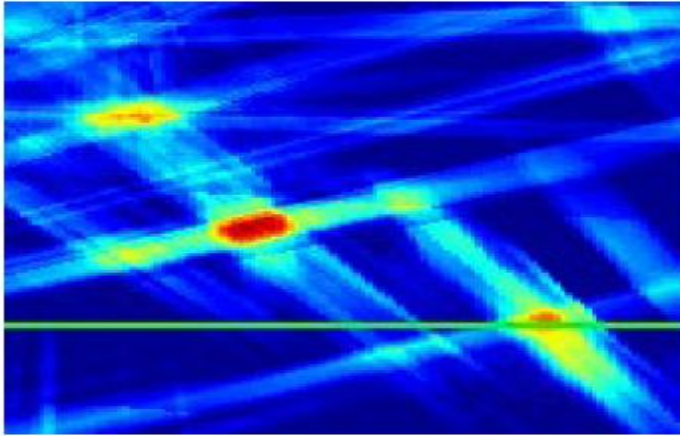


Aggregated ground occupancy mask (half of field)





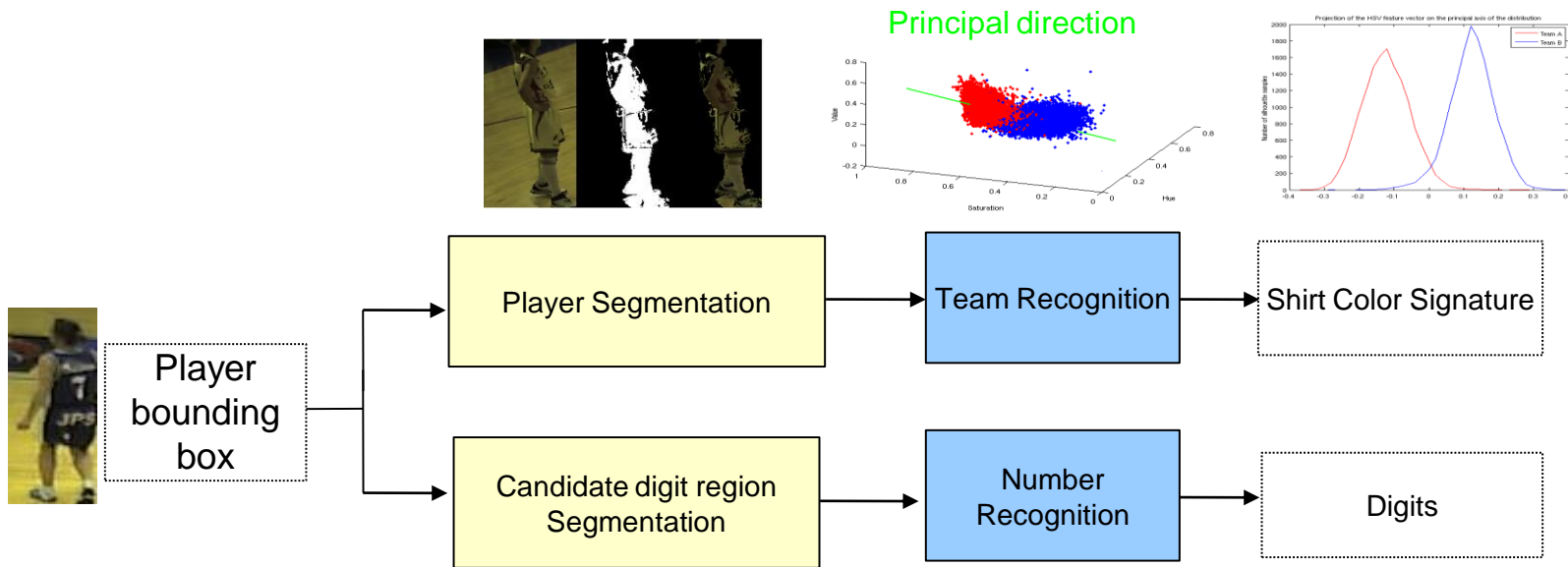
Detection through a geometrically-constrained greedy solution



- Performance improvement from 15% to 2% (errors)
- Real-time implementations (thanks to approximation)

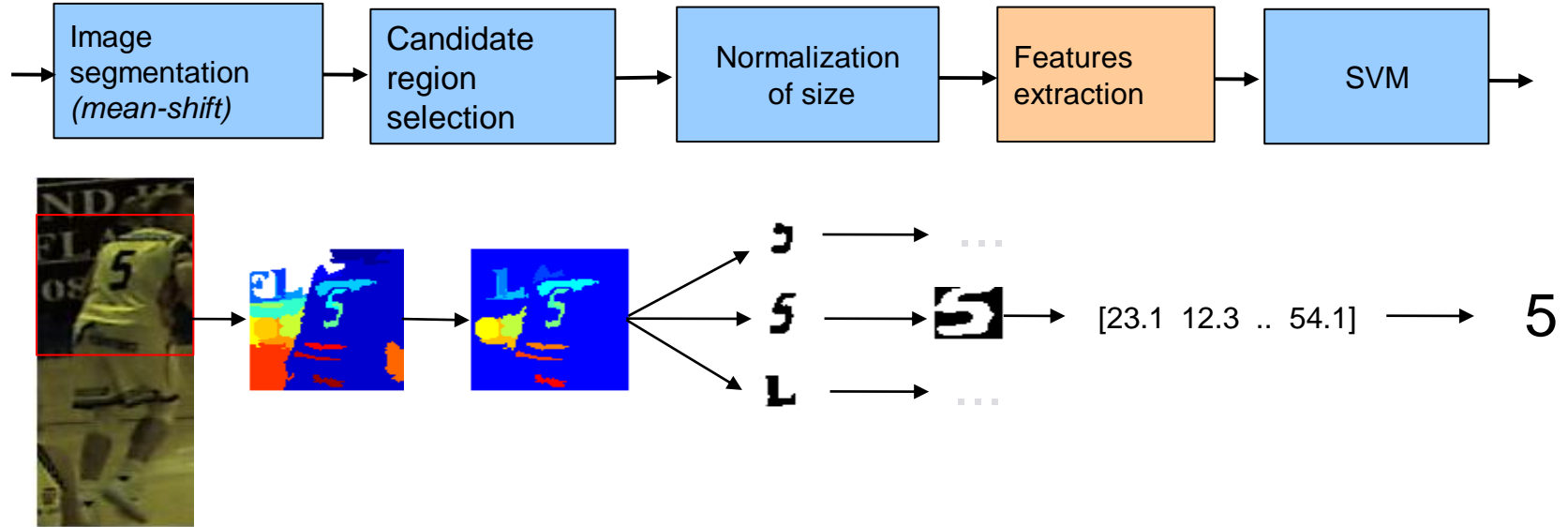


Characterizing the player appearance





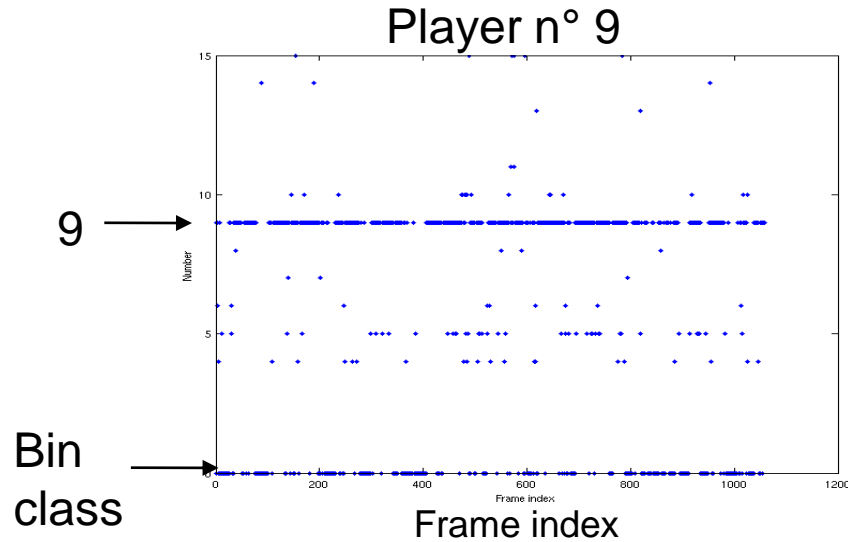
Digit recognition (1/2)





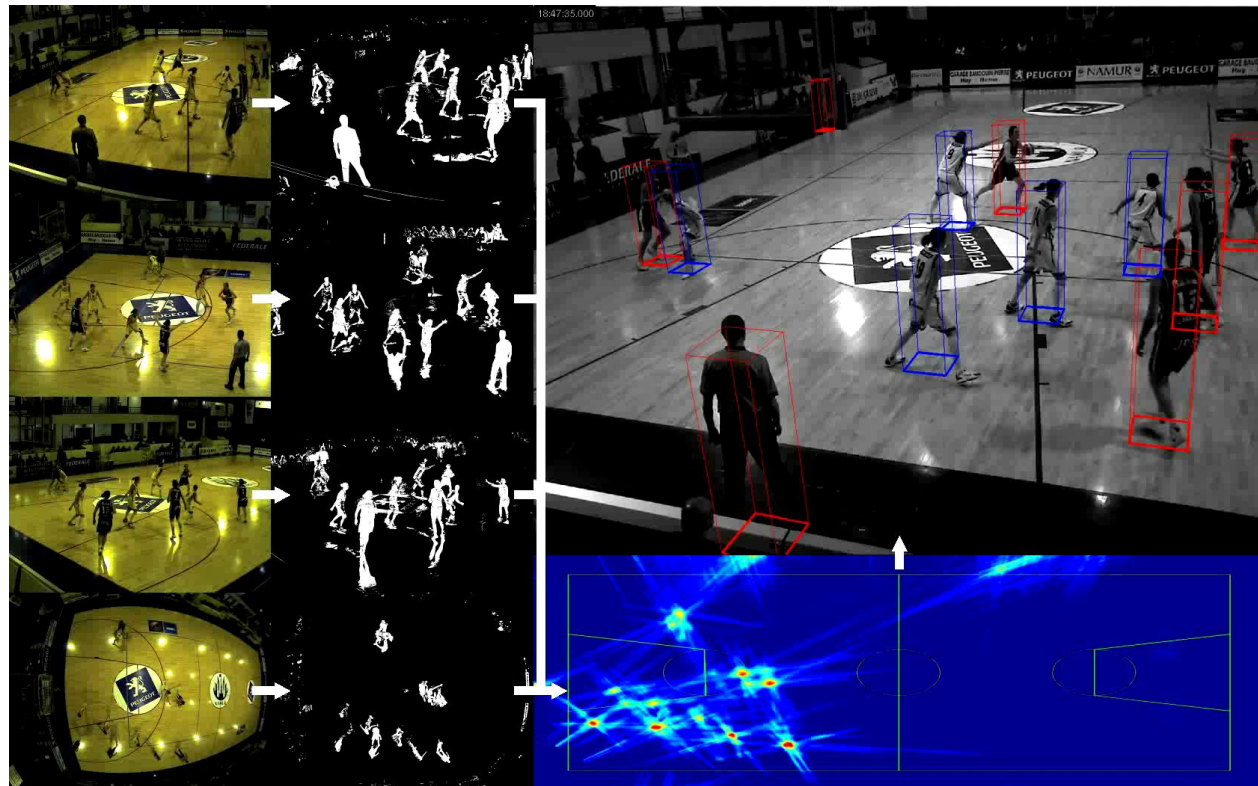
Digit recognition (2/2)

- shirt number not always visible
- need for a bin class





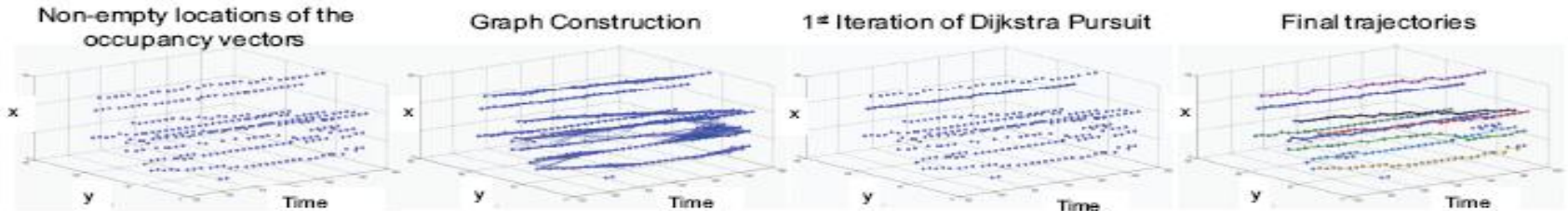
DEMO





Appearance-based multiview tracking

- Several approaches tested and compared
(see www.apidis.org for related papers from EPFL and QMUL)
- Graph-based tracking exploiting appearance features



Greedy and progressive selection of shortest path using Dijkstra (EPFL).



DEMO





Outline

- Understanding the scene
 - Player detection&tracking
 - Event recognition
- Automatic production of personalized summaries
 - 1st step: production of edited video content through viewpoint selection;
 - 2nd step: personalized summarization through time segment selection and local story organization



Event recognition

1. Objectives

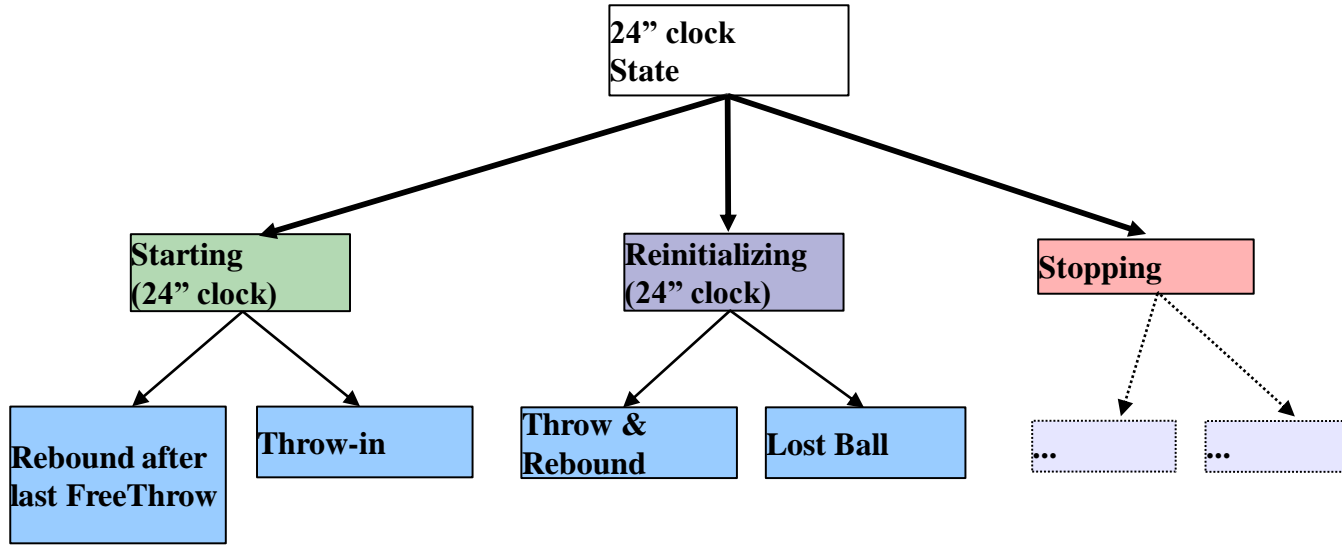
Recognize clock-events (field goals, violations, throw-in, lost balls, ...)

2. Hypothesis

24" clock and scoreboard information are available

3. Method

- 1. Tree-based organization of action*
- 2. Node decisions taken by classifiers*



- Discrimination at each node based on multi-feature classifiers
 - Clock state
 - Players or ball position
 - Next/previous action
- 92 % accuracy



Outline

- Understanding the scene
 - Player detection&tracking
 - Event recognition
- Automatic production of personalized summaries
 - 1st step: production of edited videos through viewpoint selection;
 - 2nd step: personalized summarization through time segment selection and local story organization

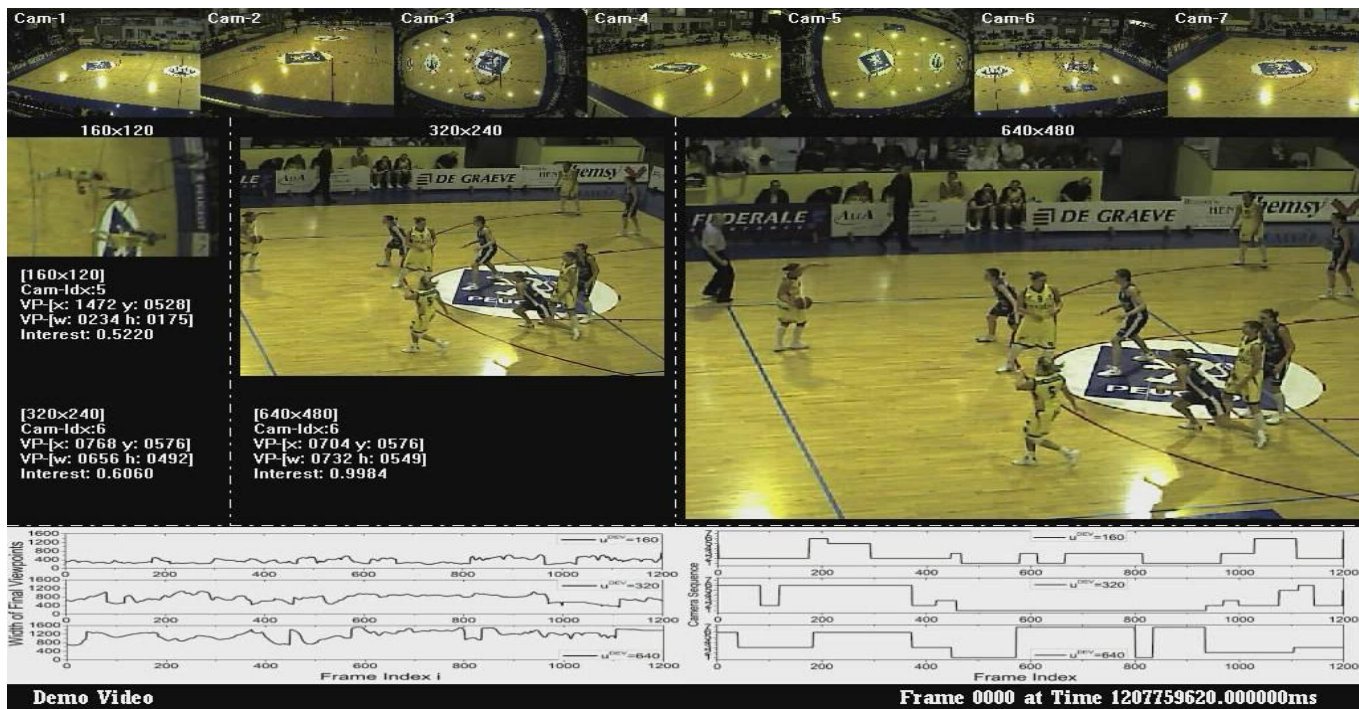


Automatic production

- **Objective:**
 - Given N cameras and a time segment, select the camera index and corresponding cropping parameters to render the scene at hand along the time segment.
- **General production rules:**
 - Completeness
 - Closeness
 - Smoothness

Owens J., 2007. Television sports production, 4th Edition, Burlington, MA, USA: Focal Press.

- Extensive subjective assessment
- Real-time implementation!
- *DEMO*





Outline

- Understanding the scene
 - Player detection&tracking
 - Event recognition
- Automatic production of personalized summaries
 - 1st step: production of edited videos through viewpoint selection;
 - 2nd step: personalized summarization through time segment selection and local story organization



Automatic summarization

- Select the temporal segments to render.
- Define how to tell the story (replays, etc).
- Account for user preferences
 - Preferred player/action;
 - Narrative preferences;
 - Summary duration;
 - Device resolution.

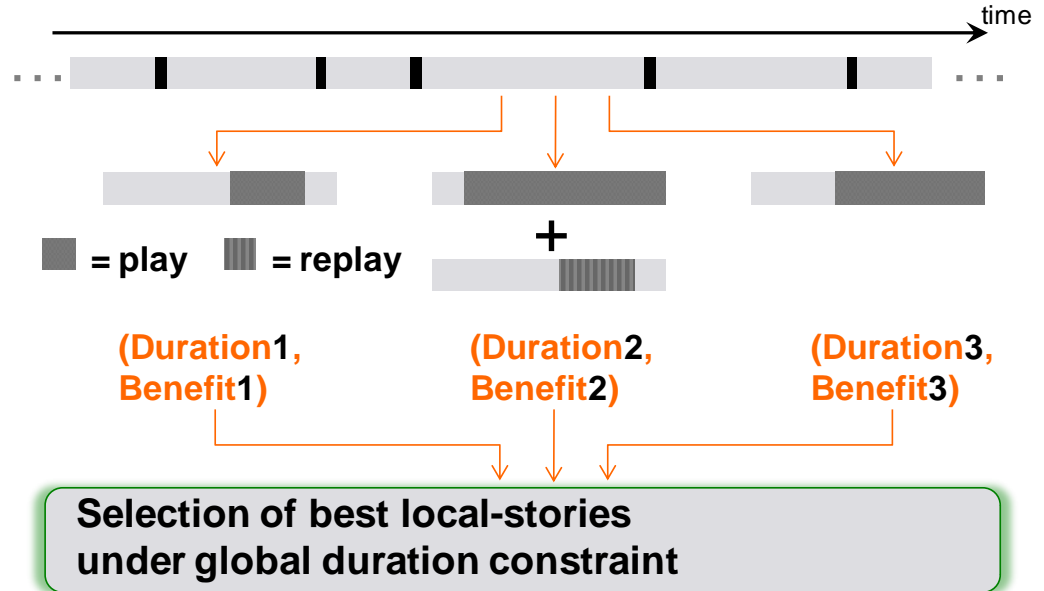


Resource allocation framework

Content-adaptive segmentation

Multiple narrative options for each segment

Benefit of each option depends on content and user preferences



Prototype on the booth.



Conclusions

- Multiview acquisition systems offers both
 - Effective scene analysis solutions (for automatic logging?);
 - Rich media content to produce visually pleasant content.
- Analysis directly impacts content personalization capabilities
 - Resource allocation framework easily takes advantage of richer semantic knowledge.



**Thanks
... and questions?**