Recognition of sport players' numbers using fast color segmentation

Cédric Verleysen

UCL, ICTEAM / ELEN

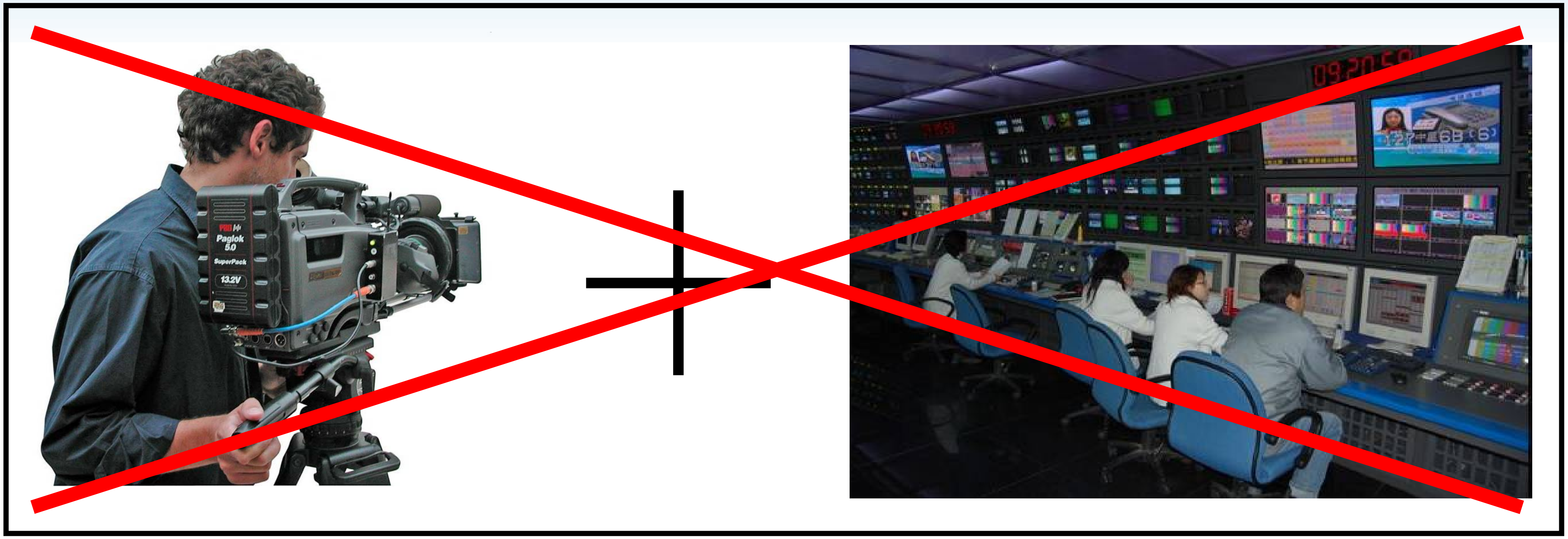
Organization of the presentation

- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

Organization of the presentation

- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

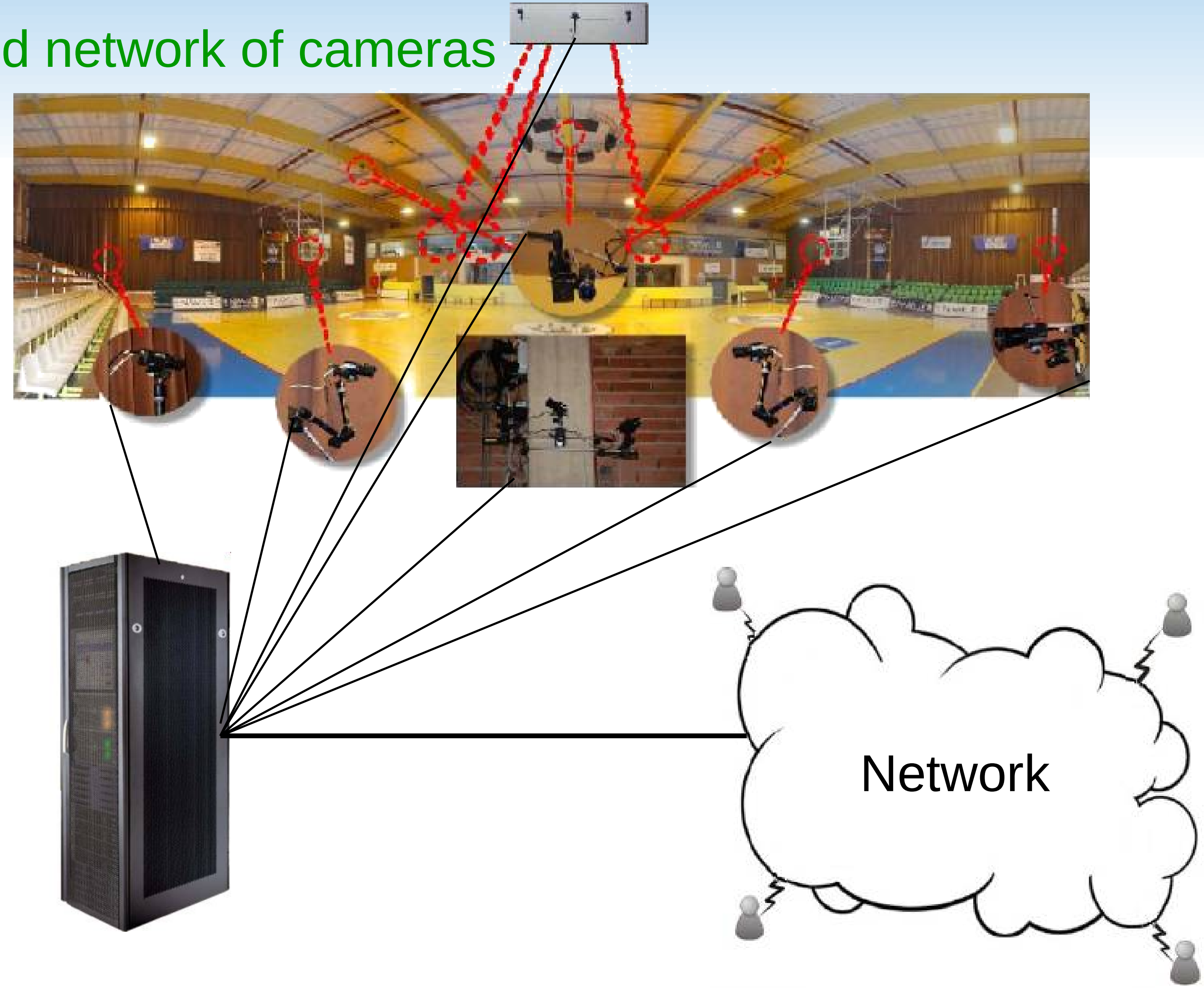
Context: autonomous production



Autonomous production

Context: acquisition infrastructure

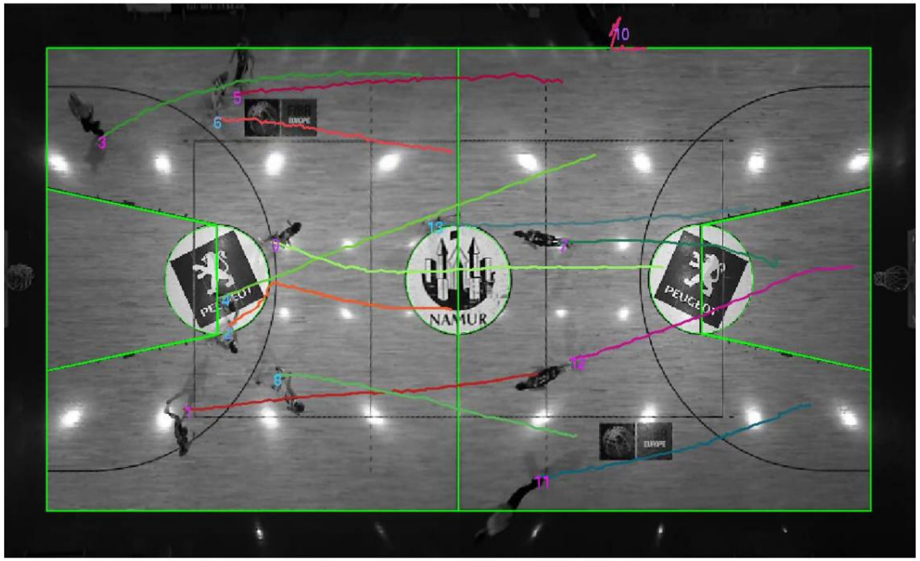
Distributed network of cameras



Context: why number recognition?

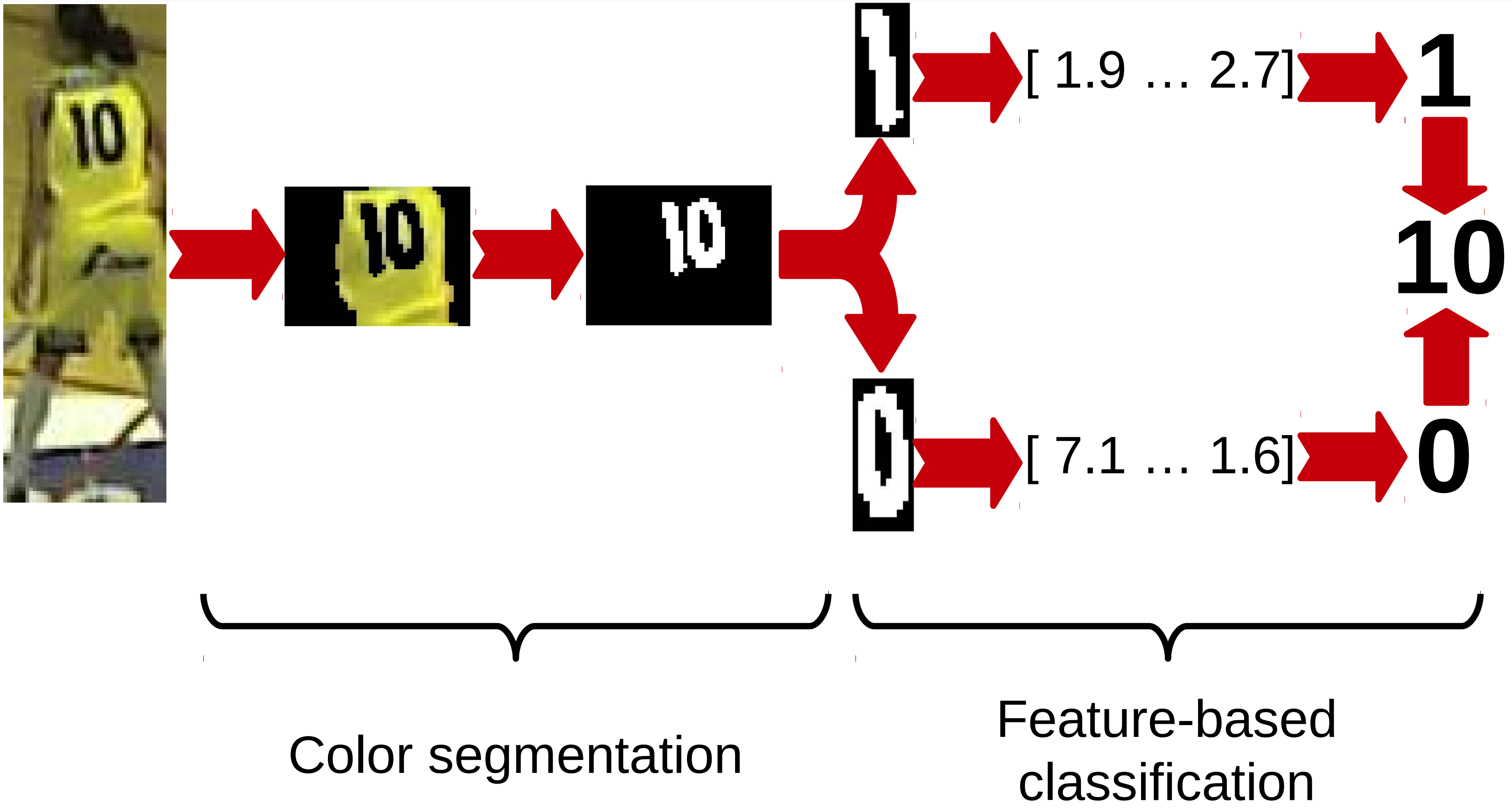
Tracking of players

Control of active cameras



etc...

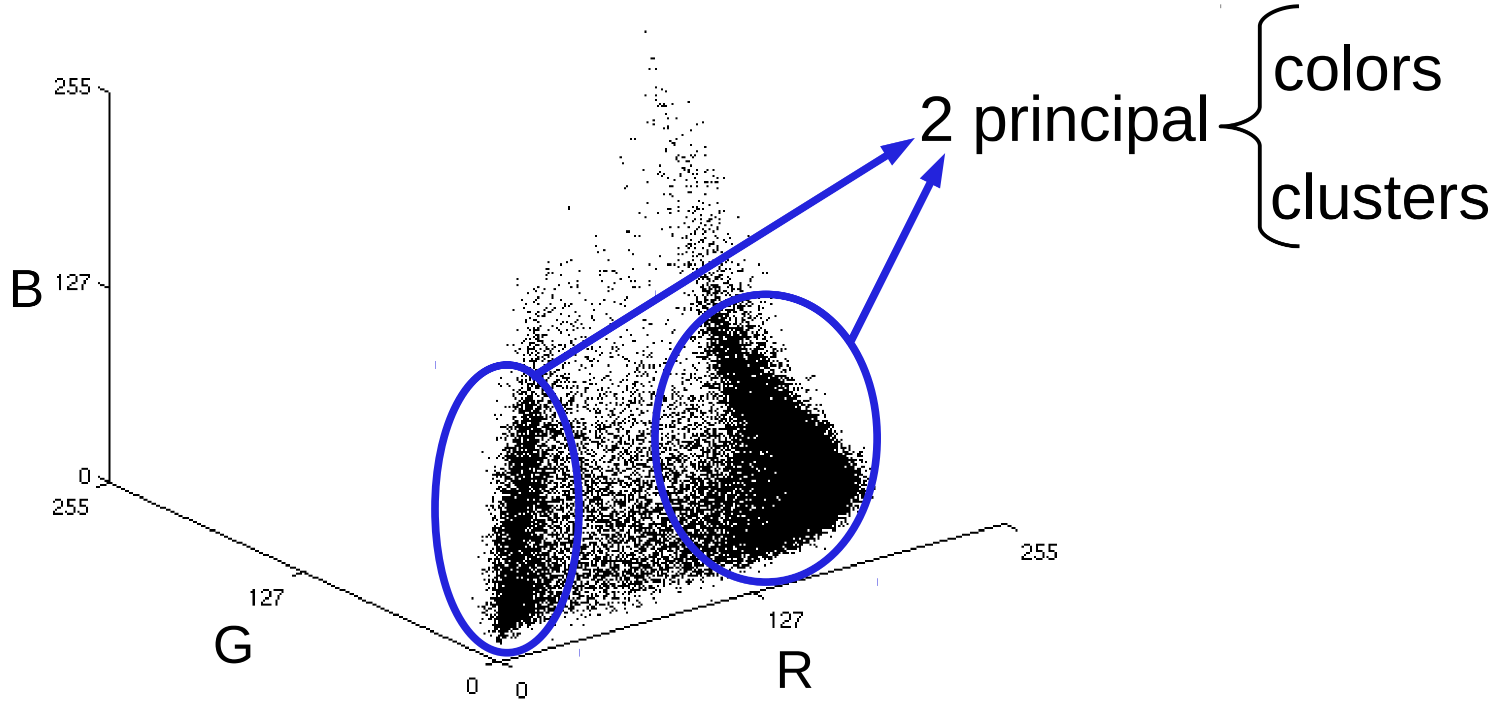
Color segmentation enables to isolate numbers for number recognition



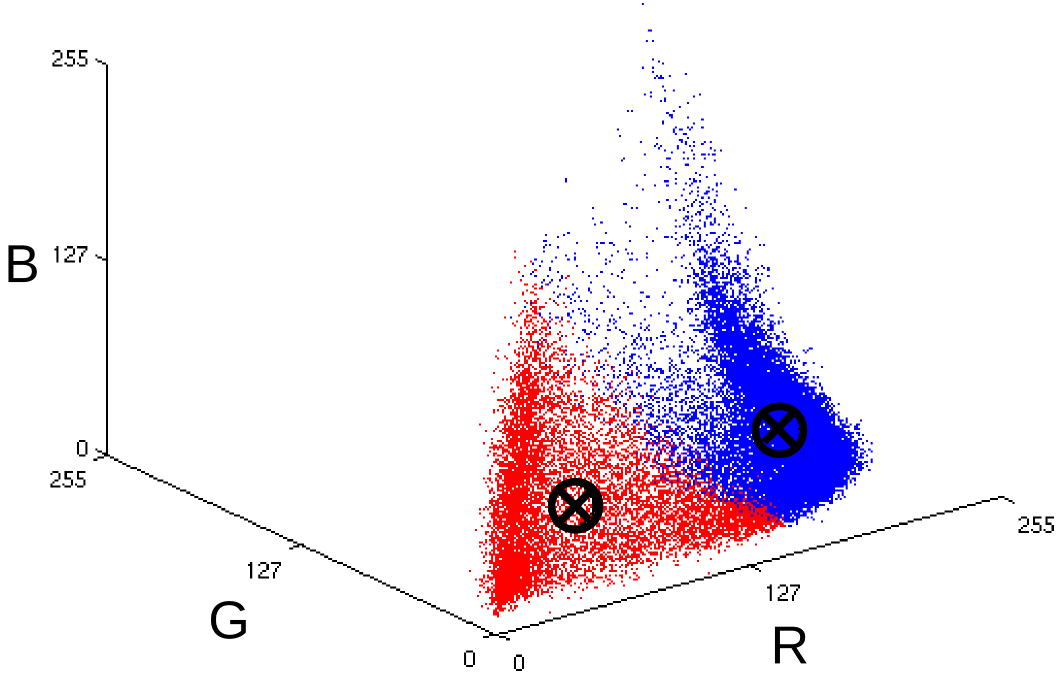
Organization of the presentation

- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

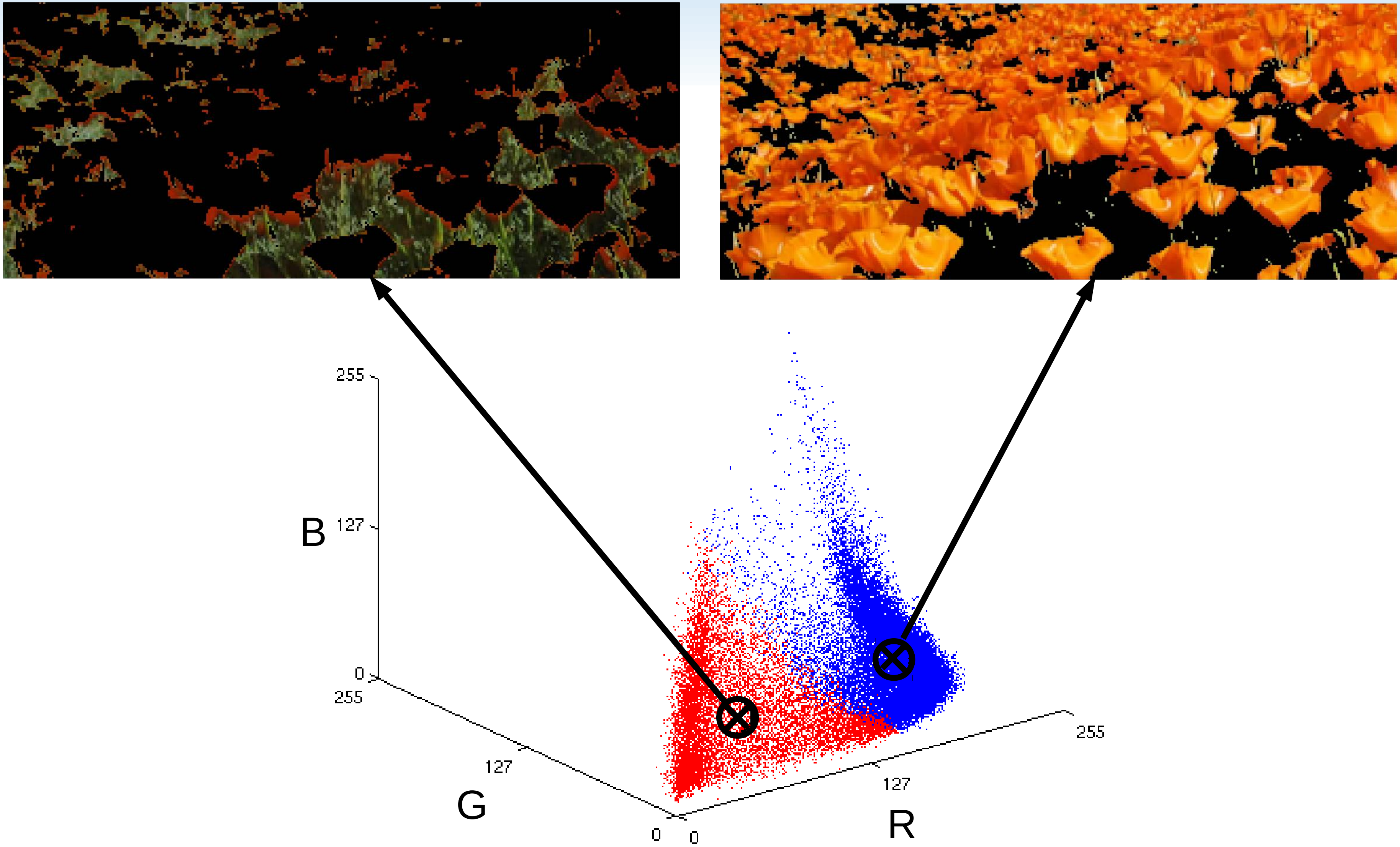
How does color segmentation work ?



K-means enables to partition N observations into K clusters



Each cluster represents a specific color

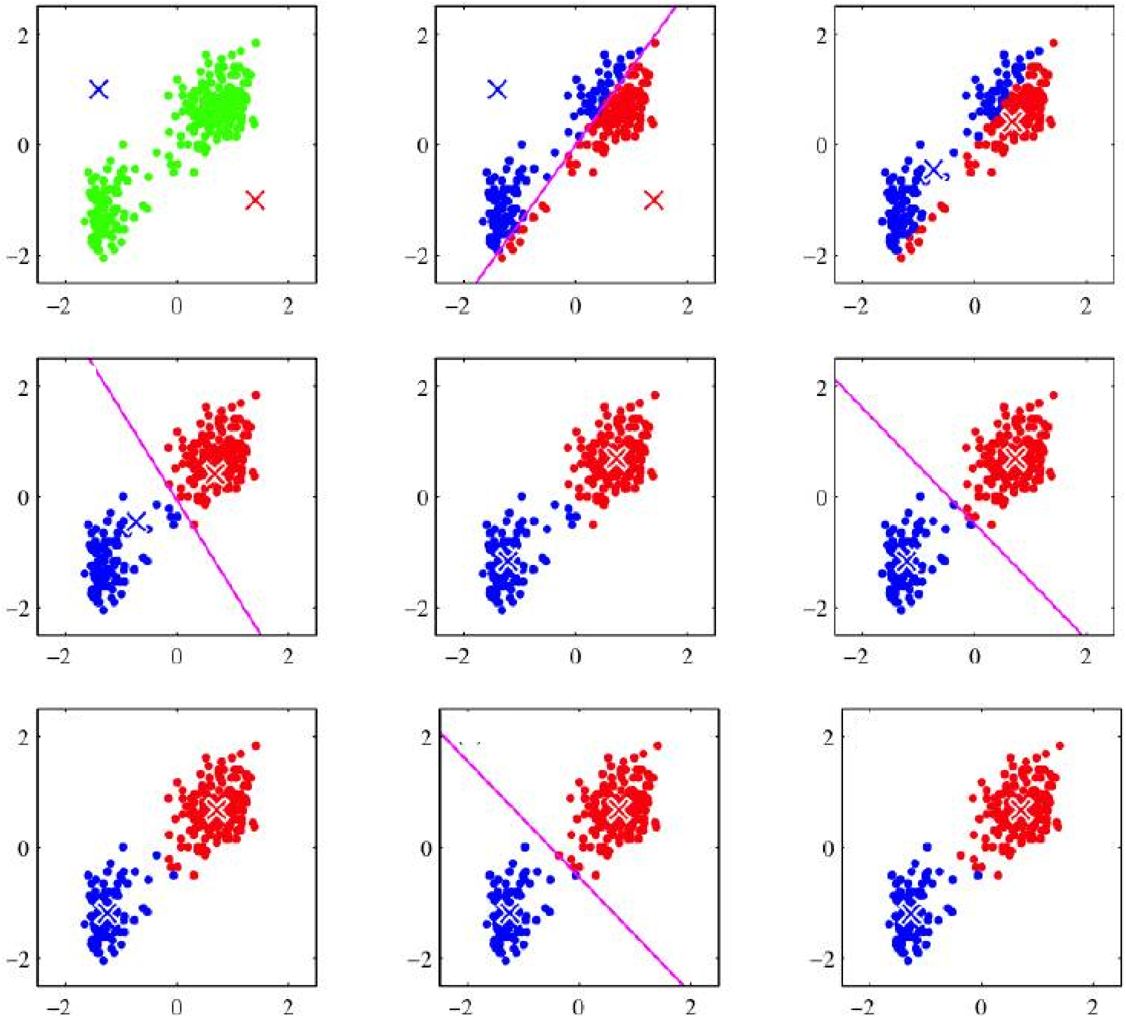


K-means problem setup

- Have N data points $\vec{x}_1, \ldots, \vec{x}_N$ with $\vec{x}_i \in \Re^D$
- Will build K clusters, presumably K << N
- Each cluster k has a cluster center ("centroid") $\vec{\mu}_k \in \Re^D$
- Have a dissimilarity measure $\mathcal{V} : \Re^D \times \Re^D \longmapsto \Re$
- Problem is to assign a cluster label k_n to $\vec{x}_1, \ldots, \vec{x}_N$ such that

$$k_n = \text{argmin}_k \mathcal{V}(\vec{x}_n, \vec{\mu}_k).$$

K-means algorithm



1) For *l = 1, ... , N*, assign $\vec{x}_n$ to its closest centroid $\vec{\mu}_k$

2) For *k = 1, ... , K*, compute as the mean of its assigned data points $\{\vec{x}_n \ : \ k_n = k\}$

3) Go to 1) until a stopping criterion (number of iterations, centroids don't change,...) is met

K-means issues

- Number of clusters K has to be known in advance
- To extract a specific object, its color has to be known

⇒ **Necessity of apriori** *[Ravichandran, IJCAM, 2009]*

- **Computational complexity** of Θ(I K N D)

D → Dimension

N → Data points

K → Clusters

I → Iterations

Proposed solutions to k-means issues

- **Learn apriori on multiple images of the same object**

→ Complexity Θ(I K N D) increases because N represents ONE pixel of ONE image

- **Each image is represented by a vector of features (<# pix.)**

→ Complexity Θ(I K N D) decreases dramatically because N represents ONE image times the number of features (D)

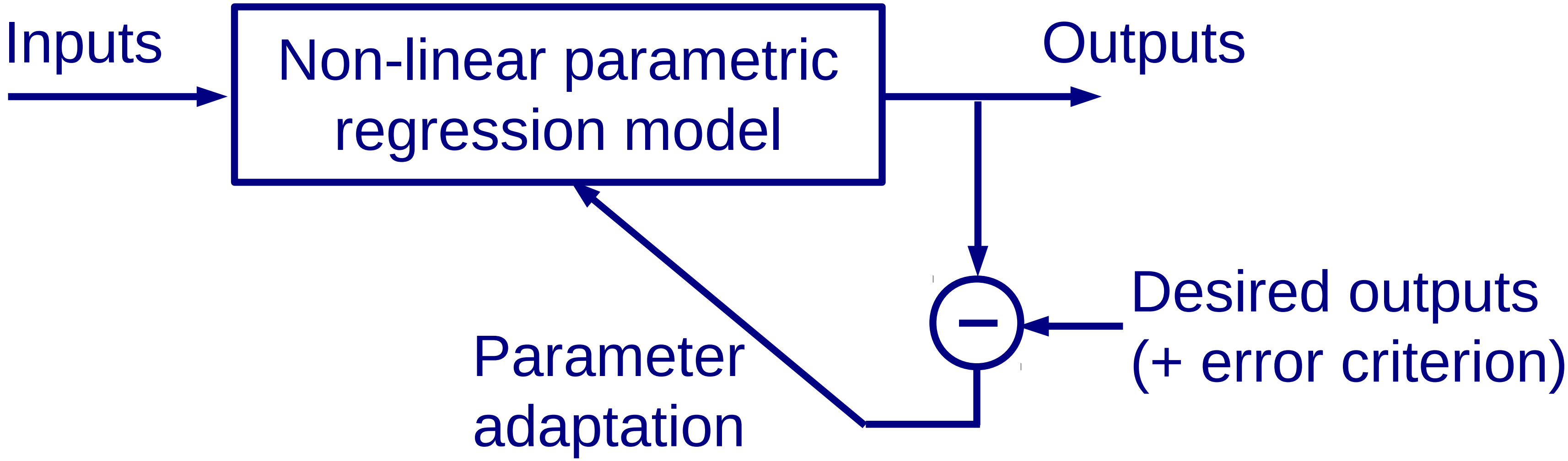
→ LEARNS APRIORI + SPEEDS UP COMPUTATIONS

Organization of the presentation

- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

Once the object extracted, its recognition can be done with a feature-based classifier

- Various supervised machine learning techniques

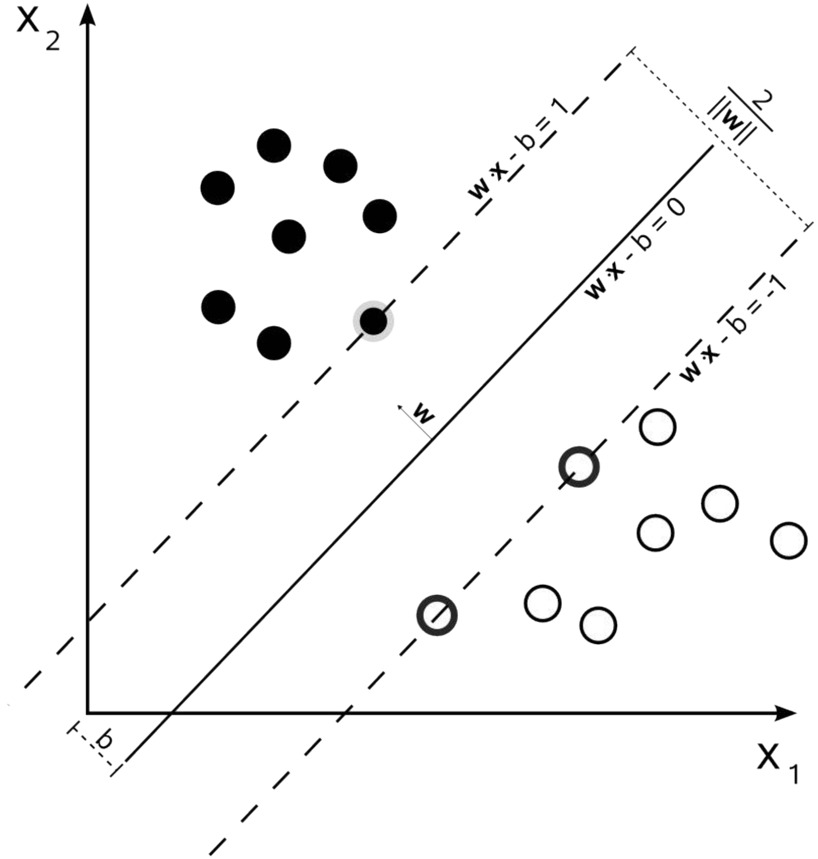


- If input = 256x256 8-bits color image = $256^{(3.256.256)}$ possibilities

Necessity to work with features!

Support vector machine is a common classifier

- Unseparable data can become separable in higher dimensions
- The larger the margin, the lower the generalization error

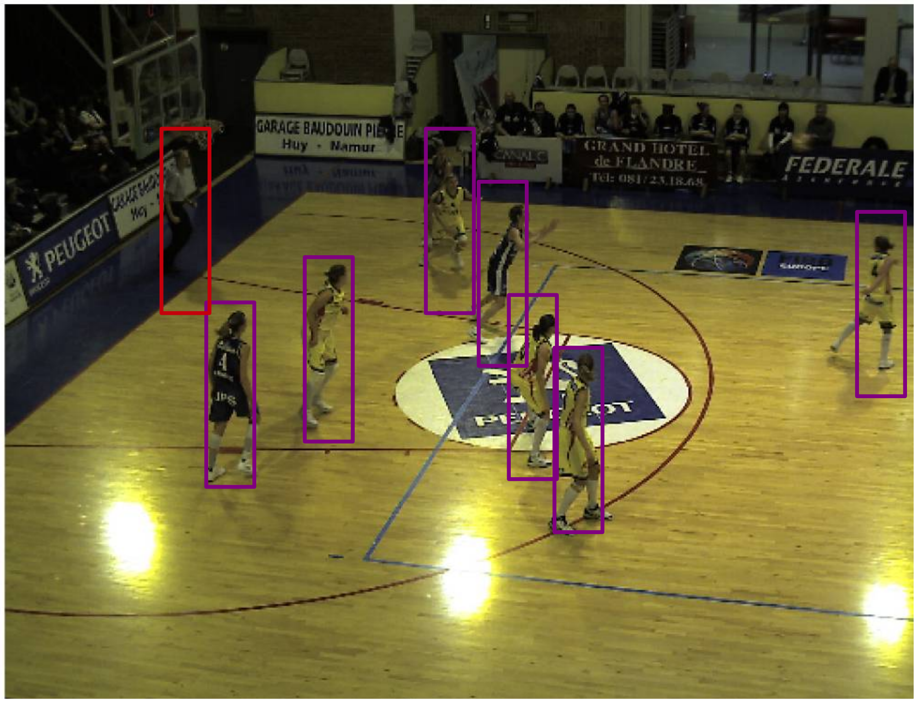


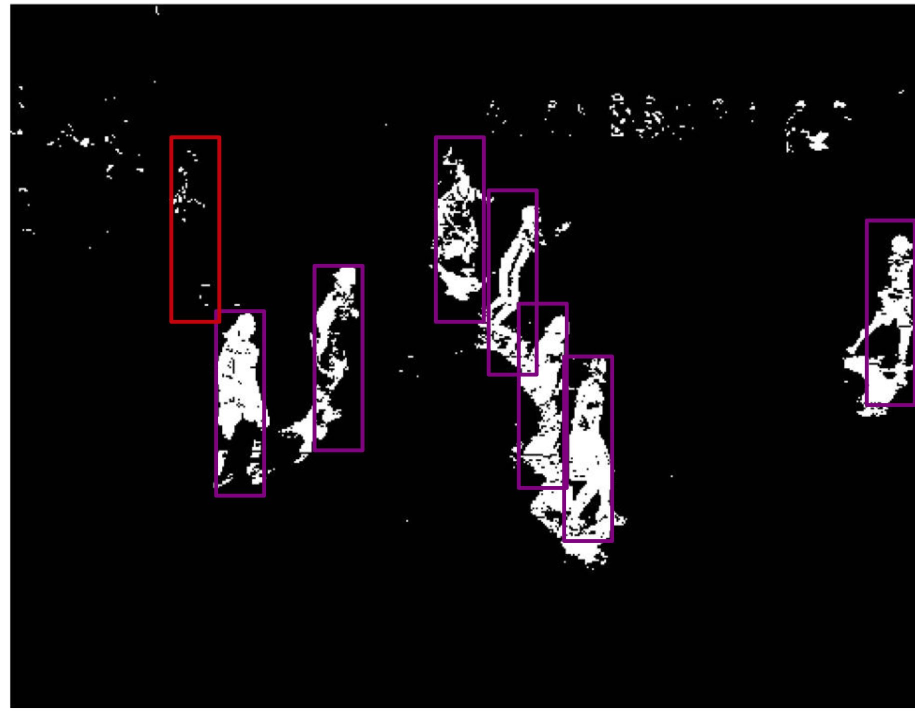
Organization of the presentation

- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

Only players are interesting...

- Inputs (YUV422 or YUV420):

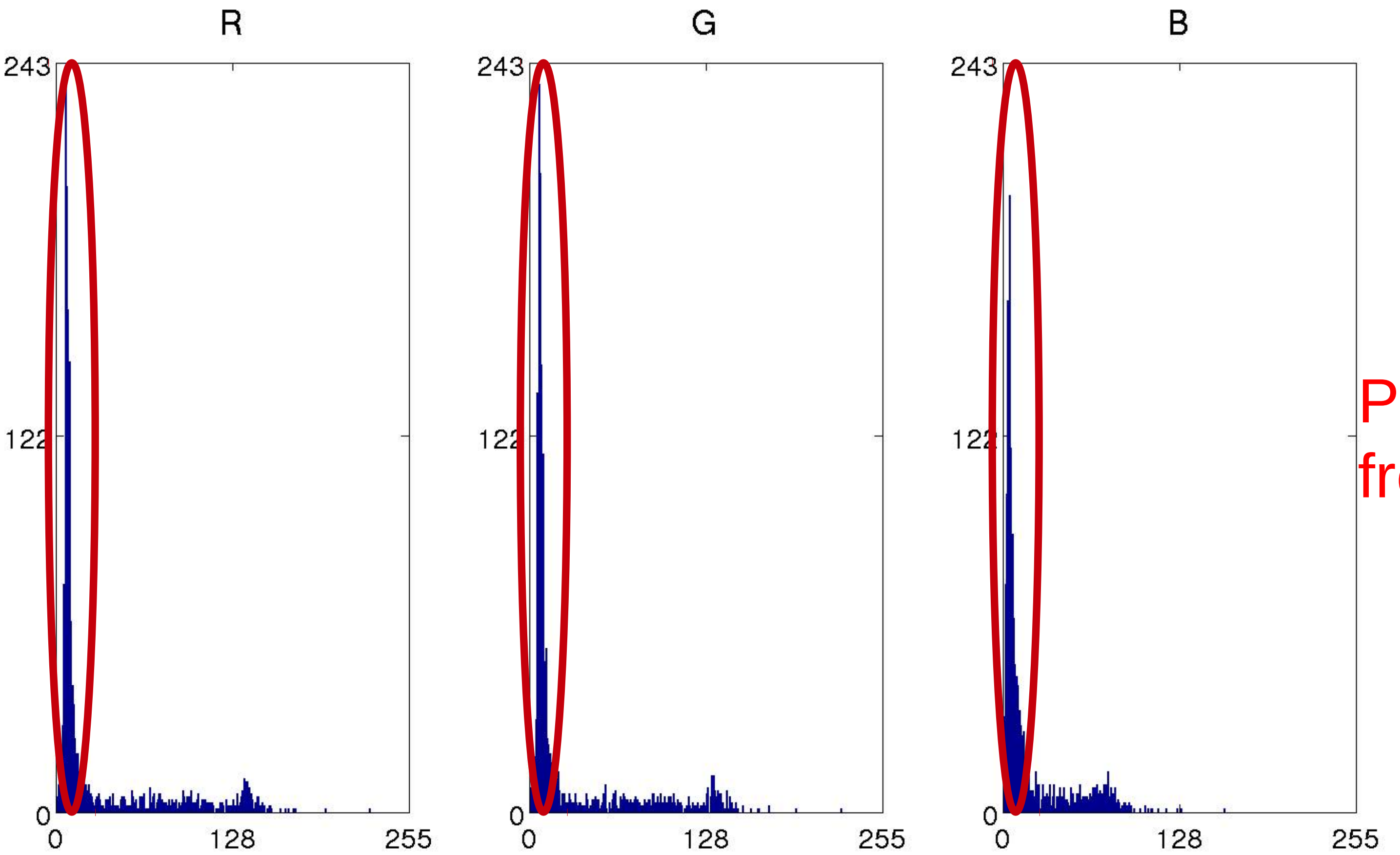




Reject referees!

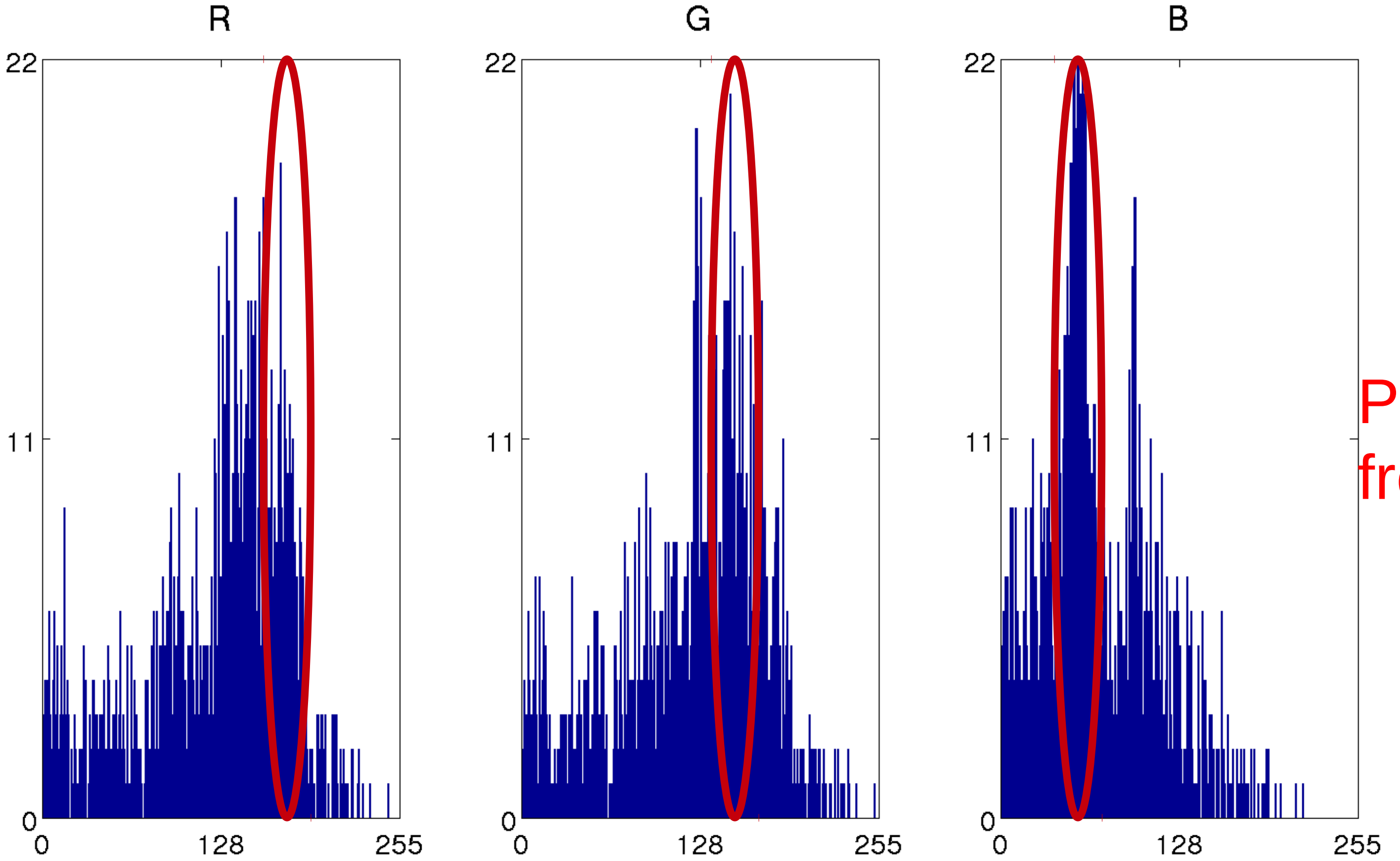
Referees always have a black pant

- Histograms without background



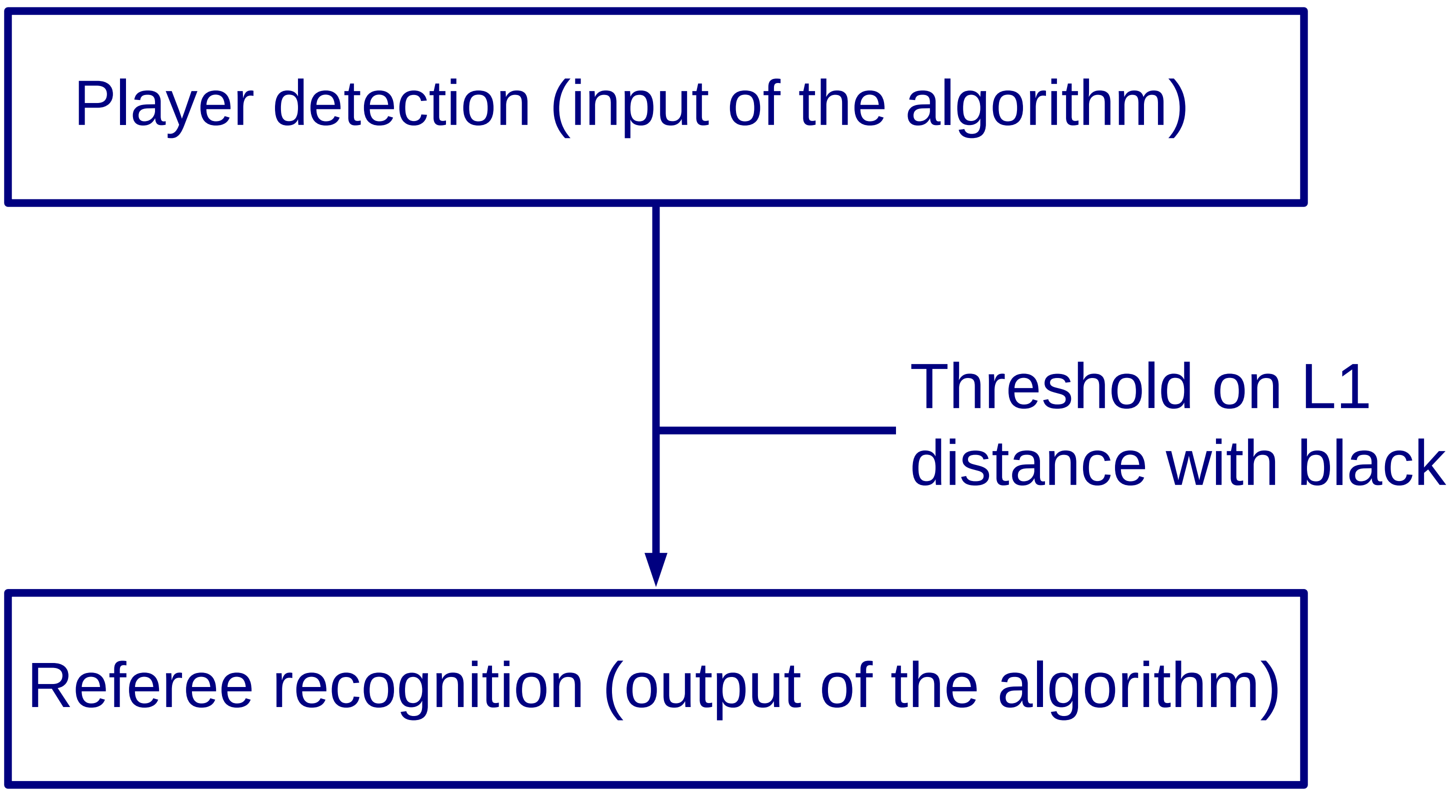
While players wear shorts

- Histograms without background



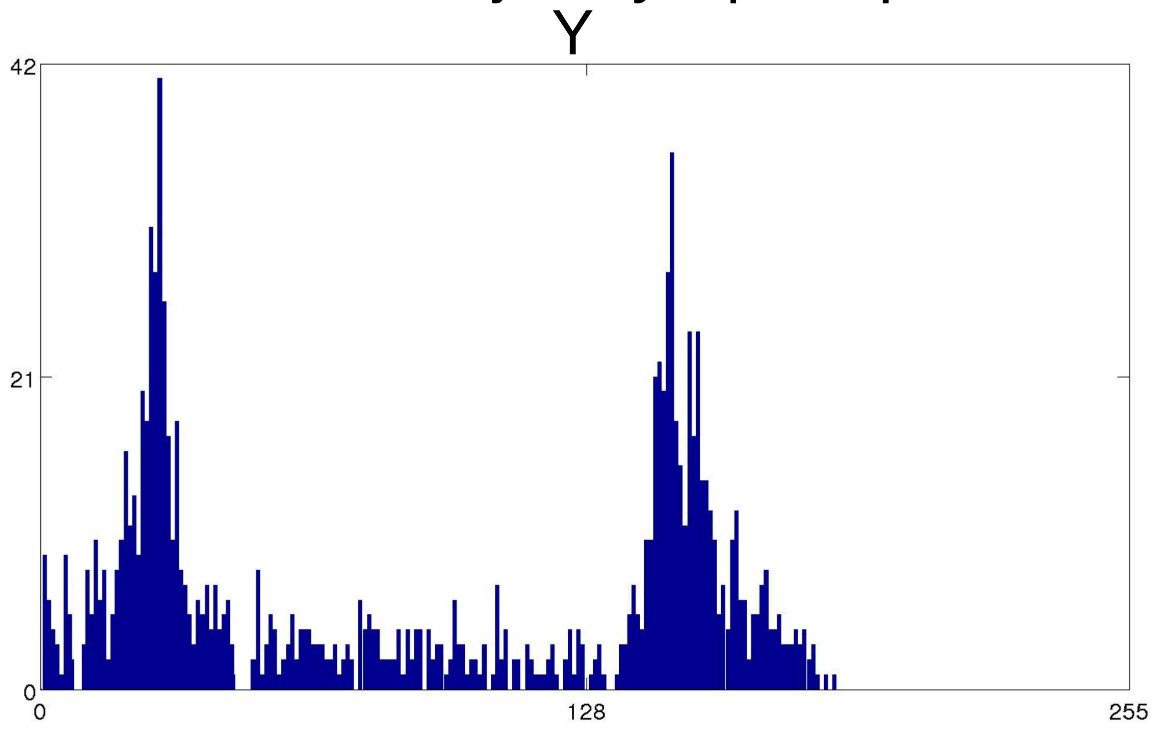
Peaks are FAR from (0,0,0)

Threshold the L_1 distance between peaks and wanted color is efficient and fast



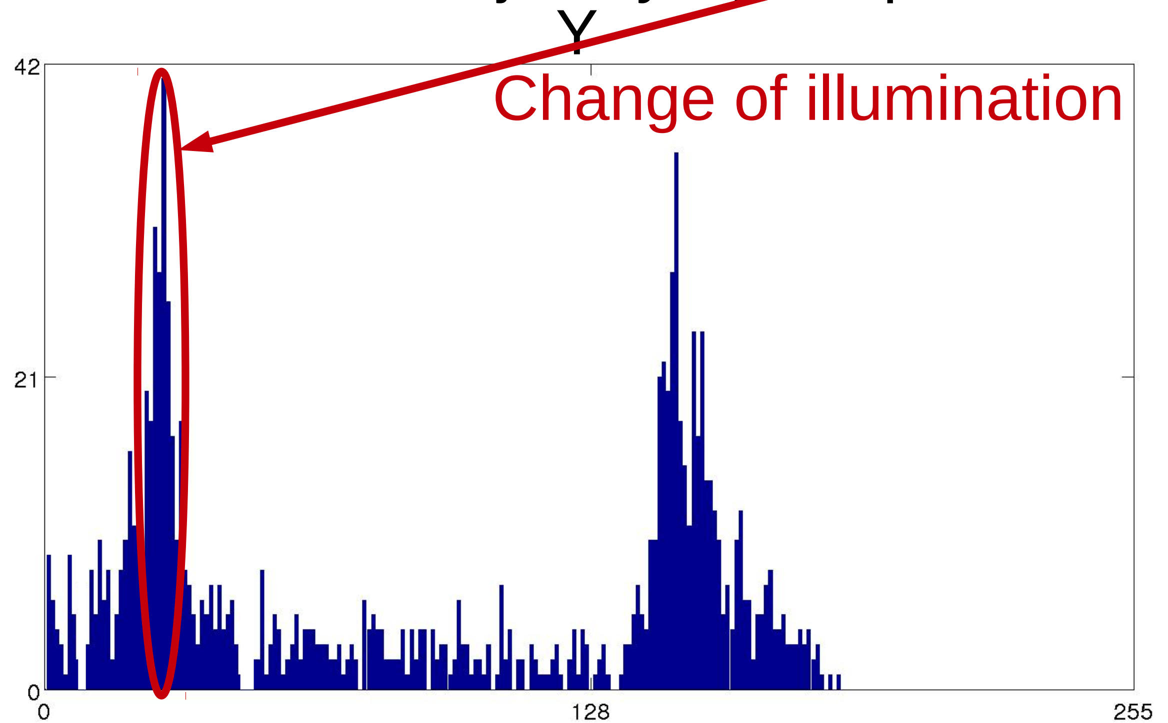
Principal color components can be noisy

1) Features to describe jerseys' principal color :



Illumination changes are concentrated

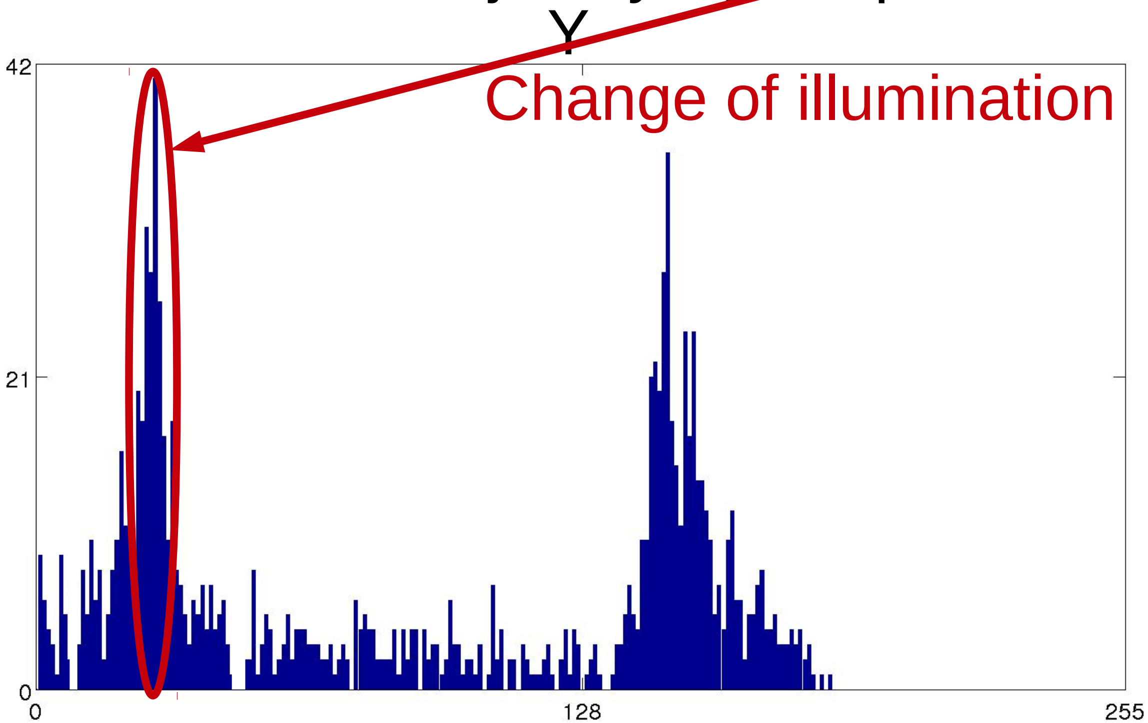
1) Features to describe jerseys' principal color :



$$Non\_isolated\_peak = \arg\max_{i \in \{peak_1,\, peak_2\}} (peak\_value_i \cdot peak\_basis_i)$$

Teams colors are learnt using NON-ISOLATED histogram peak values of the jerseys

1) Features to describe jerseys' principal color :



2) Accumulation of these features in YUV space and K-means

3) Planar separation between teams determinated by centroids

Recognition of the red team (untuned thresholds)

Video

Organization of the presentation

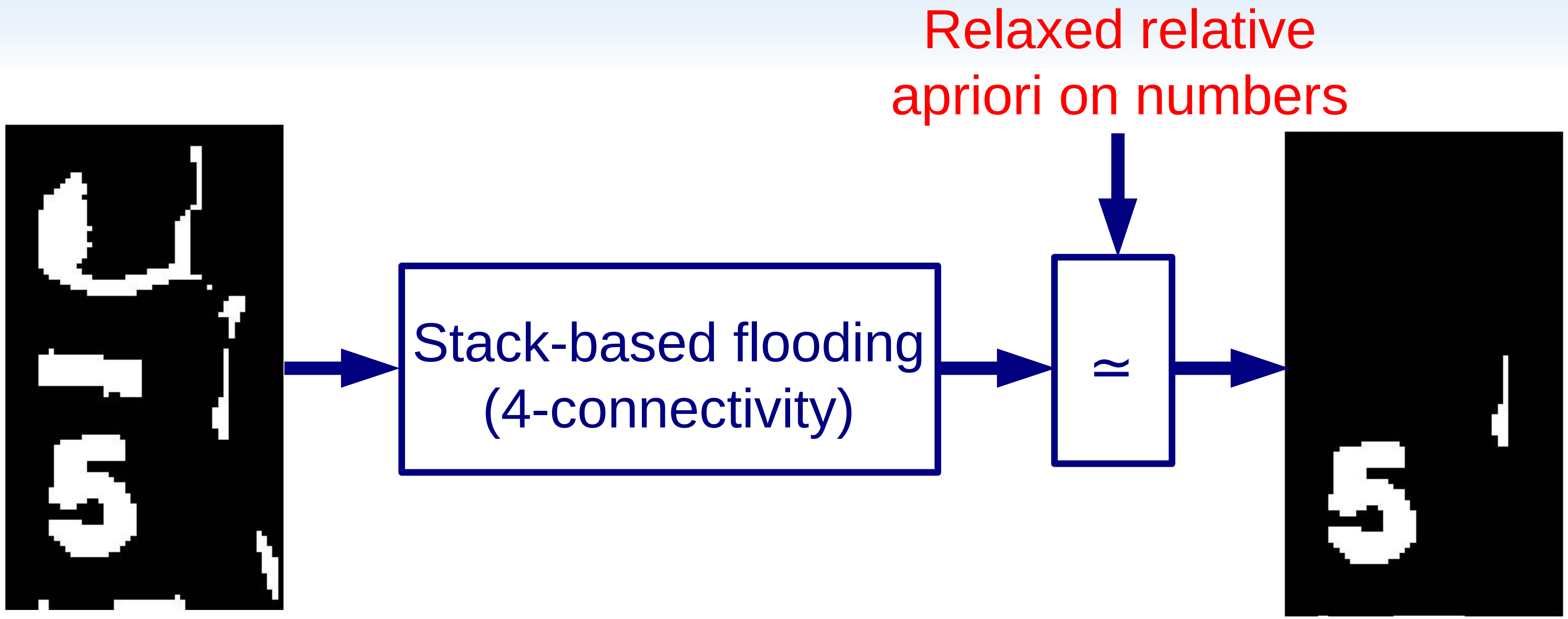
- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

After knowing the color of the jersey, we can segment the number

- Number should be the second principal color on the jersey

Remove jersey's color and apply principal color extraction

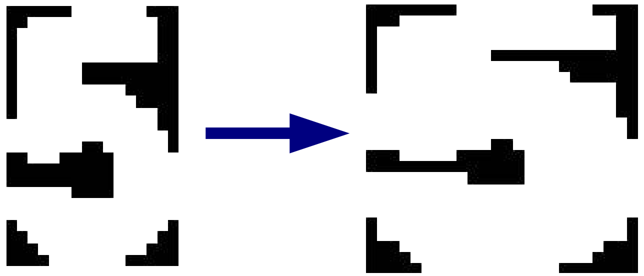
Relaxed constraints reject most of the non-number parts



- Number relative width
- Number relative height
- Number relative density

To classify numbers, some features are necessary

- Constant size pictures

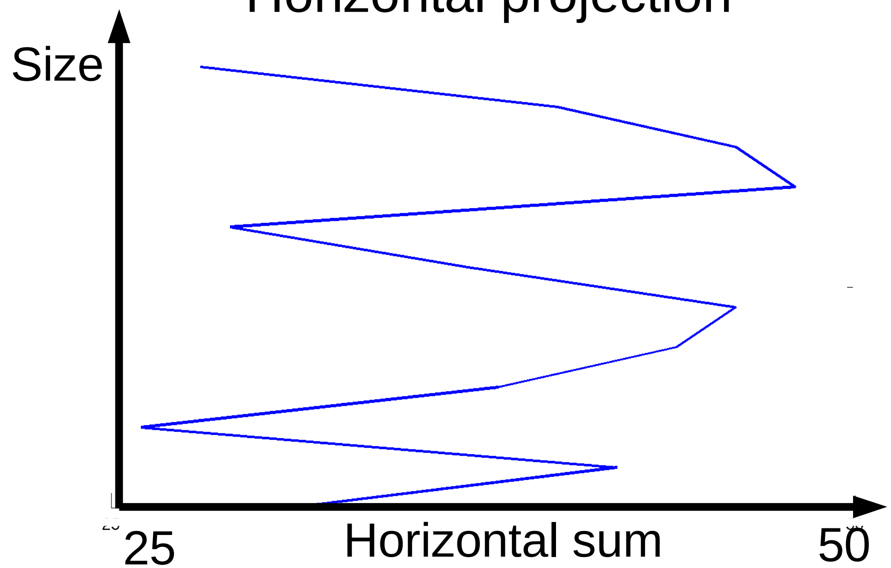


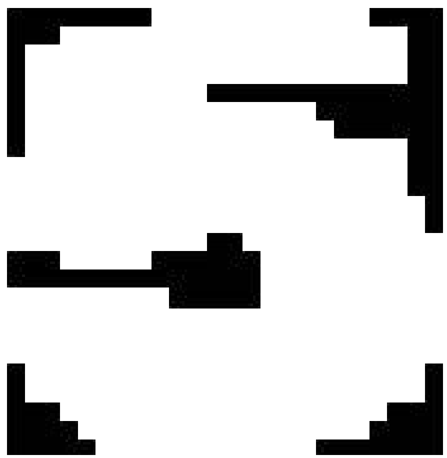
- Choosen features *[Delannay and al., 3rd ACM Dist. Cam., 2009]*:
  - Ratio height/width (before resizing)
  - Number of holes *[Dey and al., 13th conf. VLSI design, 2000]*
  - Central moments m_{01} , m_{10} , m_{02}, m_{20} and m_{22}

To classify numbers, some features are necessary

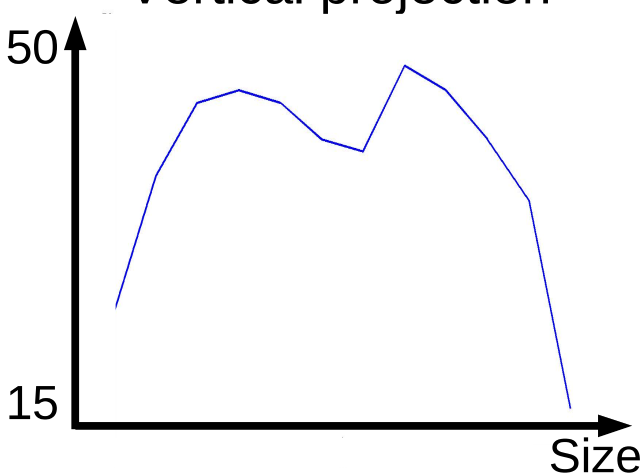
- Horizontal and vertical projections

Horizontal projection

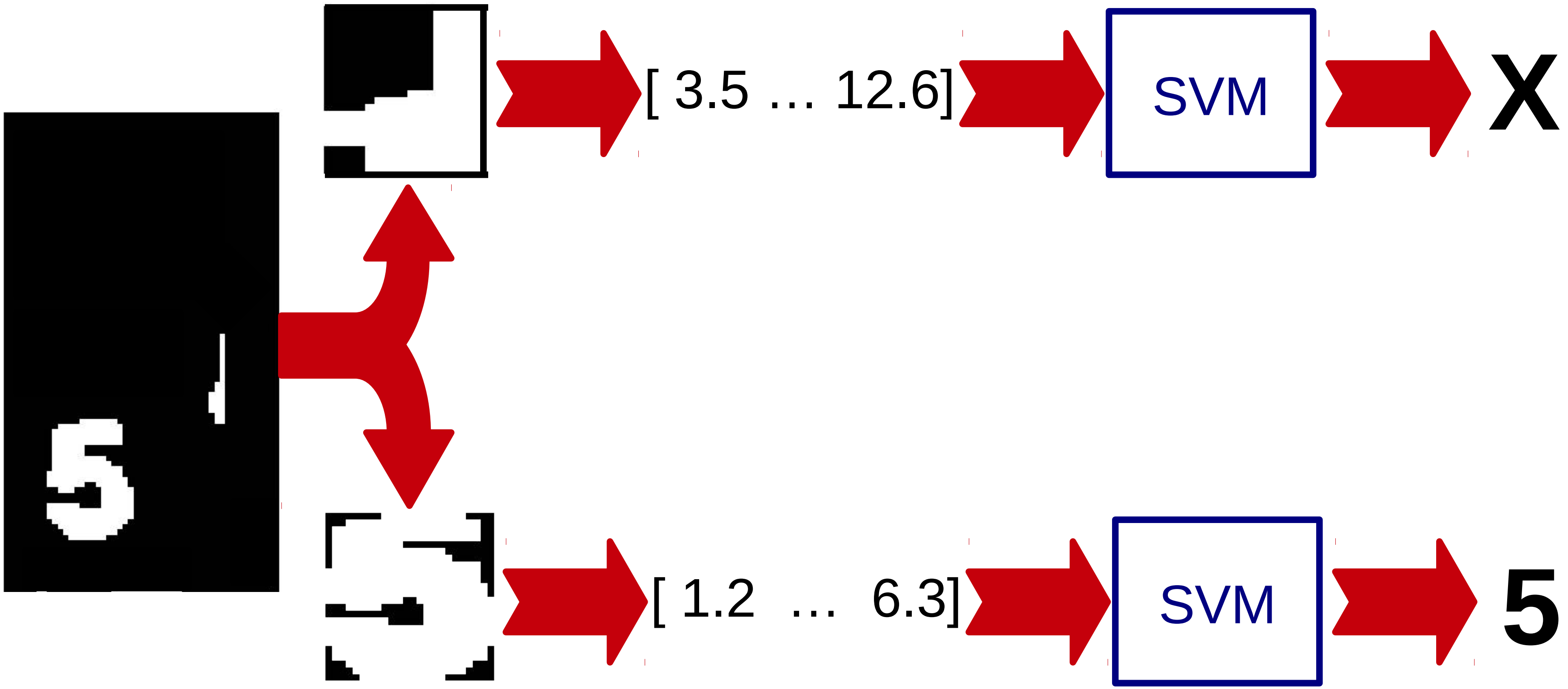




Vertical projection



SVM discriminates between numbers



Organization of the presentation

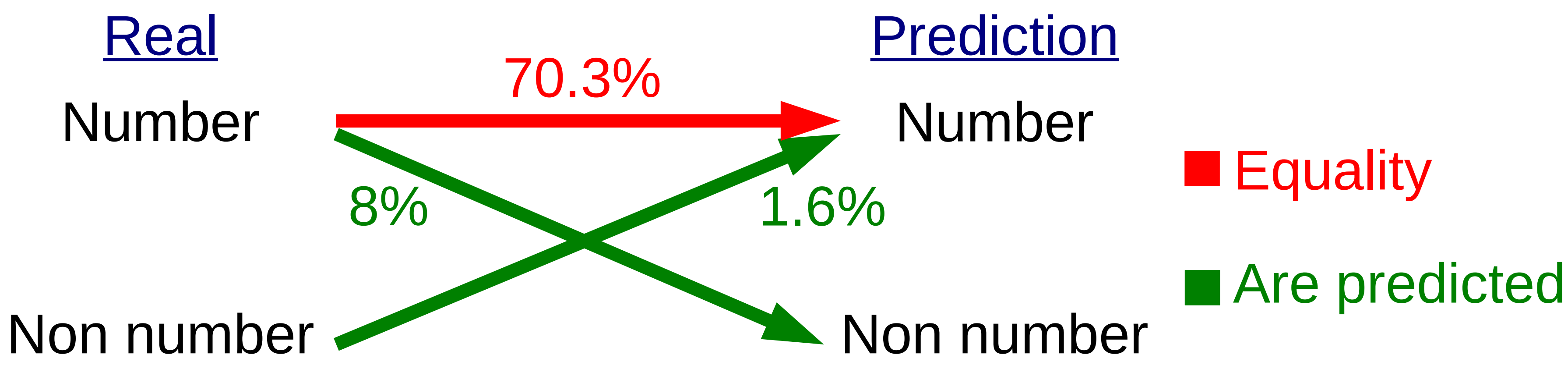
- Context
- General principles
  - *Color segmentation*
  - *Feature-based classification*
- Validation on a basketball game
  - *Team recognition*
  - *Number recognition*
- Performances

Player's caracterization runs in real-time

- C program using Intel® **I**ntegrated **P**erformance **P**rimitives
- Worst-case computation time: 14 people on the same frame

9 ms per frame (Intel I7, 3 GHz, 8 Gb RAM)

- Performances of the classifier (SVM) **on the trained template**:

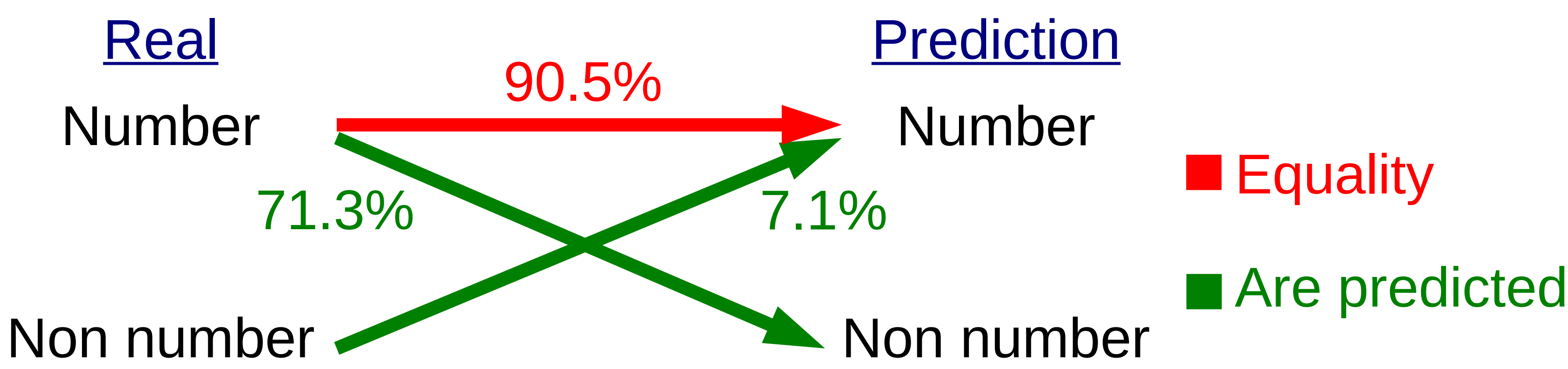


Good recognition, but numbers are often considered as non numbers

- C program using Intel® **I**ntegrated **P**erformance **P**rimitives
- Worst-case computation time: 14 people on the same frame

9 ms per frame (Intel I7, 3 GHz, 8 Gb RAM)

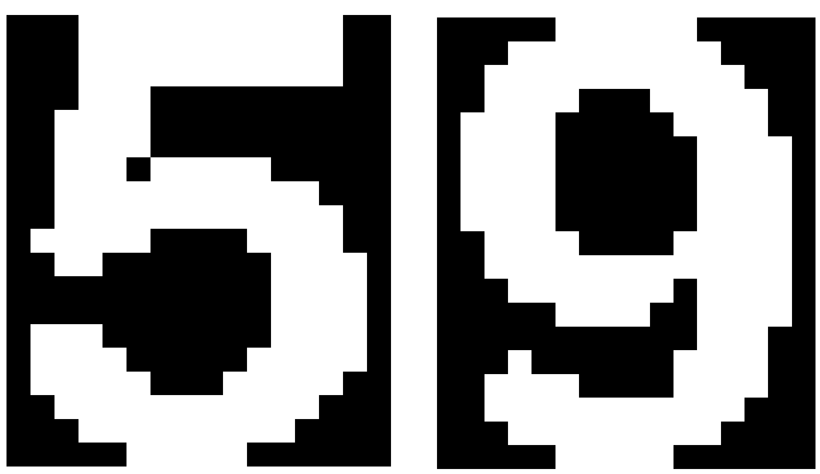
- Performances of the classifier (SVM) **on a random match**:



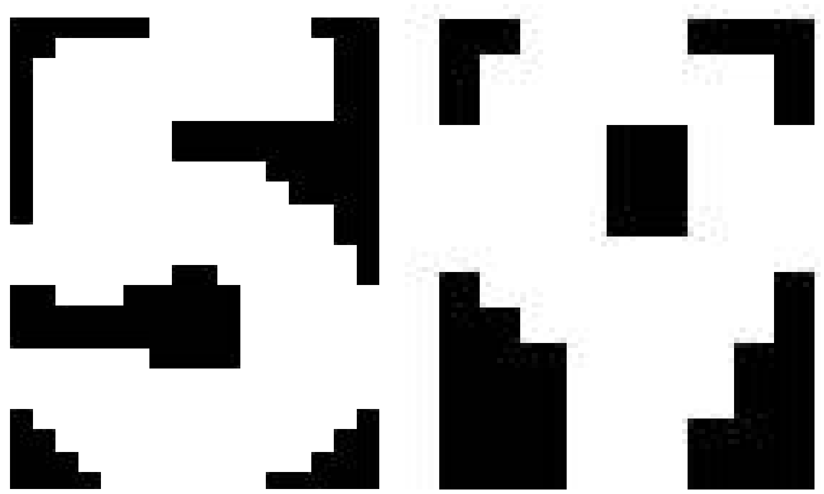
False negatives comes from the training

- Font of numbers varies from a game to another !
- Train various transformations (rotation, sheering, resizing,...) of **one font is not sufficient** for generalization

Training



Real game



Recognition of player 5 of the red team

Video

Conclusion and perspectives

- Color segmentation enables to extract a specific object if we know an apriori on its color(s)
- Apriori can be learnt from images of the same object (more robust if colorimetry adjustment has already been done)
- Color feature extraction dramatically speeds-up segmentation
- Non-isolated histogram peaks are robust color features
- Mixing color segmentation and feature-based classification can give a powerful OCR
- Feature-based classifier has to be trained with various fonts of number and with very discriminative number features

Some references

- K.S. Ravichandran and B. Ananthi, "Color skin segmentation using k-means cluster", International Journal of Computational and Applied Mathematics (IJCAM), Vol. 4(2), pp.153-157, 2009.
- M. Luo, Y.F. Ma and H.J. Zhang, "A spatial constrained k-means approach to image segmentation", The 4th Pacific Rim Conference on Multimedia, Communications and Signal Processing, 2003.
- D. Delannay, N. Danhier and C. De Vleeschouwer, "Detection and Recognition of Sports(wo)men from Multiple Views", Third ACM/IEEE International Conference on Distributed Smart Cameras, 2009.
- S. Dey, B.B. Bhattacharya, M.K. Kundu and T. Acharya, "A fast algorithm for computing the Euler number of an image and its VLSI implementation", Thirteenth International Conference on VLSI Design, pp. 330-341, 2000.
- F. Chen and C. De Vleeschouwer, "Autonomous production of basket-ball videos from multi-sensored data with personalized viewpoints", The 10th IEEE International workshop for multimedia interactive services, pp.81-84, 2009.