A probabilistic pixel-based approach to detect humans in video streams

Sébastien Piérard, Antoine Lejeune
and Marc Van Droogenbroeck

INTELSIG, Montefiore Institute, University of Liège, Belgium

May, 26th 2011

Outline

1. The detection of humans in video streams as a silhouette classification problem
2. Design of a new robust probabilistic pixel-based approach to classify silhouettes
3. Data, results, and conclusions

Outline

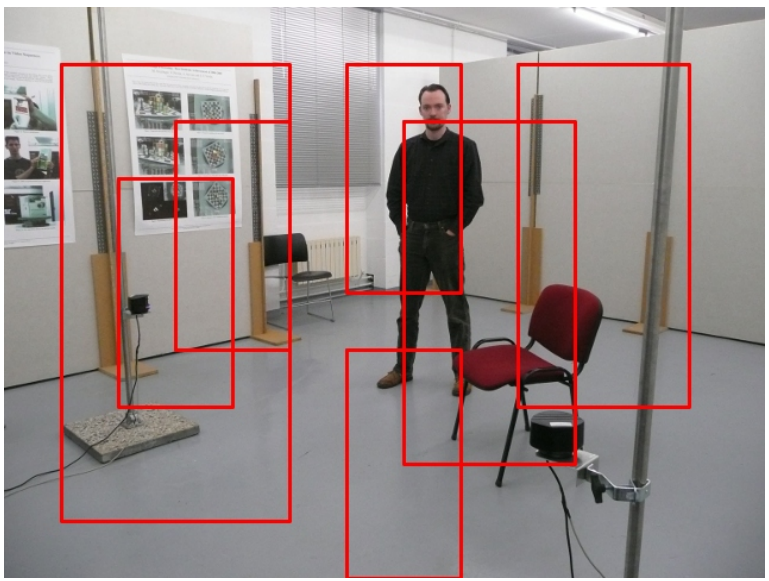
Detecting humans

Detecting humans is useful for video surveillance, automatic video scene analysis and interpretation, home entertainment, etc.

Our goal is to detect only humans, nothing else !
$\hookrightarrow$ A binary classification $\{$ human , non-human $\}$ is required.

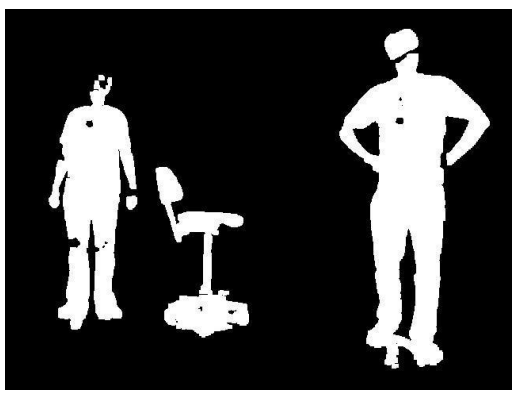
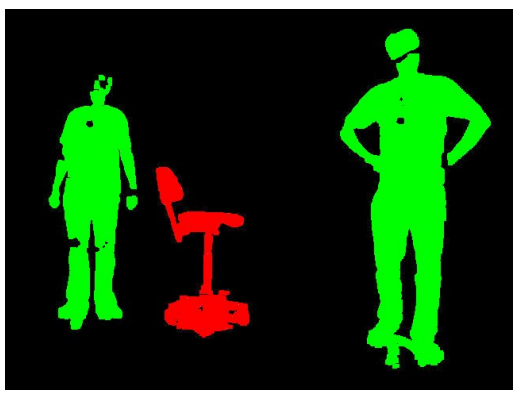
1st approach : detection in still images

- Example : Histograms of oriented gradients (HOG) (Dalal & Triggs, 2005)
  - Based on appearance (colors and textures)
  - A lot of detection windows have to be considered per image
    $\hookrightarrow$ computationally expensive
    $\hookrightarrow$ It is difficult to obtain a low false alarm rate per image



2nd approach : detection in video streams

- Example : the work of Barnich *et al.* (2006)
  - Takes advantage of the temporal information (motion) by the use of a background subtraction algorithm (*eg* ViBe)
  - Based on geometric information (the external contours)
  - Only a few blobs have to be considered per image
    $\hookrightarrow$ It is possible to obtain low false alarm rates

Outline

Background subtraction : defects of silhouettes

Moving persons or objects merged, carried objects detected, noisy contours, holes, shadows detected as foreground, . . .

$\Longrightarrow$ We need a robust silhouette description method !

Describing silhouettes

A wide variety of shape description techniques exists.

contour-based / region-based

Region-based methods are less sensitive to noise, since the noise affects only the contours in binary images.

global / local

Local description methods split shapes into smaller components, to limit the influence of defects to a few components.

$\Longrightarrow$ a **region-based**, **local description** method is preferable

Our silhouette description method

Our silhouette description method is the simplest region-based local description method one can imagine :

the set of all pixels included in the shape

But, how can we classify sets ? Unfortunately, there is no machine learning method designed for this task. Let's do it by ourselves !

Classification of compound objects

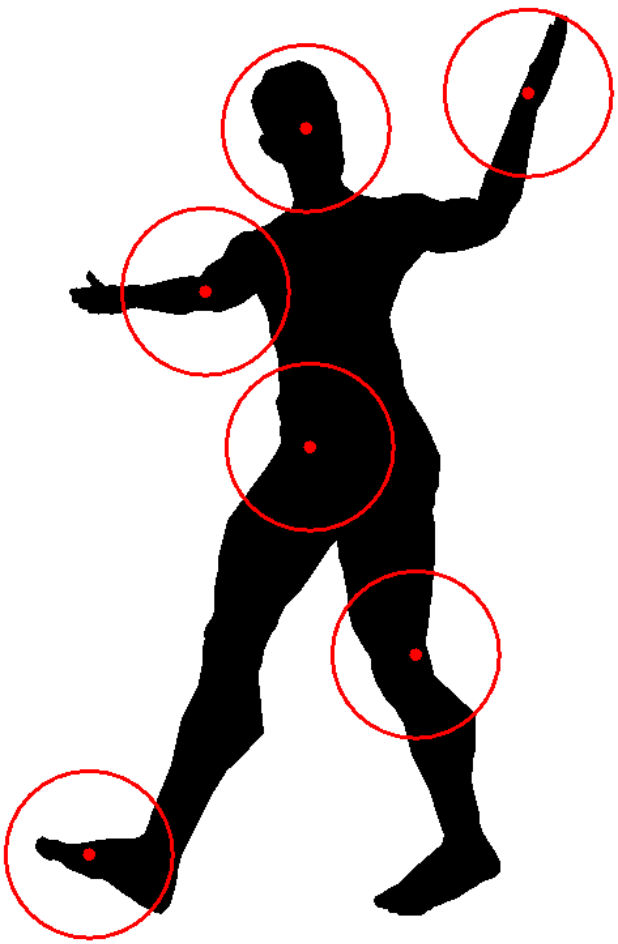
Our method is inspired by the work of Barnich *et al.* (2006)

- Each pixel plays the role of an "expert". It decides if it is part of a human silhouette or not, and gives the probability for the decision to be correct.
- The opinions given by the experts are merged by a weighted voting mechanism. The weight given to an expert depends on the probability for its decision to be correct.

The experts can be implemented by machine learning methods.

Example

Let $p_+(x)$ be the probability for a pixel x (an expert) to belong to a human silhouette.



$\rightarrow p_+(x) = 10\,\%$

$\rightarrow p_+(x) = 90\,\%$

$\rightarrow p_+(x) = 60\,\%$

$\rightarrow p_+(x) = 50\,\%$

$\rightarrow p_+(x) = 60\,\%$

$\rightarrow p_+(x) = 40\,\%$

$\Longrightarrow$ This silhouette is human !

Using the ExtRaTrees to implement the experts (1/2)

x is a pixel (an expert).
$p_+(x)$ is the probability for x to belong to a human silhouette.
$\Pi_+(x)$ is the proportion of trees voting for the class "human".
n_+ is the total amount of human pixels in the learning set.
n_- is the total amount of non-human pixels in the learning set.

- We assume that the human silhouettes and the non-human silhouettes occur at the same frequency. In this case, we propose the following estimator for $p_+(x)$:

$$\widehat{p_+(x)} = \frac{n_- \Pi_+(x)}{n_+ + (n_- - n_+)\Pi_+(x)} \tag{1}$$

Using the ExtRaTrees to implement the experts (2/2)

- Following Bayes' decision rule, the class of a pixel x is

$$\mathrm{y}(x) = \mathrm{sign}\left(\widehat{p_+(x)} - \frac{1}{2}\right) \tag{2}$$

- The probability for this decision to be correct is given by

$$\mathrm{r}(x) = \frac{1}{2} + \left|\widehat{p_+(x)} - \frac{1}{2}\right| \tag{3}$$

- We use the following weighted decision rule for a silhouette s

$$\mathrm{y}(s) = \mathrm{sign}\left(\sum_{x \in s} \mathrm{y}(x)\, w(x)\right) \tag{4}$$

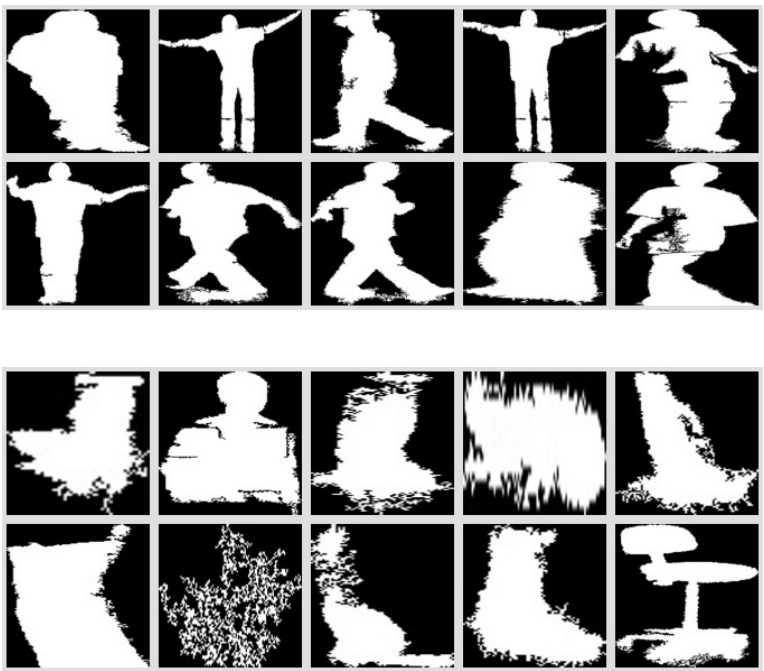
where $w(x)$ is a weight that depends on $\mathrm{r}(x)$.

Outline

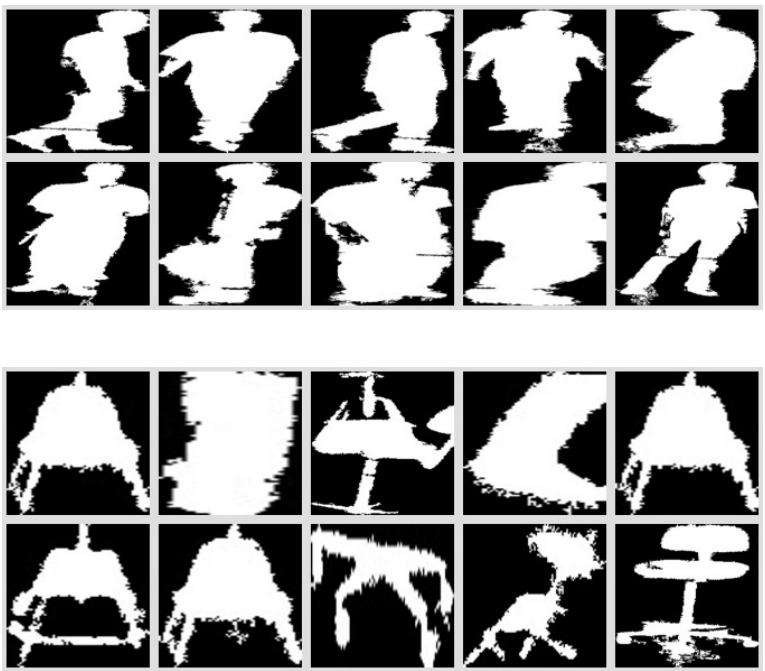
The datasets

- We used the same data as Barnich *et al.*.
- Silhouettes are resized to 100×100 pixels.

Learning set
(1245 +, 1924 -)

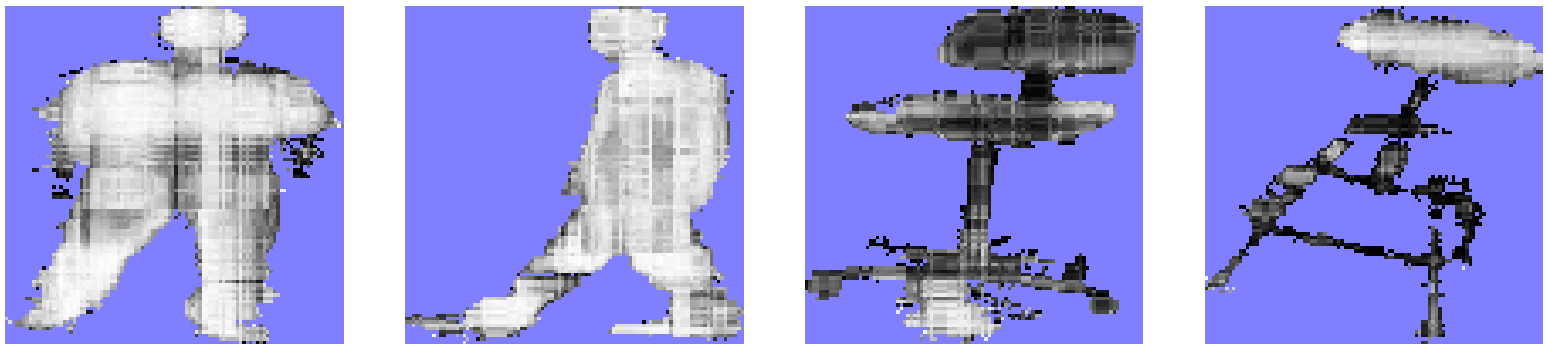


Test set
(1535 +, 1212 -)



Results

- Probability maps $\widehat{p_+(x)}$ can be drawn.



- We tried three different weighting strategies ($\mathrm{r}(x) \longmapsto w(x)$). All led to similar results (similar ROC curves).
- Correct classification rates around 90 % can be obtained for both human and non-human silhouettes !

Conclusions

We have proposed a new system for the detection of humans, suited for video streams. Our approach has been designed to rely on geometric information, and to be robust to noise.

1. A background subtraction algorithm extracts silhouettes.
2. A probabilistic information $p_+(\cdot)$ is estimated for each pixel.
3. This information is used to decide whether the silhouette is that of a human.

The results show that our approach is promising for the detection of humans in video streams. Finding the optimal neighborhood used for the description of a pixel is left for future work.