Contents

Adaboost Approach to Detection of Motion Patterns

- image-based and motion-based information are used simultaneously
- detection of short-term motion patterns rather than on tracking over extended periods of time
- closely related to the Adaboost object detection

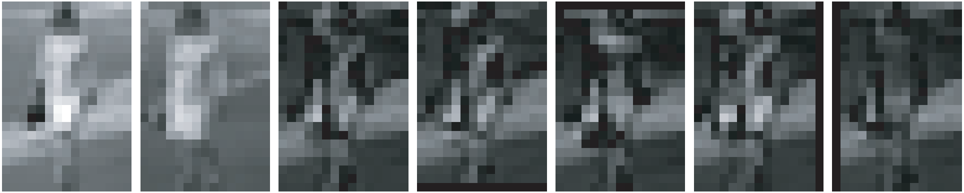
- pedestrian motion
- small set of simple rectangle filters trained on a set of examples
- multiscale
- filters work with short temporal image sequences

- motion detected as temporal differences in corresponding image blocks
- size of blocks — analysis scale
- blocks of different sizes

- computational efficiency
- motion direction derived from differences between shifted image blocks
- shift by ψ pixels – defined with respect to detection scale
- image frame acquired at time t, $t + \delta t$

- Five highly relevant

$$
\begin{aligned}
\Delta &= \mathrm{abs}(I_t - I_{t+1})\ , \\
U &= \mathrm{abs}(I_t - I_{t+\delta t} \uparrow)\ , \\
D &= \mathrm{abs}(I_t - I_{t+\delta t} \downarrow)\ , \\
L &= \mathrm{abs}(I_t - I_{t+\delta t} \leftarrow)\ , \\
R &= \mathrm{abs}(I_t - I_{t+\delta t} \rightarrow)\ ,
\end{aligned}
\tag{16.37}
$$



(a) I_t (b) $I_{t+\delta t}$ (c) Δ (d) U (e) D (f) L (g) R

Figure 16.12: Motion and appearance difference images derived according to equation (16.37). Image R has the lowest energy and as such, corresponds to the right-to-left direction of motion.

Filters f_k measure magnitude of motion

$$f_k = r_k(S) \tag{16.38}$$

Several filter types ...

$$f_i = r_i(\Delta) - r_i(S) \tag{16.39}$$

... likelihood that region is moving in a tested direction $\uparrow, \downarrow, \leftarrow$, or $\rightarrow$

S is one of the difference images $\{U, D, L, R\}$

r_i is a single rectangle sum within the detection window.

Motion shear can be determined using filters

$$f_j = \phi_j(S) \tag{16.40}$$

Filters f_m ... detecting image patterns of expected static image properties

$$f_m = \phi(I_t) \tag{16.41}$$

- filters $f_\bullet$ – from integral image
- filters $f_\bullet$ can be of any size, aspect ratio, or position
  (as long as they fit in image block)
- large number of filters
- best subset ... **to separate moving objects with motion-specific properties from the rest of the image**

- Classifier C – linear combination of selected features
- after AdaBoost training phase – thresholded sum of features

$$
\begin{aligned}
C(I_t, I_{t+\delta t}) &= 1 \quad \text{if} \quad \sum_{s=1}^{N} F_s\big(I_t, I_{t+\delta t}\big) > \theta\,, \\
&= 0 \quad \text{otherwise.}
\end{aligned}
\tag{16.42}
$$

- Feature F_s – thresholded image

$$F_s(I_t, I_{t+\delta t}) = \alpha \quad \text{if} \quad f_s(I_t, I_{t+\delta t}, \Delta, U, D, L, R) > t_s \ , \\ = \beta \quad \text{otherwise,} \tag{16.43}$$

- $t_s \in \mathcal{R}$ is a feature threshold
- f_s is one of filters $f_\bullet$

- N features f_s are selected using AdaBoost process from all considered filters
- these filters are a function of one or more parameters $I_t, I_{t+\delta t}, \Delta, U, D, L$, and/or R

- α, β, t_s, and θ – computed during the AdaBoost training process

- each of N rounds chooses from the full set of motion and appearance features
- $\rightarrow$ a mix of features balancing the appearance and motion descriptors is selected

- motion-invariant detection of object motion speed is achieved via different shifts ψ
- obtained during training – scaling all training samples to a pre-determined base resolution (i.e., bounding block size with respect to the pixel counts in the x and y directions)
- e.g., base resolution of 20 × 15 pixels was used by Viola/Jones
- multi-scale behavior achieved by operating on image pyramids

$$
\begin{aligned}
\Delta^l &= \mathrm{abs}(I_t^l - I_{t+1}^l) , \\
U^l &= \mathrm{abs}(I_t^l - I_{t+\delta t}^l \uparrow) , \\
D^l &= \mathrm{abs}(I_t^l - I_{t+\delta t}^l \downarrow) , \\
L^l &= \mathrm{abs}(I_t^l - I_{t+\delta t}^l \leftarrow) , \\
R^l &= \mathrm{abs}(I_t^l - I_{t+\delta t}^l \rightarrow) ,
\end{aligned}
\tag{16.44}
$$

- l ... pyramid level

- features computed from the pyramidal representations in a scale-invariant fashion
- scale factor of 0.8 for successive pyramid levels shown to work – all the way down to the pre-determined size of the base-resolution image block (20 × 15 pixels in the discussed case).

- Once features selected, a boosted cascade of classifiers
- Simple classifiers with high detection rates and relatively high false positive rates are employed in early stages
- More complex classifiers using larger numbers of features are used in the later cascade stages
- Each stage of the cascade attempts to reduce both the detection and the false positive rates
- $\rightarrow$ goal of reducing false positive rate more rapidly than detection rate

- Example application – pedestrians walking
- sequences of 2,000 frames
- Each of the cascade classifiers trained on 2,250 positive and 2,250 negative examples
- each example – two 20×15 image windows from two consecutive image frames ($\delta t = 1$)

- positive examples – scaled bounding boxes of pedestrians
- negative examples – no pedestrians
- feature selection – 54,624 filters

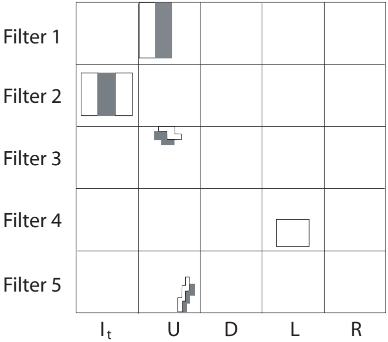


Figure 16.13: The first 5 features identified by the feature selection process for detecting walking pedestrians. The features reflect that the pedestrians were centered in the training images, tend to be different from the background, and four of them use the motion-difference images. *(Adapted from Viola03.)*

- motion information was crucial for the achieved performance
- dynamic pedestrian detector clearly outperformed the static pedestrian detector

Figure 16.14: Example results of pedestrian detection using the dynamic pedestrian detector. *Courtesy of P. Viola, Microsoft Live Labs and M. Jones, Mitsubishi Electric Research Labs, ©2003 IEEE [**?**].*