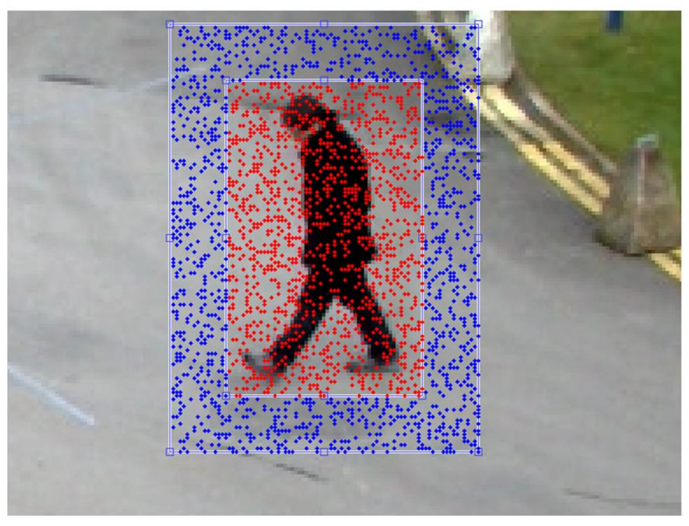
Haar-like Features and Adaptive Feature Extraction for Visual Tracking

Seunghoon Park, Adviser: Bohyung Han

A visual tracking algorithm was developed by integrating Haar-like features with adaptive feature extraction. The experiments show that the proposed tracker solves two knotty problems of visual tracking partially: varying illumination and occlusion.

Problem Description



A classification problem

- $\frac{N}{2}$ samples from target box
- $\frac{N}{2}$ samples from background box
- d-dimensional feature vector at each point

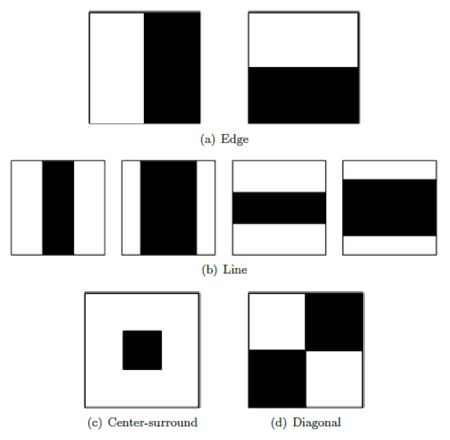
$$\mathbf{x}_i \in \mathbb{R}^d, \quad i = 1, 2, \ldots, N$$

- classifier F

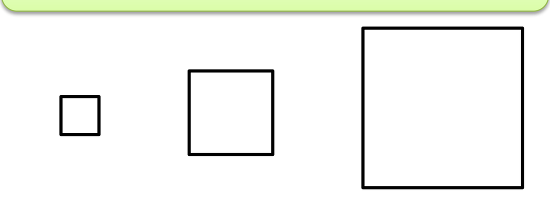
$$F : \mathbf{x} \rightarrow \begin{cases} +1, & \text{target box} \\ -1, & \text{background box} \end{cases}$$

24 Haar-like Features

8 Prototypes



3 Different Sizes



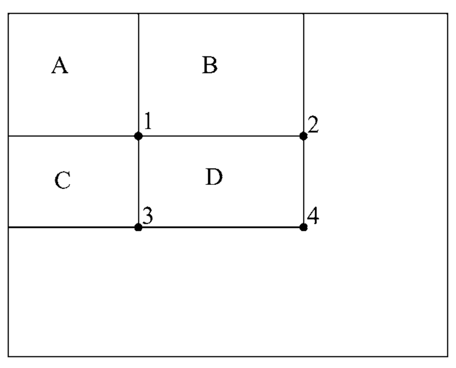
5x5 11x11 21x21

Feature Value

$$\sum_{(r,c)\in \text{white rectangle}} \text{pixel}(r,c) - \sum_{(r,c)\in \text{black rectangle}} \text{pixel}(r,c)$$

Integral Image

- aka summed area table
- can speed up the computation of feature values
- The sum of a rectangle can be achieved through only 4 memory access



D = (4+1) − (2+3)

Ensemble Tracking

Weak Classifier

- A weak classifier: $h(\mathbf{x}) = sign(\mathbf{h}^T\mathbf{x})$

where $\mathbf{h}$ is a hyperplane computed using weighted least square regression:

$$\mathbf{h} = (\mathbf{A}^T\mathbf{W}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}^T\mathbf{W}\mathbf{y}$$

- $\mathbf{A}$: $N \times (d+1)$ matrix whose each row is $[\mathbf{x}_i, 1]$
- $\mathbf{w}$: $N \times N$ diagonal matrix whose diagonal element is w_i, weight for each sample
- $\mathbf{y}$: N labels, $y_i \in \{+1, -1\}$

Strong Classifier

- Consists of T weak classifiers
- A strong classifier: $H(\mathbf{x}) = \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x})$
- Weight of each weak classifier: $\alpha_t = \frac{1}{2} log \frac{1 - err}{err}$

where $err = \sum_{i=1}^{N} w_i |h_t(\mathbf{x}_i) - y_i|$

- Whenever a weak classifier is made, weights of samples are updated: $w_i = w_i e^{(\alpha_t |h_t(\mathbf{x}_i) - y_i|)}$

Mean-shift

- Get a confidence map L

$$L(r,c) = H(\mathbf{x}(r,c)), \quad (r,c) \in \{\text{target box}\} \cup \{\text{background box}\}$$

- The location changes of boxes in each iteration

$$\Delta r = \frac{\sum_{(r_i,c_i)\in L} L(r_i,c_i) \times (r_i - r)}{\sum_{(r_i,c_i)\in L} L(r_i,c_i)}, \quad \Delta c = \frac{\sum_{(r_i,c_i)\in L} L(r_i,c_i) \times (c_i - c)}{\sum_{(r_i,c_i)\in L} L(r_i,c_i)}$$

- Repeated until the new location converges
- New N samples in the next frame
- Re-label new samples based on the new location

Update

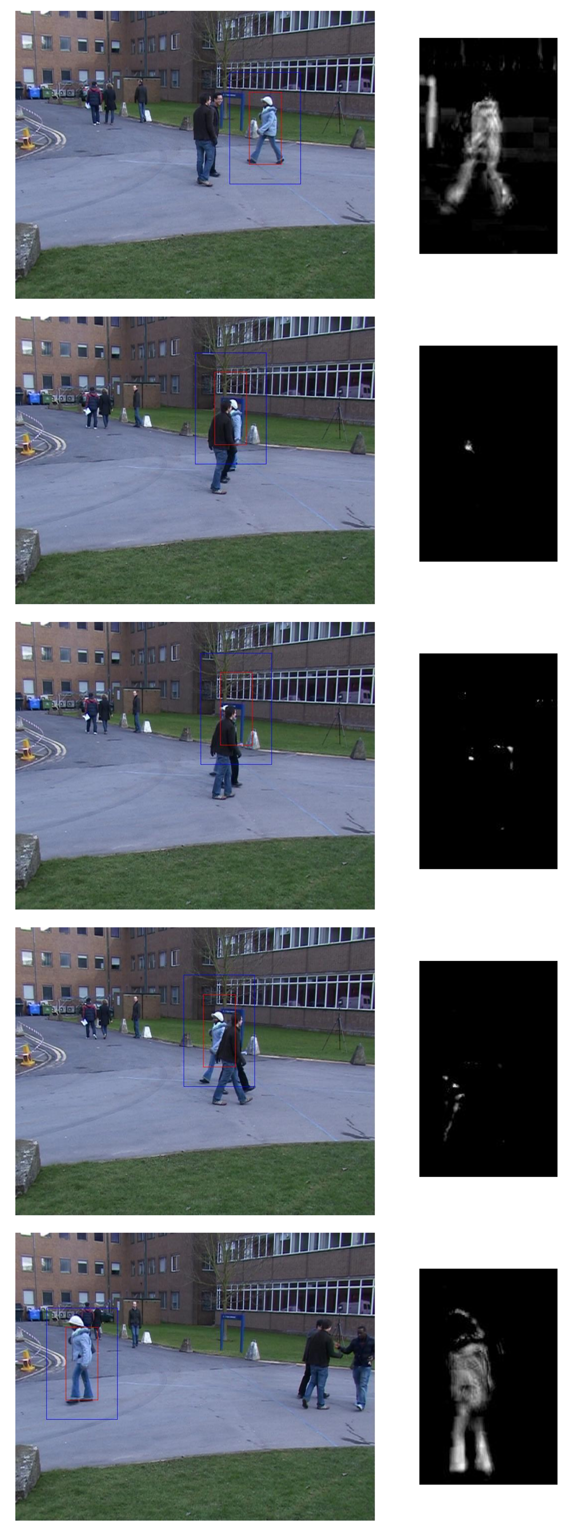
- Remain K (< T) weak classifiers to be robust to an occlusion
- A weak classifier which has the minimal err is selected during K iterations
- Make new T-K weak classifiers

Experimental Results

Short Occlusion

A successful tracking in case of a short occlusion

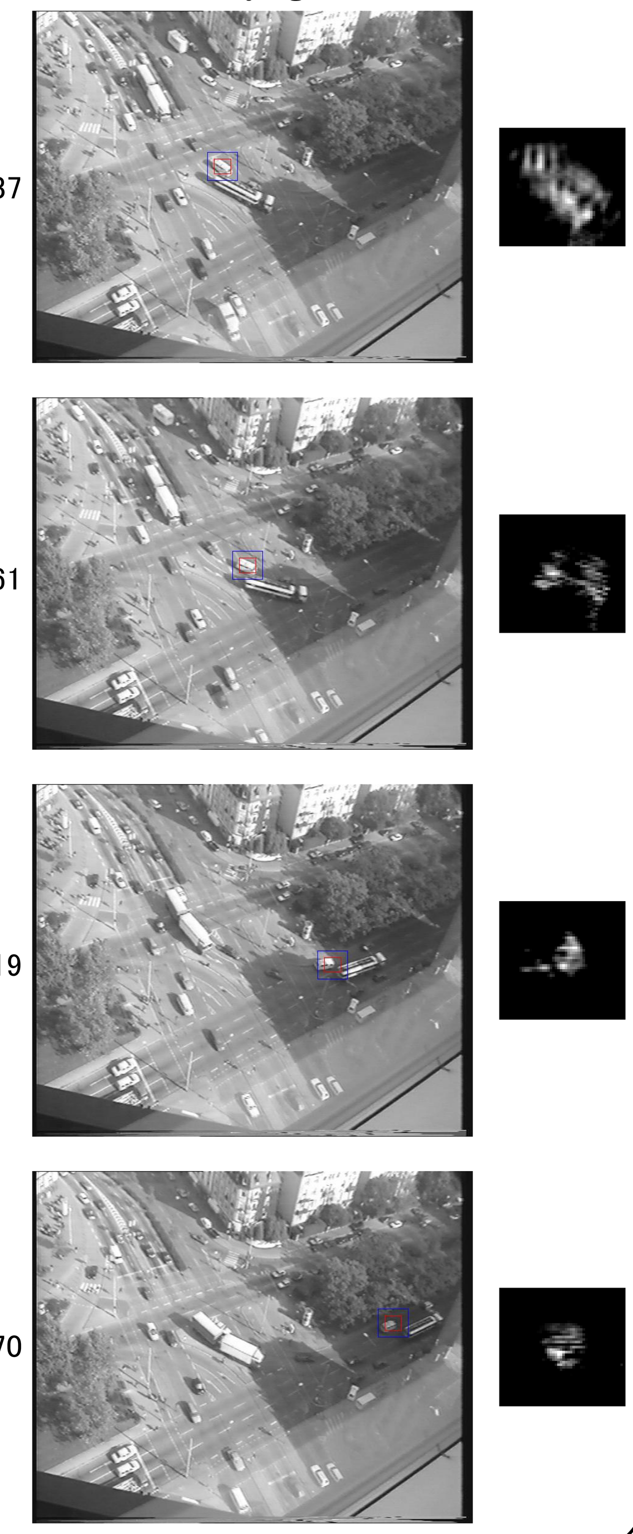
Frame 139
Frame 144
Frame 145
Frame 146
Frame 162



Varying Illumination

A successful tracking in case of varying illumination

Frame 137
Frame 161
Frame 219
Frame 270

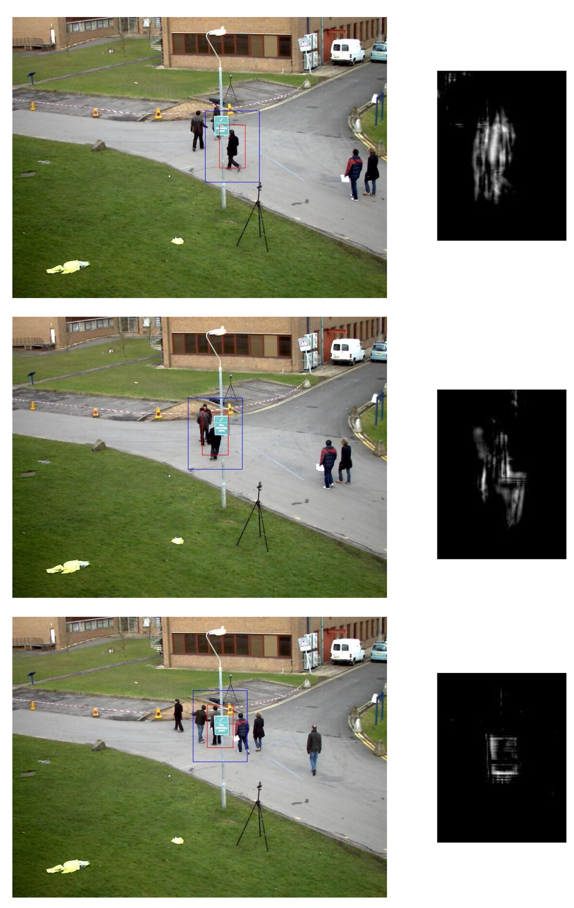


Limitations

Long Occlusion

Get stuck with an obstacle

Frame 21
Frame 30
Frame 65



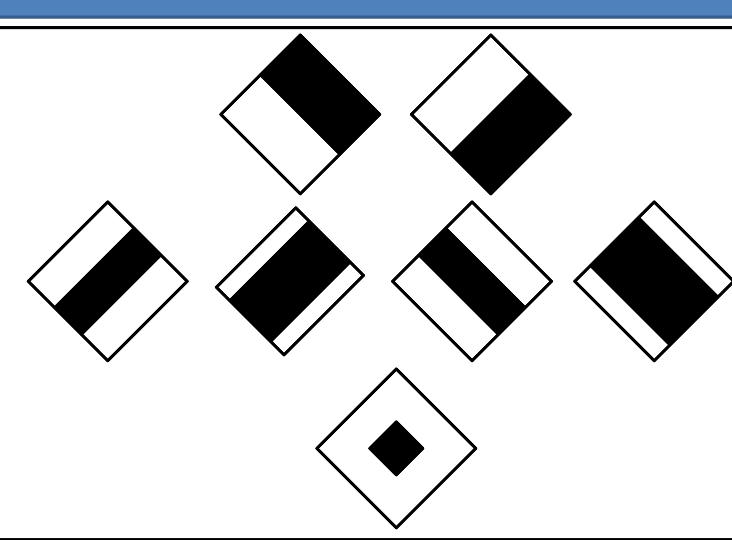
Edge Phenomena

Look like the edge of the target sometimes

Frame 169
Frame 186
Frame 188

Future works

- Rotated Haar-like features
- Initialization of tracking
- Varying size of target



Conclusion

◆ The combination of Haar-like features and ensemble tracking can improve the performance of a visual tracker in terms of a short occlusion and varying illumination.

◆ More works are necessary to succeed in tracking in case of a long occlusion and to add extensions such as initialization of tracking and varying size of target.