Real-Time and Event-Triggered Object Detection Recognition, and Tracking

Désirée Blizzard, Somayeh Davar& Arash Mohammadi
Concordia Institute for Information System Engineering, Montreal, QC, Canada

Intelligent Signal And Information Processing

Motivations/Objectives

Motivation:

- This work is motivated by the urgent need for development of an Event-Based surveillance system for real-time detection, recognition, and tracking of a Person-of-Interest (PoI) using an IP PTZ network camera.

Objectives:

- Design/Formulate Event-based Tracking problem within (non-linear) Bayesian filtering framework in combination with Convolutional Neural Networks (CNN).
- Fuse/Combine recently developed *Event-based Tracking* methodologies with *Deep Learning* mechanisms to be implemented within a practical setting.

Contribution:

- Develop/Implement a *Non-linear, Event-based, and Adaptive* tracking framework for identifying the PoI based on the given signature, adopted CNNs, and particle filters.

Material And Method

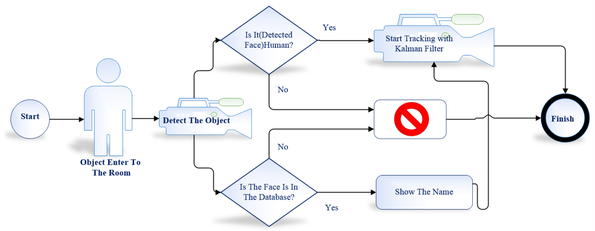

Figure 1: Block diagram of the Proposed Framework.

- A training set is prepared consisting of N images $\{\boldsymbol{P}_i, l_i\}$, for, $(1 \leq i \leq N)$, where $\boldsymbol{P}_i \in \mathbb{R}^3$ represents the input image and l_i denotes its associated label indicating whether or not $\boldsymbol{P}_i$ contains the PoI.
- An initial detection algorithm such as aggregate channel features (ACF) is used to compute proposals for the reference region ($\mathcal{B}^*$).
- We model the detected PoI in frame $\boldsymbol{P}_i$ as the state vector denoted by $\boldsymbol{x}_i$.
- The proposals computed by the initial detection module are provided as input to the CNN which provides as an output the probability of having the PoI in the given image frame.
- We consider a typical CNN structure consisting of convolutional layers equipped with a non-linear activation function, sub-sampling layers, fully connected layers, and a logistic regression layer.
- The output of the CNN module is given by:

$$p_i = f(\boldsymbol{x}_i(\mathcal{B}^*), \boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ denotes the parameters of the CNN.

IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Boston, MA, USA, 2017.

Facial Recognition

- We model the detected PoI as the state vector denoted by $\boldsymbol{x}_k$ and adopt a very weak model $\boldsymbol{F}_k$ for evolution of the state over time k.
- Establishing a database of the PoIs where for each identified PoI, several pictures are taken indicating a range of facial expressions.
- Viola-Jones algorithm is used for face detection. Each picture is changed to gray scale and resized such that only 150-by-150 image of the face remains.

Figure 2: Facial Detection Recognition

- Facial detection module is launched to detect the face in a given frame by classifying the extracted features via support vector machines (SVM) and a predict function.

Histogram of Gradients

- The histogram of oriented gradients (HOG) features are extracted from each face in our databases.
- This HOG characterizes the captured image by separating it in a large quantity of cells and quantifying its local intensity gradients and edge detection.

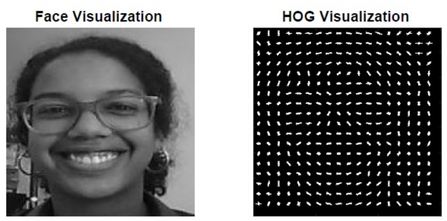

Figure 3: Extraction of the (HOG) features

- Another cell vector is initialized at the same time, storing the names of the subjects.

Implemented Event-based Tracking Framework

Implemented Convolutional Neural Networks (CNN):

- In this project, we use a minimum probability score of 30 percent to filter out false-positives.
- To improve the accuracy of human detection algorithm, CNNs are used on the propositions made by the ACF detector.
- Four main layers are considered in the incorporated CNN, namely, convolutional layer, Rectifier Linear Unit (RELU) layer, pooling layer, and the fully-connect layer.
- If the CNN module confirms the input detection score, an event is triggered and the image is provided to the tracking module. Otherwise, the proposals are eliminated to reduce the number of false positives.
- Tracking is based on Nonlinear Particle Filtering approach (Fig. 4) where constant velocity model is considered to describe trajectory of the PoI.

Experimental Setup:

- We use an IP PTZ network camera of model DCS-5010L.
- We use/adopt a pre-trained ACF detector which is trained based on the INRIA pedestrian database.
- Identification of the Faces: The face is detected in the frame with name of POI. This method characterizes the captured image by separating it in a large quantity of cells.
- The detector's results have to be filtered out depending on their probabilistic score which determines accuracy of human prediction process.

Figure 4: POI tracking using the Kalman Filter

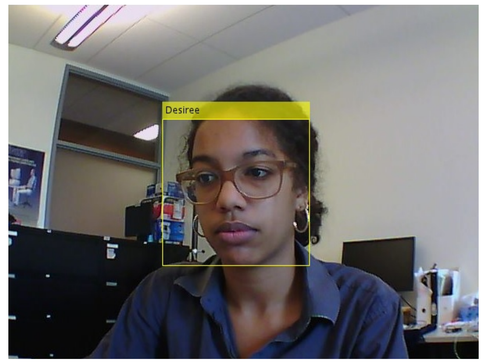

Figure 5: Facial Detection Recognition with PoI

Conclusion/Future Research

- This content outlines our ongoing research work for *practical implementation of an event-based surveillance framework*, for *real-time detection and tracking of person(s)-of-interest* via a connected group of network cameras in an indoor environment.
- We considered *detect-then-track strategy* where within tracking module we used recursive Bayesian formulation.
- Our main contribution is cascade of detection sub-systems and particle filtering where *(i) A fast detection scheme with reduced computational complexity initiates the process; (ii) An adopted CNN improves the detection accuracy, and; (iii) A non-linear filter performs the tracking task.*

Future Research:

- The ultimate goal is large-scale implementation of event-triggering estimation/tracking framework with learning capabilities in a practical setting for *real-time Recognition, detection and tracking* via networked IP cameras.

References

[1] A. Mohammadi and K.N. Plataniotis, "Event-Based Estimation With Information-Based Triggering and Unsupervised Update," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp 4924-4939, 2017.

[2] S. Davar and A. Mohammadi, "Event-based Particle Filtering with Point and Set-valued Measurements," 25th European Signal Processing Conference (EUSIPCO), 2017, In Press.