Generic Object Crowd Tracking by Multi-Task Learning

Wenhan Luo
http://www.iis.ee.ic.ac.uk/~whluo
Tae-Kyun Kim
http://www.iis.ee.ic.ac.uk/~tkkim

Department of Electrical and Electronic Engineering,
Imperial College, London, UK

Multiple Object Tracking (MOT) is an important problem for its various applications. In general, approaches for MOT can be categorised into two types, sequential ones and batch ones. Sequential ones utilise observations from frames up to the time while batch ones use observations from all frames of a video. For the significant progresses achieved in the pedestrian detection field, most existing work for MOT follows the batch fashion (i.e. it employs a pedestrian detector to carry out detection in each frame in advance, and then handles MOT as a data association problem by treating the detection responses of all the frames as observations). However, little attention has been paid to detection and tracking of multiple objects of an arbitrary type. In this paper, we tackle the problem of tracking multiple objects without limitation of the type of the objects. Furthermore, we show how this problem can be formulated within the Multiple Task Learning (MTL) framework [3].

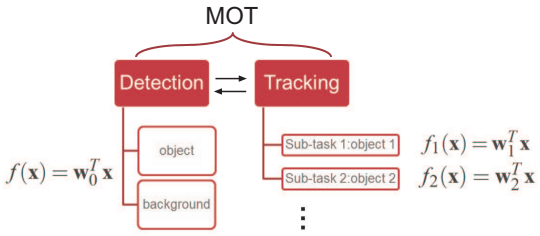

Figure 1: Task decomposition of our MOT problem.

By adopting the strategy of tracking-by-detection, we decompose our problem into two main tasks, detection and tracking, and further decompose the tracking task into multiple sub-tasks, each of which corresponds to tracking an individual object. Fig. 1 shows the decomposition of our task. In the stage of fulfilling the detection task, a binary detector is learnt to detect objects in images. For the tracking task, multiple trackers are learnt on top of the detector to trace detected objects in subsequent frames. To prevent trackers from drifting away from targets, the detector is utilised to anchor the trackers by the proposed Mean Regularised Joint Feature Learning algorithm. At the same time, the trackers are jointly learnt by sharing common features to capture the relatedness among multiple tasks. To further improve the performance, we use a smoothness term which globally considers all the labelled and unlabelled data. In the following, details of detection and tracking will be given.

The detection task is completed by posing it as a linear Laplacian SVM [2] optimisation problem. We firstly reject some sliding windows which are impossible to be objects by the objectness measurement [1] to accelerate the whole procedure. Then we construct a graph treating all the samples $\mathbf{X}$ as vertices and similarities among objects as edges. Let us write the detector as $f(\mathbf{x}) = \mathbf{w}_0^T\mathbf{x}$, then we have the following object function to minimise,

$$\begin{aligned} \min_{\mathbf{w}_0,\varepsilon_i} \quad & \gamma_1\|\mathbf{w}_0\|^2 + \gamma_2\mathbf{w}_0^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{w}_0 + \gamma_3\sum_{i=1}^{n_l}\varepsilon_i \\ s.t. \quad & y_i\mathbf{w}_0^T\mathbf{x} \geq 1-\varepsilon_i,\ i=1,2,...,n_l \\ & \varepsilon_i \geq 0,\ i=1,2,...,n_l \end{aligned} \tag{1}$$

where $\mathbf{L}$ is the Laplacian matrix calculated from the graph and $\{\varepsilon_i,\ i=1,2,...,n_l\}$ are slack variables. Following the primal-dual formulation, we derive a quadratic optimisation problem as,

$$\begin{aligned} \max_{\alpha\in\mathbb{R}^{n_l}} \quad & \sum_{i=1}^{n_l}\alpha_i - \frac{1}{2}\alpha^T\mathbf{Q}\alpha \\ s.t. \quad & 0 \leq \alpha_i \leq \gamma_3,\ i=1,2,...,n_l \end{aligned} \tag{2}$$

where $\mathbf{Q} = \mathbf{Y}^T\mathbf{J}^T\mathbf{X}^T(2\gamma_1\mathbf{I}+2\gamma_2\mathbf{L}\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{J}\mathbf{Y}$, $\mathbf{J} = [\mathbf{I}\ \mathbf{0}]^T$ is a $n\times n_l$ matrix with $\mathbf{I}$ as the $n_l\times n_l$ identity matrix, $\mathbf{Y} = diag(y_1,...,y_{n_l}) \in \mathbb{R}^{n_l\times n_l}$ and $\alpha = [\alpha_1,...,\alpha_{n_l}]^T \in \mathbb{R}^{n_l}$ are Lagrangian multipliers. Solving this problem we can obtain α, and the detector is $\mathbf{w}_0 = (2\gamma_1\mathbf{I}+2\gamma_2\mathbf{L}\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{J}\mathbf{Y}\alpha$.

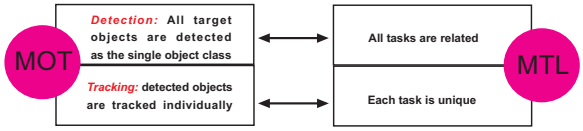

Figure 2: Formulation of the MOT problem into MTL.

To track multiple objects, we formulate the problem within MTL framework. The motivation for it can be illustrated by Fig 2. Writing each tracker as $f_t(\mathbf{x}) = \mathbf{w}_t^T\mathbf{x}$ and all the trackers $[\mathbf{w}_1,\cdots,\mathbf{w}_T]$ as a matrix $\mathbf{W}\in\mathbb{R}^{d\times T}$, we propose the Mean Regularised Joint Feature Learning algorithm which minimises the following objective function with regard to $\mathbf{W}$,

$$\begin{aligned} \min_{\mathbf{W}\in\mathbb{R}^{d\times T}} \frac{1}{2}\sum_{t=1}^{T}\|\mathbf{J}_t^T\mathbf{X}_t^T\mathbf{w}_t - \mathbf{Y}_t\|^2 + \rho_1\|\mathbf{W}\|_{2,1} + \frac{\rho_2}{2}\sum_{t=1}^{T}\|\mathbf{w}_t - \mathbf{w}_0\|^2 \\ + \frac{\rho_3}{2}\sum_{t=1}^{T}\mathbf{w}_t^T\mathbf{X}_t\mathbf{L}_t\mathbf{X}_t^T\mathbf{w}_t \end{aligned} \tag{3}$$

In the above function, the first term is the loss from labelled data of each task. In this term $\mathbf{X}_t$ is the combination of labelled samples and unlabelled samples for a sub-task t, $\mathbf{J}_t$ is a matrix which chooses only the labelled data to calculate the loss. $\mathbf{Y}_t$ is the label vector of the task t (we give the neutral label 0 to the unlabelled data). The second term is the joint feature learning term. We learn the features shared by multiple tasks via the penalty of $\|\mathbf{W}\|_{2,1}$, the $\ell_{2,1}$ norm of $\mathbf{W}$. This regularisation term can result in that only some rows of $\mathbf{W}$ are non-zero, which correspond to the features shared by all sub-tasks. The third term is the regularisation to relate the two main tasks. This regularisation term benefits the trackers in two aspects. Firstly, as $\|\mathbf{w}_t\|^2 = \|\mathbf{w}_t-\mathbf{w}_0+\mathbf{w}_0\|^2 \leq \|\mathbf{w}_t-\mathbf{w}_0\|^2+\|\mathbf{w}_0\|^2$, and we have minimised $\|\mathbf{w}_0\|^2$ in the detection stage, thus minimising $\|\mathbf{w}_t-\mathbf{w}_0\|^2$ equals minimising $\|\mathbf{w}_t\|^2$, further improving the generalisation ability of each tracker. Secondly, this term can prevent trackers from drifting to the background as we enforce each tracker to be close to the detector. The last term is the combination of smoothness for each tracker/task. $\mathbf{L}_t$ is the Laplacian matrix associated with the graph of the task t. This smoothness term enables the tracker to view the labelled and unlabelled samples (candidates) together. This composite optimisation problem can be solved by adopting the Accelerated Gradient Method (AGM) [4].

We test our approach on four challenging data sets (including a publicly available data set) to evaluate its performance. Results compared with our self baselines and some other counterparts show that the proposed method significantly outperforms the state-of-the-art methods.

Our conclusion is that by decomposing our problem into two main tasks and representing their relation via the proposed Mean Regularised Joint Feature Learning algorithm, we can effectively derive the desired list of detected and tracked objects in frames.

[1] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *CVPR*, 2010.

[2] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR*, 7:2399–2434, 2006.

[3] R. Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

[4] Y. Nesterov. Gradient methods for minimizing composite objective function. core discussion papers 2007076, université catholique de louvain. *Center for Operations Research and Econometrics (CORE)*, 2007.