

Long Term Connectivity in CRF-based Multi-Person Tracking

Bhargavi K M
 Metechde branch
 T-Johns Institute of Technology
 Bangalore, India

Abstract:- This paper explores an alternative approach that relies on longer-term connectivities between pairs of detections for multi-person tracking. We formulate tracking as a labeling problem in a Conditional Random Field (CRF) framework, where we target the minimization of an energy function defined upon pairs of detections and labels. We presented a CRF model for detection-based multi-person tracking. Contrarily to other methods, it exploits longer-term connectivities between pairs of detections. Moreover, it relies on pairwise similarity and dissimilarity factors defined at the detection level, based on position, color and also visual motion cues, along with a feature-specific factor weighting scheme that accounts for feature reliability. The model also incorporates a label field prior penalizing unrealistic solutions, leveraging on track and scene characteristics like duration and start/end zones.

1. INTRODUCTION

QR code Automated tracking of multiple people is a central problem in computer vision. It is particularly interesting in video surveillance contexts, where tracking the position of people over time might benefit tasks such as group and social behavior analysis, pose estimation or abnormality detection, to name a few. Nonetheless, multi-person tracking remains a challenging task, especially in single camera settings, notably due to sensor noise, changing backgrounds, high crowding, occlusions, clutter and appearance similarity between individuals.

Tracking-by-detection methods have become increasingly popular. These methods aim at automatically associating human detections across frames, such that each set of associated detections

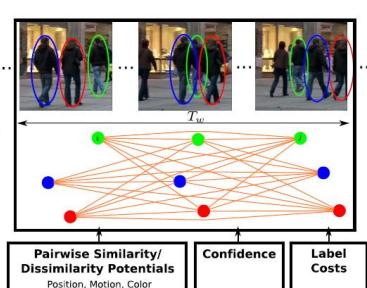
two nodes within the shown temporal window are displayed). Pairwise feature similarity/dissimilarity potentials, confidence scores and label costs are used to build the energy function to optimize for solving the labelling problem within the proposed CRF framework.

Compared to background modelling-based approaches, tracking-by-detection is more robust to changing backgrounds and moving cameras. However, human detection is not without weaknesses: detectors usually produce false alarms and they miss detect objects. Several existing approaches address these issues by initially linking detections with high confidence to build track fragments or tracklets, and then finding an optimal association of such tracklets. Although obtaining impressive results on several datasets, these approaches ultimately rely on low-level associations that are limited to neighboring time instants and reduced sets of features (color and adjacency). Hence, a number of higher-level refinements with different sets of features and tracklet representations are required in order to associate tracklets into longer trajectories.

Here we explore an alternative approach that relies on longer-term connectivities between pairs of detections for multi-person tracking. We formulate tracking as a labeling problem in a Conditional Random Field (CRF) framework, where we target the minimization of an energy function defined upon pairs of detections and labels. Our approach is summarized contrarily to existing approaches, the pairwise links between detections are not limited to detections pairs in adjacent frames, but between frames within a time interval T_w . Hence, the notion of tracklets is not explicitly needed to compute features for tracking, allowing us to keep the optimization at the detection level. One important advantage of our modelling scheme is that it allows to directly learn the pairwise potential parameters from the data in an unsupervised and incremental fashion. To that end, we propose a criterion to first collect relevant detection pairs to measure their similarity/dissimilarity statistics and learn model parameters that are sensitive to the time interval between detection pairs. Then, at a successive optimization round, we can leverage on intermediate track information to gather more reliable statistics and exploit them to estimate accurate model parameters.

1.1 Objective and scope of the project

The Automated tracking of multiple people is a central problem in computer vision. It is particularly interesting in video surveillance contexts, where tracking the



univocally belongs to one individual in the scene.

Detections in incoming frames are represented as observation nodes. Pairs of labels/observations within a temporal window are linked to form the labelling graph, thus exploiting longer-term connectivities (note: for clarity, only links having their

position of people over time might benefit tasks such as group and social behaviour analysis, pose estimation or abnormality detection, to name a few. Nonetheless, multi-person tracking remains a challenging task, especially in single camera settings, notably due to sensor noise, changing backgrounds, high crowding, occlusions, clutter and appearance similarity between individuals

Used in

- video surveillance
- group and social behavior analysis
- pose estimation or abnormality detection

2. PROBLEM DEFINITION

- Compared to some existing CRF approach for tracking a novel aspect of existing framework is that the energy function includes higher order terms in the form of label costs.
- In this tracking framework, this translates into penalizing the complexity of the labeling, mostly based on the fact that sufficiently long tracks should start and end in specific areas of the scenario.
- In this existing system it is difficult to track the person with same label or identity, whenever the person moved into the hidden areas or during occlusion.

3. EXISTING SYSTEM

The existing approaches address issues by initially linking detections with high confidence to build track fragments or tracklets and then finding an optimal association of such tracklets. Although obtaining impressive results on several datasets, these approaches ultimately rely on low-level associations that are limited to neighboring time instants and reduced sets of features (color and adjacency). Hence, a number of higher-level refinements with different sets of features and tracklet representations are required in order to associate tracklets into longer trajectories. Here tracking is a labeling problem in a Conditional Random Field (CRF) framework, where we target the minimization of an energy function defined upon pairs of detections and labels. Contrarily to existing approaches, the pairwise links between detections are not limited to pairs of detections in adjacent frames, but between frames within a time interval.

To summarize, the project addresses the multi-person tracking problem within a tracking-by-detection approach and makes contributions in the following directions:

- 1) A CRF framework formulated in terms of similarity/ dissimilarity pairwise factors between detections and additional higher-order potentials defined in terms of label costs. Differently from existing CRF frameworks, our method considers long-term connectivity between pairs of detections. Note however that long-term temporal connectivity alone is generally not sufficient to guarantee good results, and needs to be exploited in conjunction with the other contributions described below: visual motion, confidence weights, time-sensitive parameters with unsupervised learning from tracklets.

2) A novel potential based on visual motion features. Visual motion allows incorporating motion cues at the bottom association level, i.e., the detection level, rather than through tracklet hypothesizing.

3) A set of confidence scores for each feature-based potential and pair of detections. The proposed confidence scores model the reliability of the feature considering spatio-temporal reasoning such as occlusions between detections.

4) In similarity/dissimilarity formulation, the parameters defining the pairwise factors can be learned in an unsupervised fashion from detections or from tracklets, leading to accurate time-interval dependent factor terms.

4. PROPOSED SYSTEM

Sparse Gaussian Conditional Random Fields:

4.1 How to overcome problems

We propose a new second-order active set method for solving the sparse Gaussian CRF. Such algorithms have previously been applied to the Gaussian MRF and a general analysis of such methods (showing quadratic convergence) is presented. The method here largely mirrors in the approach for the Gaussian MRF, but the precise formulation is significantly more involved, owing to the complexity of gradient term in the likelihood. Despite being a second-order method, we show that the resulting algorithm is faster (to reach any accuracy) than previously proposed approaches, and several orders of magnitude faster at achieving solutions to high accuracy. The sparse Gaussian conditional random field enjoys many benefits of existing methods for learning high dimensional Gaussian graphical models; we believe that the advances put forward in this project make the model significantly more practical for large-scale problems, and also significantly advance our theoretical understanding of the method. Furthermore, the empirical results presented here on tracking of objects and humans using image processing.

4.2 Advantages of proposed system

Furthermore, to take into account not only the actual feature distance value but also its reliability, we exploit a set of confidence scores per feature to characterize how trustable the pairwise distances are.

These scores ultimately allow to re-weight the contribution of each feature based on spatio-temporal cues, and to rely on the most reliable pairwise links for labeling.

This is important near occlusion situations, where thanks to long-term connectivity, the labeling can count on cleaner detections just before or after occlusion to propagate labels directly to the noisier detections obtained during occlusion instead of through adjacent drift-prone frame to-frame pairwise links only. And Low Label costs.

Finally, a significant advantage of the sparse Gaussian CRF approach is that the Sparsity pattern of the resulting model can be interpreted directly as conditional dependencies between variables, and thus the Sparsity pattern itself can be very informative.

- This proposed SGCRF can take more size of inputs i.e. in the form of video sequence.
- Accurate results can be obtained even if the persons to be tracked are hidden or occluded and thus providing long term connectivity.
- Even if we discard some frames for detection we can get the accurate results by detecting the target person and thus the complexity of energy function can be reduced.

5. CONCLUSION

This paper presents a SGCRF model for detection-based multi-person tracking. Contrarily to other methods, it exploits longer-term connectivities between pairs of detections. Moreover, it relies on pair wise similarity and dissimilarity factors defined at the detection level, based on position, color and also visual motion cues, along with a feature-specific factor weighting scheme that accounts for feature reliability. The model also incorporates a label field prior penalizing unrealistic solutions, leveraging on track and scene characteristics like duration and start/end zones. Experiments on public datasets and comparisons with state-of-the-art approaches validated the different modeling steps, such as the use of a long time horizon T_w with a higher density of connections that better constrains the models and provides more pair wise comparisons to assess the labeling, or an unsupervised learning scheme of time-interval sensitive model parameters.

6. REFERENCES

- [1] A. Andriyenko and K. Schindler, "Multi-target tracking by continuous energy minimization," in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 1265–1272.
- [2] A. Andriyenko, K. Schindler, and S. Roth, "Discrete-continuous optimization for multi-target tracking," in Proc. IEEE Conf. CVPR, Jun. 2012, pp. 1926–1933.
- [3] S. Bak, D. P. Chau, J. Badie, E. Corvee, F. Bremond, and M. Thonnat, "Multi-target tracking by discriminative analysis on Riemannian manifold," in Proc. IEEE ICIP, Sep./Oct. 2012, pp. 1605–1608.
- [4] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 3457–3464.
- [5] J. Berclaz, F. Fleuret, and P. Fua, "Multi-camera tracking and atypical motion detection with behavioral maps," in Proc. ECCV, 2008, pp. 112–125.
- [6] J. Berclaz, F. Fleuret, and P. Fua, "Multiple object tracking using flow linear programming," in Proc. 12th IEEE Int. Workshop Winter-PETS, Dec. 2009, pp. 1–8.