In 16th Portuguese Conference on Pattern Recognition (RecPad 2010)
Vila Real, Portugal, October 2010

Multiple Hypothesis Group Tracking in Video Sequences

David Antunes[1], David Martins de Matos[2], José Gaspar[3]

[1,2] L^2F - INESC-ID, R. Alves Redol 9, Lisboa, Portugal
[1,3] Institute for Systems and Robotics, IST/UTL, Av. Rovisco Pais 1, Lisboa, Portugal
[1] davidmiguel@antunes.net, [2] david.matos@l2f.inesc-id.pt, [3] jag@isr.ist.utl.pt

***Abstract*—**[123] **In this paper the Multiple Hypothesis Tracking with group tracking is applied to tracking targets in video sequences. The width/height ratio is used as a target feature to help in target/detection association. Group tracking is tested in real video sequences (PETS 2001) and shows that is possible to keep linked the otherwise interrupted tracks.**

I. Introduction

The Multiple Hypothesis Tracking (MHT) algorithm, proposed by Reid [6], is fundamental in the field of Multi-Target Tracking [1]. In this paper we describe an application of the MHT enhanced with group tracking. Group tracking is necessary to maintain the tracks of several targets when they come together for several frames. The target $width/height$ ratio feature is used to help in the association problem. The MHT has also been applied, with group tracking, to track persons with a robot's range finder by [4].

II. Frame Processing

In order to detect targets in the frame we start by detecting active pixels as in [3]. To detect active pixels in frame I_i the last frame I_{i-1} is used. The difference frame I_D is calculated as $I_D = |I_i - I_{i-1}|$. A threshold T_1 is applied to I_D in order to generate a binary matrix I_B with active pixels. $I_B(x, y) = 1$ if $I_D(x, y) >= T_1$.

In order to avoid many false alarms, we apply morphological processing based on the local density of the active pixels. Each pixel of the filtered image, $I_M(x, y)$ is considered active if the number of active pixels in the neighborhood, $I_B(w(x, y))$ where $w(x, y)$ denotes the window centered at (x, y), is larger than a threshold. Agglomerative hierarchical clustering [2] is then applied to I_M to generate clusters of active pixels. Each cluster j is represented by its centroid, c_j which is termed a detection (measurement). We denote $Z(k) = \{c_j : j \in 1..N\}$ the set of measurements observed in frame k, and $Z^k = \{Z(1), Z(2), ..., Z(k)\}$ the set of all measurements up to time k.

III. Tracking

Target information includes its (x, y) position and, an additional feature, the $width/height$ (w/h) ratio, to help solve associations. No motion model is used to predict the position of targets in the future frames. The targets' gates are circular around the target. The number of the last frame where the target was detected is also maintained, for track termination. The MHT is implemented with clustering [6].

[123]hello

A. Probability evaluation

Let Ω_i^k denote an hypothesis at time k, where i is the hypothesis number. The probability of a new hypothesis in time k given the measurements Z^k is denoted by $P(\Omega_i^k|Z^k)$, and is calculated recursively as follows [5]:

$$P(\Omega_i^k|Z^k) = P(\psi_i^k|\Omega_p^{k-1}, Z(k)) \cdot P(\Omega_p^{k-1}|Z^{k-1}) \quad (1)$$

where Ω_p^{k-1} is the parent hypothesis, and ψ_i^k represents the associations in the current frame. The second term on the right-hand side (RHS) of the equation corresponds to the probability of the parent hypothesis in the hypothesis tree, and the first term on the RHS of the equation represents the probability of the assignments in the current hypothesis. The modeling of $P(\psi_i^k|\Omega_i^{k-1}, Z(k))$ is introduced in the following.

Considering a hypothesis Ω_i^{k-1} and a new set of detections $Z(k)$, each detection may have several different origins: an existing target was detected, the detection is a false detection, or a new target is detected. For each existing target: it may have not been detected in the current frame, it may be associated with a detection, or its track may be terminated (for lack of detections for several frames).

The probabilities of false detection, new target, and a target not being detected are constants, respectively, P_{FA}, P_{NT}, P_{MD}. Let $d_{M,T}$ be the distance between the detection D and the target T, the probability of association of a measurement D with a target T ($P_{D,T}$) is $1/d_{D,T}$, or 0 if the detection if not within the target's gate. A track is terminated if its target is not detected after N_{nd} frames. A track termination has no influence on the hypothesis probability. Let N_{FA}, N_{NT}, N_{MD} be the number of false alarms, new tracks, and missed target detections, then:

$$P(\psi_i^k|\Omega_i^{k-1}, Z(k)) = (P_{FA})^{N_{FA}} \cdot (P_{NT})^{N_{NT}} \cdot (P_{MD})^{N_{MT}} \cdot \prod_{(D,T) \in \psi_i^k} P_{D,T}$$

where (D, T) is an assignment of detection D to track T.

B. Group tracking

At any frame, any two targets or target groups may merge together. Only two targets or target groups are allowed to merge at each frame to prevent a combinatorial explosion of hypothesis. The groups are formed when several targets (or groups of targets) are associated with the same detection, so each detection may have been caused by one or more targets and groups of targets. At any frame any group may split into several subgroups (or single targets).

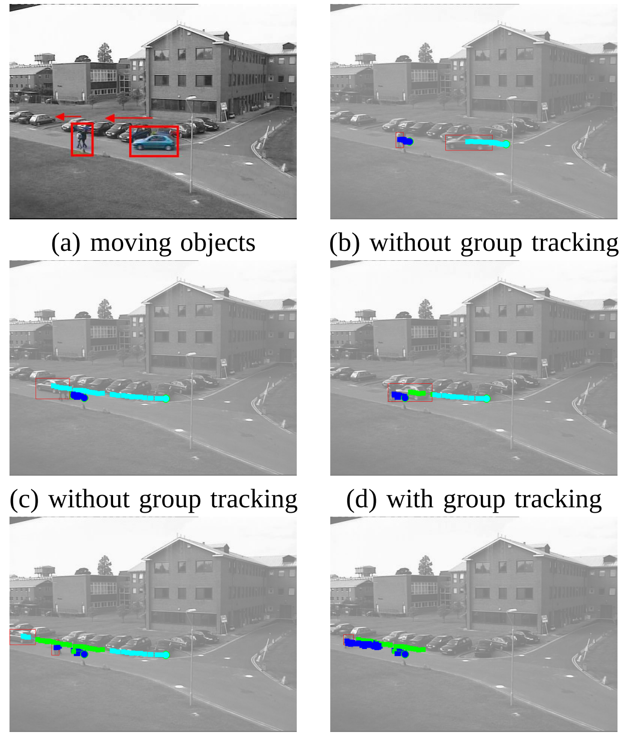
Group creation/merging and splitting have no direct influence in the hypothesis probability. The group operations which better explain the input data will be selected through greater probability. For example, if in Ω_i^{k-1} there where two close tracks and in $Z(t)$ there is only one measurement in that area, then the hypothesis that the two targets form a group will be selected. This hypothesis will have greater probability than, for example, one where only a target is detected.

C. Target Features

The $width/height$ ratio of each target is used to solve association problems. It is particularly useful in problems involving certain types of targets, most noticeably people and vehicles. Each target has a ratio which is only updated when the target is the sole responsible for a single detection (i.e. the target is not in a group) because a group of targets has a different ratio than each of its elements. In certain frames a target's ratio may suffer an abrupt change due to the process of cluster detection. To attenuate this effect the targets' ratio is calculated as a floating average.

In a hypothesis Ω_i^k, a change in a target's T ratio R^T modifies the hypothesis probability by $C_T = 1/(1 + |R^T_{\psi_i^k} - R^T_{\Omega_i^{k-1}}| \cdot F)$. The factor F adjusts the weight of ratio changes in hypotheses probability calculation. Then:

$$P(\psi_i^k|\Omega_i^{k-1}, Z(k)) = (P_{FA})^{N_{FA}} \cdot (P_{NT})^{N_{NT}} \cdot (P_{MD})^{N_{MT}} \cdot \prod_{(D,T)\in\psi_i^k} P_{D,T} \cdot C_T$$



(a) moving objects (b) without group tracking
(c) without group tracking (d) with group tracking
(e) with group tracking (f) with group tracking

Fig. 1: Comparing not-having vs having group tracking.

IV. Results

Without group tracking, when two targets get close together they are included in the same cluster of active pixels. They may remain in the same cluster for several frames. After some time of the two tracks being tracked, only one will remain active and the other track will be terminated for lack of detections. When the group separates, one of the targets will generate a new track.

Figure 1 shows an example tracking scene (a), with two pedestrians (they will be considered as a single target, as they are always close together) on the left, and a car on the right. In (b) and (c) group tracking is not enabled: when the two tracks come together (b) they eventually become part of the same cluster (c) and one of them is terminated. In this case it is the dark blue track (people) which is terminated. When the persons separate from the car one of them will generate a new track (not shown).

Figure 1 (d), (e), and (f) illustrate the same situation with group tracking: when the two targets form the same cluster they are joined in a new group (d). The track of the group of targets is colored green. Because of the $width/height$ target feature the two targets separate from one another correctly (e). Finally after the car has left the image, the persons are still being tracked correctly (f).

V. Conclusions

The benefits of enhancing the standard MHT application with group tracking were evidenced. The $width/height$ ratio, can be very effective in discriminating between different targets as the provided example demonstrated.

Acknowledgements

This work has been partially supported by the Portuguese Government - Fundação para a Ciência e Tecnologia (ISR/IST pluriannual funding) through the PIDDAC program funds, and by the project DCCAL, PTDC / EEA-CRO / 105413 / 2008.

References

[1] Samuel S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerospace and Electronic Systems Magazine*, 19:5–18, 2004.
[2] Stephen Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241–254, 1967.
[3] Pedro Miguel Torres Mendes Jorge. *Seguimento de Pessoas e Grupos em Sinais de Video com Redes Bayesianas*. PhD thesis, 2007.
[4] Boris Lau, Kai O. Arras, and Wolfram Burgard. Tracking groups of people with a multi-model hypothesis tracker. In *ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pages 3487–3492, Piscataway, NJ, USA, 2009. IEEE Press.
[5] Manuel Mucientes and Wolfram Burgard. Multiple hypothesis tracking of clusters of people. pages 692 –697, 2006.
[6] Donald B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24:843–854, 1979.