Joint Collaborative Tracking and Multitarget Shape Updating Under Occlusion Situations

Francesco Monti, Carlo S. Regazzoni

Department of Biophysical and Electronic Engineering, University of Genova, Genova, Italy

Current Problem

Tracking multiple video objects is a challenging task due to the interaction that occurs on the image plane. The evidence of the objects is mixed and the same observations are used for more than an object with the result of tracking and model updating failures

Proposed Solution

A collaborative approach where interacting trackers share information about their understanding of the scene to avoid the use of the same observations for more than a tracker.

Overview

The proposed tracking approach is based on shape information estimated by using as observations corner information (a sampling of the shape in points of local structure).

The state of the object is composed by:

-position state $\underline{X}_p$: position of the object on the image plane

-shape state $\underline{X}_s$: coordinates, with respect to the center of the model, and persistence of points of high curvature which sample the shape of the object.

The observations $\underline{Z}$ are the coordinates of the corners extracted from the image.

There is no need to jointly estimate the state of objects that don't interact. Interacting trackers are therefore grouped in collaborative sets. The states of the objects in a set are estimated collaboratively by the trackers in the set independently from the other trackers.

The tracking algorithm consists of two steps:

-collaborative position estimation

-collaborative shape estimation

The two steps procedure is motivated by the following Bayesian decomposition where $\underline{X}^{\aleph}$ is the joint state of the objects in a set

- Divide the trackers in collaborative sets
- For each set
- Compute the position of the objects in the set
- Compute the shape of the objects in the set

$$p\left(\underline{X}^{\aleph}_{p,t},\underline{X}^{\aleph}_{s,t}\left|\underline{Z}^{\aleph}_{t},\underline{X}^{\aleph}_{p,t-1},\underline{X}^{\aleph}_{s,t-1}\right.\right)=\underbrace{p\left(\underline{X}^{\aleph}_{p,t}\left|\underline{Z}^{\aleph}_{t},\underline{X}^{\aleph}_{p,t-1},\underline{X}^{\aleph}_{s,t-1}\right.\right)}_{\text{Collaborative position estimation}}\cdot\underbrace{p\left(\underline{X}^{\aleph}_{s,t}\left|\underline{Z}^{\aleph}_{t},\underline{X}^{\aleph}_{p,t-1},\underline{X}^{\aleph}_{s,t-1},\underline{X}^{\aleph}_{p,t}\right.\right)}_{\text{Collaborative shape estimation}}$$

The states that maximize the posterior are selected as the positions and shapes at time t. The maximization can be solved sub-optimally by estimating the position state at first and, by fixing the position, estimating the shape state.

Collaborative position estimation

The collaborative position estimation is decomposed in :

-A priori position estimation

-Position likelihood estimation

A priori position estimation

By hypothesis, an object position, conditioning on its past position is independent from the positions of the objects in the set → It is possible to factor the joint probability function and use for each object its predictive model

Position Likelihood estimation

The likelihood represents the probability of occurrence of a set of observation corners given the past shapes, positions and an hypothesis on the present position.

PROPOSED SOLUTION

Likelihood estimation → *In the single object case [1] it is computed using a many to many matching algorithm that trough a voting mechanism computes the similarity, given a position hypothesis, of the cloud of observed corners with the model points. In the multiple object case, the observations of the objects are not separable. The likelihood is therefore not factorizable.*

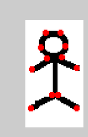
- For each object in the set
- Compute the position likelihood as if it were alone in the scene → *The maximum in the likelihood of the most visible object in the set does not depend on the observations generated by the other objects.*
- Compute the order of visibility (Q) of the object → *Estimated by computing the ratio between the number of matched observations in the position with the highest number of votes and the number of its model points: the model with the ratio nearest to 1 (i.e. the model which explain better the observations) is selected as the "most visible object".*
- Select the most visible object and store its estimated position and observations → *The position with the highest likelihood is selected as the position of the most visible object and its observation are estimated as the observations that contributed to vote to the selected position.*
- Remove the observations of the most visible object from the scene → *Not all the observations are removed. The observation that voted also for near local maxima are left since they can belong to near objects.*

[1] M. Asadi and C. Regazzoni, "Tracking using continuous shape model learning in the presence of occlusion," *EURASIP J. on Advances in Signal Processing, Special Issue on Track Before Detect Algorithms*, [in press].

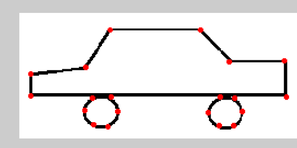
Collaborative likelihood estimation example

The likelihood estimation is explained trough an example in controlled conditions.

Objects models at time t-1

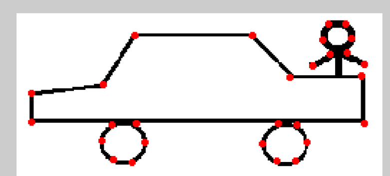


Man (12 model points)



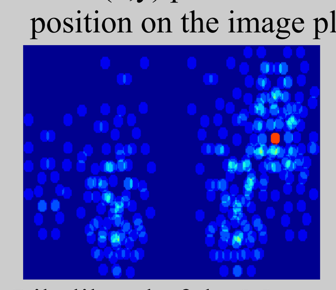
Car (20 model points)

First iteration

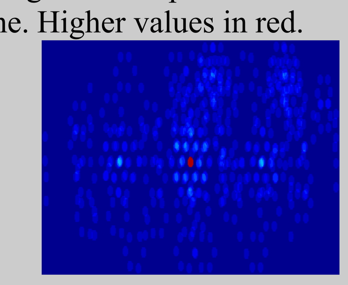


Scene at time t (observations in red) (the man is partially occluded and only 8 corners are visible)

Each (x,y) position on the figure corresponds to a position on the image plane. Higher values in red.



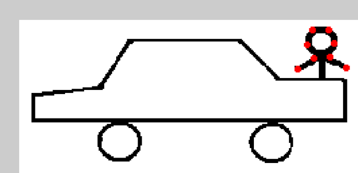
Likelihood of the Man
The maximum obtains 8 votes → Q=0.66



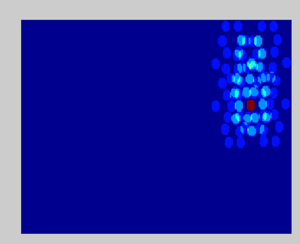
Likelihood of the Car
The maximum obtains 20 votes → Q=1

The car is therefore estimated as in foreground. Take the value of the maximum as the value of the likelihood of the car and remove the observations that contributed to this maximum.

Second iteration

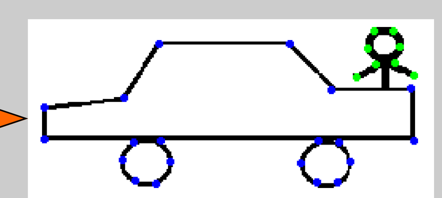


Remaining observations



Likelihood of the man using the remaining observations
Take the value of the maximum as the value of the likelihood of the man

The observations are therefore divided between the trackers. This information is exploited during the shape estimation →



Collaborative shape estimation

The collaborative shape estimation is decomposed in :

- *A priori Shape estimation*
- *Shape likelihood estimation*

A priori Shape estimation

The shape prediction model jointly predicts the shapes of the objects in the set given the shapes at time t-1 and the actual positions with the rationale that model points of different trackers that share the same position on the image plane can't increase their persistency at the same time.

The model assigns a 0 probability to the shapes that are not coherent with the shapes at time t-1 and assigns to the others an equal probability of occurrence.

The following rules are used to check if a shape is coherent with respect of the past shape

1. *A model point can increase or decrease its persistence by one*
2. *A model point that is not in the model at time t-1 can be inserted in the model with persistency 1*
3. *A model point with persistency 1 at time t-1 can increase its persistence by one or can be removed*
4. *Two model points of different trackers that, once projected on the image plane share the same position, cannot increase their persistency at the same time*

Point 4. and the Shape likelihood jointly contribute to avoid the use of the same observation for the estimation of the shape of more than a tracker.

Shape likelihood estimation

The shape likelihood represents the probability of occurrence of a set of observation corners given the past shapes, positions, the actual position and an hypothesis on the present shape.

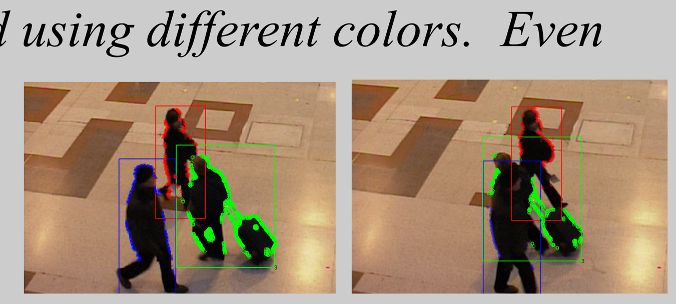
The observations have been assigned to each tracker during the position estimation procedure. Since each corner is considered as extracted independently from the others, it is possible to factor the likelihood as a product on all the points of the image plane

Each term of this product, given an a-priori estimated shape, is 0 (is 1 otherwise) if

1. *a corner is extracted in (x,y), it belongs to the tracker under analysis and in the a-priori shape there is a model point in (x,y) with decreased persistence*
2. *a corner is not extracted in (x,y) and in the a-priori shape of the corner under analysis there is a model point in (x,y) with increased persistence.*

Shape estimation results

The model points with increased persistence of each tracker are highlighted using different colors. Even during complex occlusion the observations are correctly assigned.

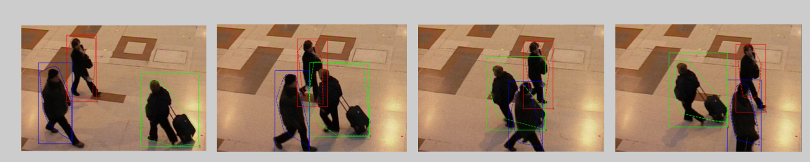


Results

Experiments have been done on different sequences containing occlusion. The bounding boxes of the objects are plotted using a solid line. The convex hull of the model using a dashed line

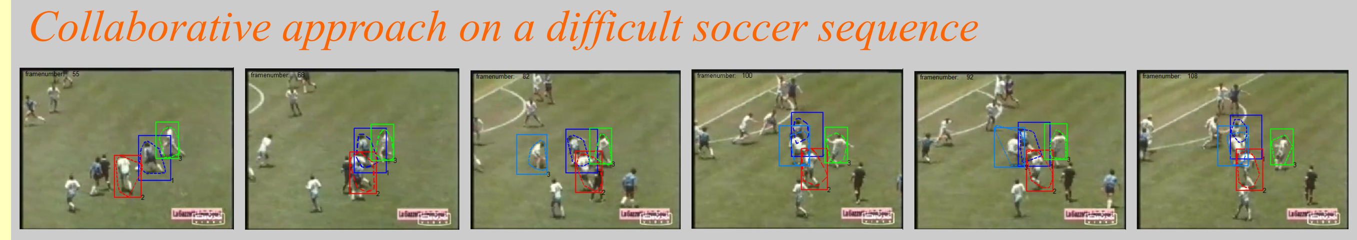
Comparison between Collaborative and non collaborative approach

Collaborative Approach →



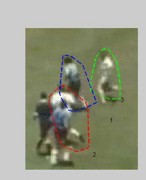
Non collaborative Approach [1] →

Collaborative approach on a difficult soccer sequence

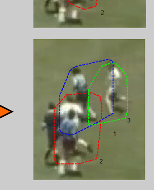


Comparison between collaborative and non collaborative shape estimation results. The convex hull of the model point is estimated more precisely by using the Collaborative approach.

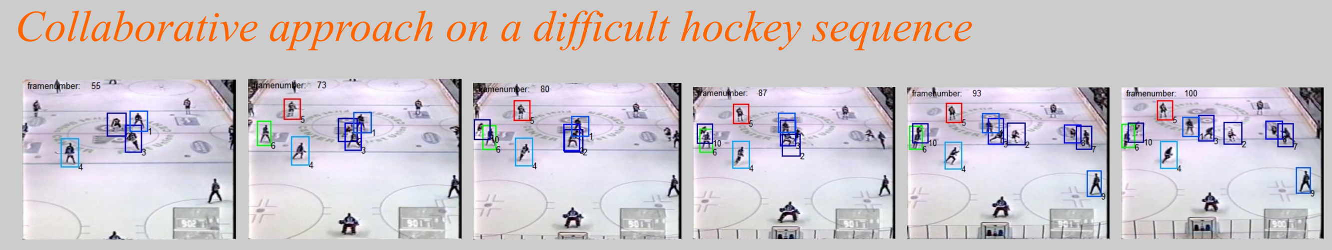
Collaborative Approach →



Non collaborative Approach [1] →



Collaborative approach on a difficult hockey sequence



The proposed collaborative approach (that does not use a priori information about the target) obtained on this sequence results comparable to state of the art methods that use offline learned models