“Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics”

Keni Bernardin and Rainer Stiefelhagen

Presented by Abhiram Iyer
ECE 285, Spring 2020

Objective

- MOT (multiple object tracking) metrics are not universal
  - Every paper or approach presents their approach to evaluating how well a MOT system performs
  - One metric can sometimes be interpreted as a combination of others
- Paper introduces 2 *universal* metrics that can be applied to any problem
  - MOTP (multiple object tracking precision)
    - Precision determined by difference in location or overlap between hypothesis *h* and object *o*
  - MOTA (multiple object tracking accuracy)
    - Overall accuracy of how well the system has performed
- Paper uses the CLEAR workshops (classification of events, activities, and relationships) to demonstrate how well the metrics are used

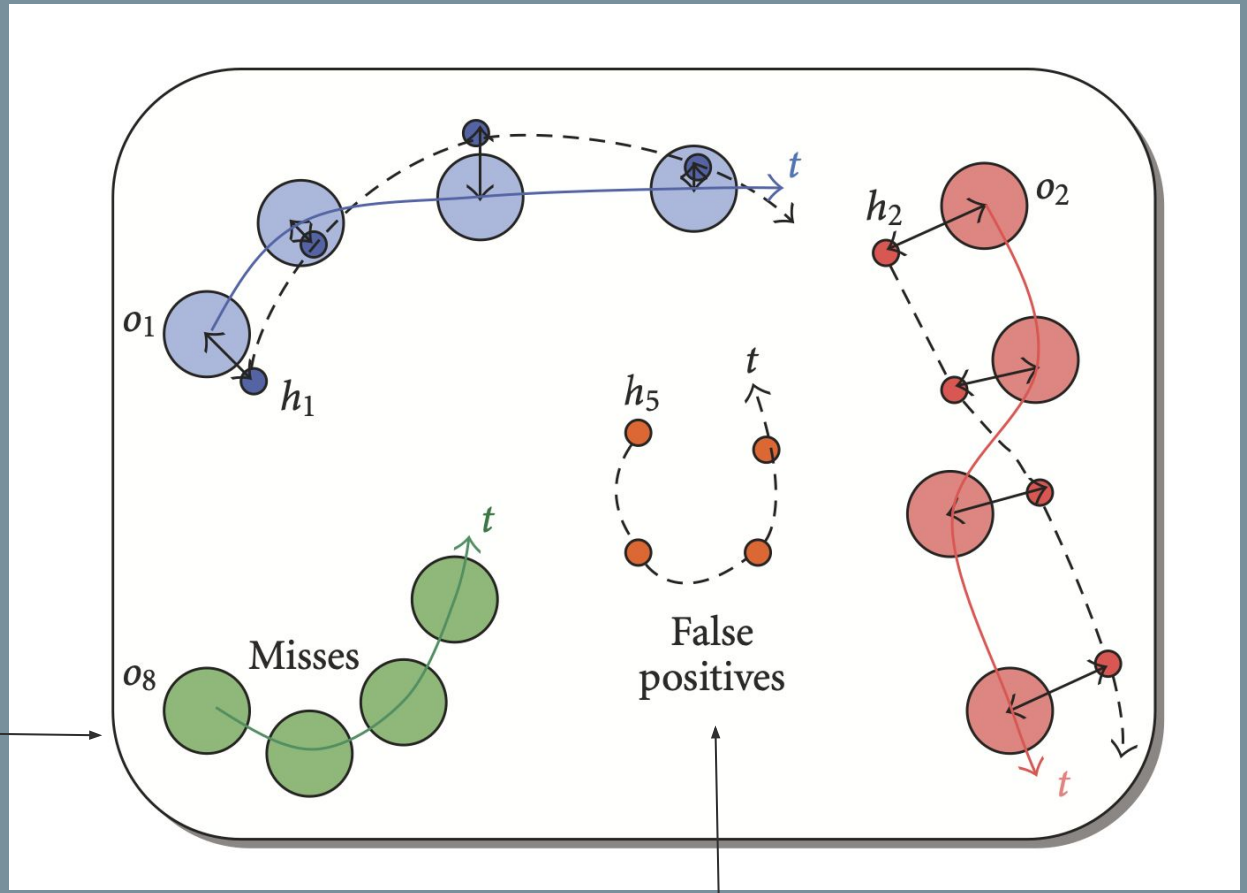
Methodology

- Must use the following criteria when designing the metrics
  - Be able to judge tracker's precision when determining object locations
  - Must consistently track object through time and produce one trajectory per object
  - Have few free parameters - helps in straightforward comparisons between experiments
  - Easily understandable and intuitive
  - Be general so we can apply them to various tasks (2D, 3D tracking, etc.)
  - Few metrics but expressive

Assume a set of hypotheses $\{h_1, ..., h_m\}$ and a set of visible objects $\{o_1, ..., o_n\}$

- For each time frame 't':
  - Establish best correspondence between hypothesis and object.
  - Compute error in object's position estimation ***(TRACKING PRECISION)***
  - Add up all correspondence errors ***(TRACKING ACCURACY)***
    - Miss = object has no hypothesis
    - False positive = hypothesis refers to a non-existent object
    - Mismatch error = hypothesis for an object changed compared to previous frames

Methodology (continued)



No hypotheses

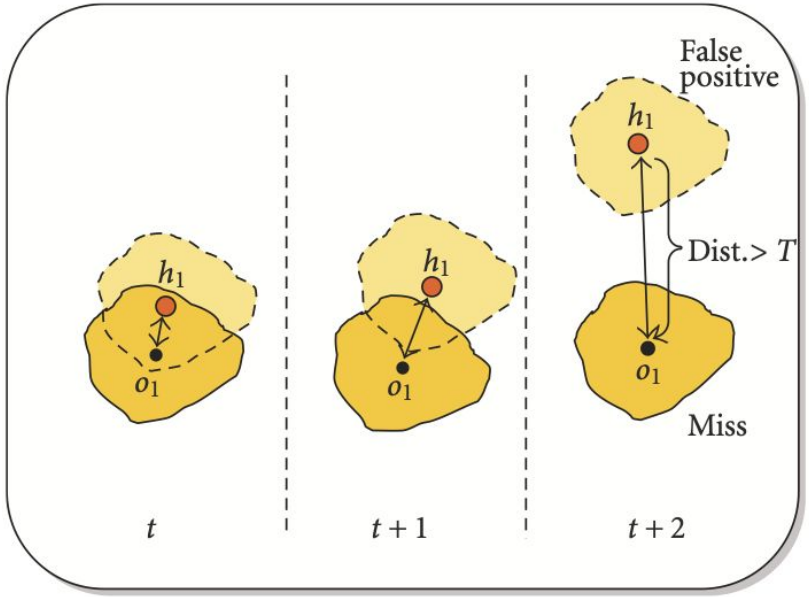
No real objects exist

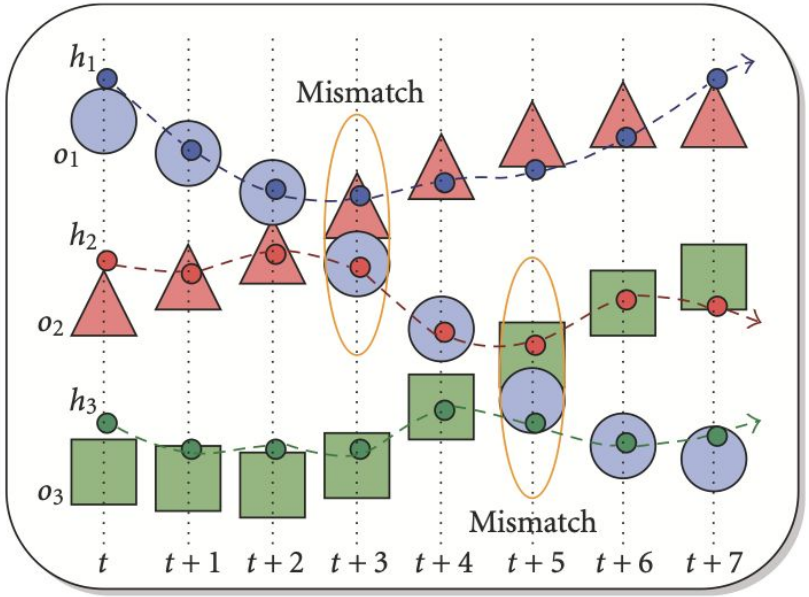
Methodology (continued)

Some more intuition when constructing MOTA and MOTP

- Correspondence between an object and a hypothesis should not be made if their distance exceeds a certain threshold ‘T’
  - At some point, an error in position estimation turns into a miss/false positive: the tracker has missed the real object and/or the hypothesis corresponds to an object that doesn’t exist
  - Can use Euclidean distance (in 2D or 3D image coordinates) or calculate overlap between objects (in which case the threshold is set to zero overlap)
- Assume that M_t corresponds to object and hypothesis mappings at time *t*
  - If a new correspondence is made at time *t+1* between o_i and h_k that contradicts a mapping (o_i, h_j) in M_t , a mismatch error is counted and (o_i, h_j) is replaced by (o_i, h_k) in M_{t+1}
  - Why? We want to use the mappings at the current time step to make a judgment on the mappings in the next time step
    - Will help us find the most likely track if there are multiple hypotheses mapped to the same object (i.e. we use the hypothesis in the last time step)

Methodology (continued)





Methodology (continued)

Algorithm (Part 1)

Initializations:

$M_0 = \{\}$

$mme_t = 0$ → mismatch errors

$fp_t = 0$ → false positive errors

$m_t = 0$ → misses

Methodology (continued)

Algorithm (Part 2)

For every time frame ‘t’:

g_t = total # of objects present at this time ‘t’

For every $\{o_i, h_j\}$ in previous mapping M_{t-1}:

- Make sure $\{o_i, h_j\}$ is still valid. If o_i is still visible and h_j still exists at time ‘t’ and if their distance < threshold (or overlap > threshold), make correspondence between o_i, h_j for time step ‘t’

Methodology (continued)

Algorithm (Part 3)

For every time frame ‘t’:

For objects which no correspondence was made:

- Find a matching hypothesis using the minimum weight assignment problem (problem can be solved with Munkres’ algorithm in polynomial time)

- If correspondence $\{o_i, h_k\}$ is made that contradicts a mapping $\{o_i, h_j\}$ in M_{t-1}, replace $\{o_i, h_j\}$ with $\{o_i, h_k\}$ and increment mme_t

Methodology (continued)

Algorithm (Part 4)

For every time frame ‘t’:

All the matches made in time step ‘t’ can be described as c_t. For each match, calculate the distance d_t^i between the object o_i and its corresponding hypothesis

All remaining hypotheses = false positives. Increment fp_t accordingly

All remaining objects = misses. Increment m_t accordingly

Methodology (continued)

Algorithm (Part 5)

$$\text{MOTP} = \frac{\sum_{i,t} d_t^i}{\sum_t c_t}.$$

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + f\,p_t + mme_t)}{\sum_t g_t}$$

Results and Analysis

- The CLEAR evaluation workshops feature tasks like tracking humans and objects in natural, unconstrained indoor and outdoor scenarios
  - Data collected to test systems that fused multimodal and multisensory data

- The paper describes how the threshold ‘T’ was set for various challenges in CLEAR
  - 3D visual person tracking: Euclidean distance between hypothesized and labeled persons on the ground plane with T = 50cm
  - 2D face tracking: overlap between hypothesized and labeled face (via bounding boxes). T = zero overlap
  - 2D person and vehicle tracking: bounding box overlap again
  - 3D acoustic and multimodal person tracking: systems expected to pinpoint 3D location of active speakers. Euclidean distance on the ground plane with T = 50cm

Results and Analysis (continued)

More intuition

- If T approaches infinity, all correspondences stay valid no matter how large the distance between object and hypothesis becomes
- If T approaches 0, all objects will eventually be considered missed, and the MOT metrics are not useful

Results and Analysis (continued)


ITC-irst


UKA

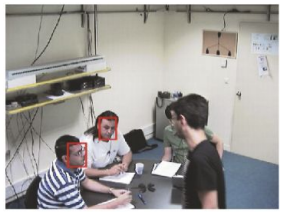

AIT


IBM


UPC

FIGURE 4: Scenes from the CLEAR seminar database used in 3D person tracking.


UKA


AIT

Results and Analysis (continued)

Site/system	MOTP	Miss rate	False pos. rate	Mismatches	MOTA
System A	92 mm	30.86%	6.99%	1139	59.66%
System B	91 mm	32.78%	5.25%	1103	59.56%
System C	141 mm	20.66%	18.58%	518	59.62%
System D	155 mm	15.09%	14.5%	378	69.58%
System E	222 mm	23.74%	20.24%	490	54.94%
System F	168 mm	27.74%	40.19%	720	30.49%
System G	147 mm	13.07%	7.78%	361	78.36%

FIGURE 7: Results for the CLEAR'07 3D multiple person tracking visual subtask.

Site/system	MOTP	Miss rate (dist > T)	Miss rate (no hypo)	MOTA
System A	246 mm	88.75%	2.28%	−79.78%
System B	88 mm	5.73%	2.57%	85.96%
System C	168 mm	15.29%	3.65%	65.44%
System D	132 mm	4.34%	0.09%	91.23%
System E	127 mm	14.32%	0%	71.36%
System F	161 mm	9.64%	0.04%	80.67%
System G	207 mm	12.21%	0.06%	75.52%

FIGURE 8: Results for the CLEAR'06 3D Single Person Tracking visual subtask

System G had the best performance in terms of accuracy, but System B had the best precision

- For single person tracking, "miss rate (hypo)" (i.e. misses resulting from failures to detect person or object) did not play much of a role → intuitively explains that single person tracking is effective in at least finding the object of interest
- System B far outperformed the others in terms of precision

Results and Analysis (continued)

Site/system	MOTP	True miss rate	True false pos. rate	Loc. error rate	A-MOTA
System A	257 mm	35.3%	11.06%	26.09%	1.45%
System B	256 mm	0%	22.01%	41.6%	−5.22%
System C	208 mm	11.2%	7.08%	18.27%	45.18%
System D	223 mm	11.17%	7.11%	29.17%	23.39%
System E	210 mm	0.7%	21.04%	23.94%	30.37%
System F	152 mm	0%	22.04%	14.96%	48.04%
System G	140 mm	8.08%	12.26%	12.52%	54.63%
System H	168 mm	25.35%	8.46%	12.51%	41.17%

FIGURE 9: Results for the CLEAR'07 3D person tracking acoustic subtask.

System G was the best system. It had the best precision and highest accuracy

Site/system	MOTP (overlap)	Miss rate	False pos. rate	Mismatch rate	MOTA
System A	0.66	42.54%	22.1%	2.29%	33.07%
System B	0.68	19.85%	10.31%	1.03%	68.81%

On first glance, System A and B performed relatively equally in terms of precision. However, the accuracy (MOTA) tells us a different story and shows that System B outperformed System A significantly

Advantages and Disadvantages

- Advantages
  - Universal way of evaluating MOT systems, since every implementation might use different evaluation metrics
  - Paper presents a general and flexible framework that can easily be applied to any MOT system
  - Framework contains only 1 “free” parameter (threshold)
    - Changing the threshold won’t completely invalidate comparisons between different experiments - the MOT metrics (MOTA and MOTP) still remain highly effective points of comparison
  - Framework is intuitive - looking at the results tells us exactly *how* one system performed better than another
- Disadvantages
  - Despite having only 1 free parameter, deciding how to adapt this parameter to a given MOT system might be tricky in some cases (i.e. in situations where Euclidean distance and overlap are not good indicators of precision)
  - Forced to do a grid search approach of deciding which threshold works best. The paper mentions that T = 50cm was chosen “intuitively”. A procedure to choose an optimal threshold just given the data is unknown.

Key Takeaway

The paper presents a clear set of MOT metrics (specifically, MOT accuracy and MOT precision) that can universally be applied to any MOT problem. These metrics are not only intuitive to understand, but also enable researchers to effectively compare the results between various problems and experiments.

Question

The paper uses a threshold measurement ‘T’ to calculate MOTP (multiple object tracking precision). When is it better to use an area overlapping threshold (i.e. how much one object overlaps with another) rather than a Euclidean distance threshold (i.e. how far away one object is from another)?

Q&A