PERFORMANCE EVALUATION FOR TRACKING ALGORITHMS USING OBJECT LABELS

Julius Popoola and Aishy Amer

Concordia University, Electrical and Computer Engineering, Montréal, Québec, Canada
E-Mail: j_popool, amer@ece.concordia.ca

ABSTRACT

This paper proposes performance measures to evaluate object tracking algorithms using object labels and sizes.. The usefulness and effectiveness of the proposed evaluation measures are shown by reporting the performance evaluation of two tracking algorithms. We then compare the application of the proposed evaluation measures with related work to demonstrate that they are more suitable.

***Index Terms*—** Performance Evaluation, Object tracking, Object detection

1. INTRODUCTION

The development of object tracking algorithms has become a popular area of research in recent years with many applications developed. An example of such application is the surveillance system, whereby the tracking algorithm enables the surveillance systems to detect deposit and removal of objects, vandalism and monitor movement in the restricted areas. This leads to the improvement in level of security provided in areas where such surveillance systems are deployed.

Since, many tracking algorithms are being proposed, objectively measuring the efficiency of a tracking algorithm has become necessary for comparing the performance of tracking algorithms. The importance of such an evaluation measure for object tracking algorithms has been recognized by the research communities and several papers on evaluation metrics have been presented in the International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), but no standard evaluation measure is available as of present.

The purpose of this paper is to propose a new performance evaluation metric for tracking algorithms. The proposed metric is called Label and Size Based Evaluation Measure (LSBEM), which uses the unique label assigned to objects and size of the object detected by the tracking algorithm as the main factors for evaluating a tracking algorithm. To demonstrate the effectiveness of LSBEM the performance of two tracking algorithms based on [1] and [2] was evaluated using several video sequences. The result of the evaluation is compared with the result obtained using the evaluation measure proposed in [3], using the same video sequence. The result of the comparison proves that LSBEM is an effective objective measure for tracking algorithm.

The contributions of our approach are as follows; firstly, four normalized measures sufficient for analyzing object tracking performance; secondly, the inclusion of unique object labels, which allows LSBEM to evaluate both object detection and object tracking performance; and finally, object size analysis to evaluate the tracking performance in relation to inaccurate object size detection

This work was supported, in part, by the Natural Science and Engineering Research Council (NSERC) of Canada.

The remainder of this paper is as follows. Section 2 discusses related work. Section 3 explains the criteria for evaluating tracking algorithm. LSBEM is introduced in Section 4, while the result and conclusion are presented and discussed in Section 5 and 6, respectively.

2. RELATED WORK REVIEW

In [3], Black et. al presented an evaluation metric that finds the correspondence between ground truths and tracked objects to compute true positive and false positive matches to calculate the following measures. Tracker Detection Rate and False Alarm Rate measures the object detection rate of tracking algorithms, the Track Detection Rate measure the rate at which individual objects are detected in relation to the ground truth. Object Tracking Error measures the accuracy of the correspondence. Track Fragmentation is the number of times a track label changes and the Tracking Success Rate is the ratio of the non-fragmented tracked objects to the total number of the ground truth objects. Occlusion Success Rate (OSR) gives the rate at which the tracking algorithm detects occlusion. The metrics presented in [3] gives a good overview of some of the fundamental features of a tracking algorithm that should be evaluated. However, inaccurate object size detection and how accurately the objects are labeled is not evaluated.

The evaluation technique presented in [4] determines the threshold from the distance matrix between the centroid of the bounding box for the ground truths and the result of the tracking algorithm. This threshold is then used to find the correspondence between the result of the tracking algorithm and the ground truth to compute False Positive Track Error, False Negative Track Error, Average Area Error, and Task Incompleteness Factor measures. The metrics presented in [4] does not measure the performance of the tracking algorithm in regards to objects labeling and occlusion detection.

The performance evaluation metrics presented in [5] is divided into frame-based and object-based metrics. For the frame-based metrics true positive, true negative, false positive, and false negative matches are computed for all frames, which is used to calculate the Tracker Detection Rate, False Alarm Rate, Detection Rate, Specificity, Accuracy, Positive Prediction, Negative Prediction, False Negative Rate and False Positive Rate. For the object-based metrics, each individual object correspondence is used to compute the true positive, false positive and total ground truth to calculate the Tracker Detection Rate, False Alarm Rate, and Object Tracking Error (OTE). The frame-based metrics provides information about how the tracking algorithm handle objects within a frame, and the object-based metric allow for measuring the performance of the tracking algorithm on how each object is tracked during the duration of the video sequence. The process of labeling, changes in the size of the object and occlusion detection was not evaluated.

3. CRITERIA TO EVALUATE TRACKING ALGORITHM

In this section we discuss the criteria to be used to evaluate tracking algorithms. The evaluation of a tracking system is based on measuring the performance of the tracking algorithm in relation to the features of the system. The fundamentals of a tracking system are the detection of objects and the extraction of the object trajectories within a video scene. For a tracking algorithm to monitor and extract the trajectory of multiple objects, the tracking algorithm should be able to differentiate between the objects. This normally accomplished by labeling objects once they are first initialized. Hence, the features of a good tracking algorithm are as follows;

1. The tracking algorithm should detect all the objects that enter or moved in the scene.
2. The tracking algorithm should differentiate between multiple objects that are present in the scene at the same time.
3. To monitor and extract the trajectory of all objects the unique label assigned to each object must be maintained for all the tracked objects.
4. The motion or lack of motion of the object should not lead to change of object label.
5. The tracking algorithm should handle occlusion and exposure without object labels changing.

After reviewing the performance evaluation metrics proposed in [3], [4], and [5], we found that these performance evaluation metrics focus on object detection with little reference to how objects are labeled. From subjectively evaluating the performance of tracking algorithms it was also discovered that multiple labels of the same object leads to significant errors in post-tracking-process. These discoveries lead to the proposal of the LSBEM.

4. PROPOSED LABEL AND SIZE BASED EVALUATION MEASURE (LSBEM)

The metrics used for calculating LSBEM is computed by comparing the unique object label and size of each tracked object to the unique label and size of the ground truth object. Thus the number of object active frames (OAF), true positive match (TPM), false positive match (FPM), and the size detection rate (SDR) are computed.

The object active frames (OAF) is the number of consecutive frames that the object is visible in the video sequence, true positive match (TPM) is the number of frames where there is a match between the tracked object and the ground truth object that share the same unique labels. False positive match (FPM) is the number of frames that the object with a unique label is not matched to a ground truth and the size detection rate (SDR) is the number of frames the size of the tracked object with a unique label is not less than 80% and not more than 20% of the matched ground truth object. Also, computed are the number of total unique object labels generated (L) , the number of matched labels (L_m), the number of non-matched labels (L_n), the total number of ground truth labels (GTL), the total number of matched ground truth objects (GTM), and the total number of ground truths objects (GTT). The metrics for LSBEM consists of Object Detection Rate (OBR), Average Size Detection Rate ($ASDR$), Label Tracking Detection Rate ($LTDR$), and Non-Label Tracking Detection Rate ($NTDR$).

4.1. Object Detection Rate (OBR)

Object Detection Rate measures the rate at which tracked objects are matched to the ground truth without reference to the labels assigned. OBR value varies between 0 and 1, 0 means poor object detection and 1 means that all ground truth objects are matched.

$$OBR = \frac{1}{GTL} \sum_{g=0}^{GTL} \frac{GTM_g}{GTT_g} \tag{1}$$

4.2. Average Size Detection Rate (ASDR)

Average Size Detection Rate measures the rate at which the tracked object's size differs from the matched ground truth. ASDR value varies between 0 and 1, where 0 means poor object size detection and 1 means accurate object size detection. The sensitivity of the size detection rate can be varied if required, for the result presented sensitivity of ASDR is set at ± 20%.

$$ASDR = \frac{1}{L} \sum_{i=0}^{L} \frac{SDR_i}{TPM_i} \tag{2}$$

4.3. Label Tracking Detection Rate (LTDR)

Label Tracking Detection Rate measures the rate at which uniquely labeled objects matches the uniquely labeled ground truth. LTRD value varies between 0 and 1, 0 implies that the tracking algorithm has poor object matching to the unique labels and 1 means accurate object tracking with the same unique label and location between the tracked object and ground truth.

$$LTDR = \frac{1}{L_m} \sum_{i=0}^{L_m} \frac{TPM_i}{OAF_i} \tag{3}$$

4.4. Non-Label Tracking Detection Rate (NTDR)

Non-Label Tracking Detection Rate is used to account for the tracked objects that are not matched. NTDR value varies between 0 and 1. If NTDR is equal to 1 this indicates that all the non-matched objects are tracked without affecting the matched objects. This situation occurs when objects are detected temporarily especially from reflective surfaces.

$$NTDR = \frac{1}{L_n} \sum_{i=0}^{L_n} \frac{FPM_i}{OAF_i} \tag{4}$$

5. RESULTS

To illustrate the efficiency of the evaluation metrics proposed the tracking performance of the tracking algorithms presented in [1] and [2] was measured using several video sequences, in different environments, and depicting various occurrence of events. The results of six video sequences are presented in this report. The descriptions of the video sequences used are as follows:

1. **Video 1:** A man moves through the hall deposits an object and the moves out of the scene.
2. **Video 2:** A man comes into the room moves around the room and then exits the room. While in the room occlusion occurred between the man and a table that is part of the background.

Table 1. Performance Evaluation using LSBEM. (Higher OBR, ASDR, LTDR and NTDR means better tracking results.)

Video	OBR	ASDR	LTDR	NTDR	Code
1	1.000	0.993	0.983	0.033	[1]
1	1.000	1.000	0.968	0.063	[2]
2	0.995	0.847	1.000	1.000	[1]
2	0.995	0.951	1.000	1.000	[2]
3	0.945	0.899	0.778	0.611	[1]
3	0.941	0.983	0.907	0.607	[2]
4	0.995	0.967	0.658	0.921	[1]
4	0.976	0.875	0.407	0.960	[2]
5	0.960	0.997	0.973	0.757	[1]
5	0.948	0.977	0.867	0.711	[1]*
5	0.962	0.996	0.926	0.289	[2]
6	0.903	0.949	0.845	0.851	[1]
6	0.744	0.915	0.796	0.936	[2]

3. **Video 3:** The scene of an hallway where two men both entered in the scene and through the hallway and one man deposition an object.
4. **Video 4:** One woman walked across the scene of the camera, she was lost for a short period because she walked behind a pillar in the middle of the scene.
5. **Video 5:** Two people walking in opposite directions completely occlude each other when they meet.
6. **Video 6:** Three people moving in different direction experience occlusion and exposure.

The result of evaluating the tracking algorithms presented in [1] and [2] using the proposed metrics is shown in Table 1.

5.1. Comparing tracking methods

For Video 1, 100% of the ground truth objects where matched for both [1] and [2]. But the value of LTDR shows that [1] correctly matched 98% of the uniquely labeled objects to the uniquely labeled ground truth objects and [2] correctly matched 97% of the uniquely labeled objects to the uniquely labeled ground truth. Since 100% of the ground truth objects were matched, the value of LTDR is close to 100% and the value of NTDR is very low, this indicates that a few objects were detected in some frames that are not part of the ground truth. From subjective evaluation, it was found that these additional objects where as a result of the split detection that occurs when the object was deposited. With [1]'s ASDR value at 99% and the ASDR value of [2] is 100%, [2] is selected as the best algorithm for video 1.

For video 2, the value of OBR, LTDR and NTDR are the same for both [1] and [2]. The value of OBR shows that 99% of the ground truth objects were matched to the tracked objects. The unit value of LTDR indicates that 100% of the uniquely labeled objects match the uniquely labeled ground truth, and the value of NTDR shows that 100% of the uniquely labeled objects were not matched to ground truth objects that were tracked independently. The unity value of NTDR is as a result of objects that temporarily appear on the scene. Examples of such objects are due to the appearance of shadows and reflections, as shown in (a) of Fig. 1. The better algorithm for video 2 is [2] because 95% of the objects matched to the ground truth were within the 20% size range.

(a) (b)

Fig. 1. (a) Tracking output of [2] from video 1 and (b) Tracking output of [2] from video 6.

For Video 4, the effect of the object lost for a short period of time is very evident in value of LTDR, because the woman walking across the scene was correctly tracked with the correct label until she was lost for a short period of time, after that she was relabeled.

For video 5, the algorithm presented in [1] performed better than [2] comparing the first and third line of video 5 results in Table 1. When the occlusion correcting step in [1] was disabled the LTDR value for [1] reflects the change in performance, as depicted in Fig. 2 and second line for video 5 in Table 1 marked with asterisks.

(a) Before Occlusion (b) After Occlusion

(c) Before Occlusion (d) After Occlusion

Fig. 2. Tracking output of[1] for Video 5 (a) and (b) and tracking output of [2] for Video 5 (c) and (d).

The performance of the tracking algorithm in relation to recovery from occlusion is the cause for change in the evaluation result. When exposure occurred as show in Fig. 2 [2] was able to retain the object label while [1] assigns one of the objects a new label as shown in Fig 2. The ability of LSBEM to detect and account for such occurrence is demonstrated in the value of LTDR for video 5. The low value of NTDR for video 5 is also a demonstration of the effectiveness of the LSBEM, because of the algorithm presented in [2] merges all object labels when occlusion occurs making it difficult to compare with ground truths during the partial occlusion stage.

For video 6, the algorithm presented in [1] proved to be better than [2] with higher values for OBR, ASDR, and LTDR. This is because video 6 consists of frames where three people occluded as shown in (b) of Fig. 1 and the occlusion lasted for 60 frames, which [1] is better designed to handle. As shown in Table 1 the value of NTDR for the algorithm presented in [1] is less than [2] for most video sequences. This is because [2] has a feature for correcting splits introduce by segmentation errors, thereby reducing the number of non-matched regions to the ground truth.

Table 2. Performance Evaluation using Technique presented in [3]. (Higher TRDR, ATDR, OSR, and TSR mean better tracking and Lower FAR, AOTE, and TF means better tracking.)

Video	TRDR	FAR	ATDR	AOTE	TSR	TF	OSR	Code
1	1.000	0.010	0.983	1.407	1.00	0	0	[1]
1	1.000	0.020	0.968	1.387	1.00	0	0	[2]
2	0.995	0.240	1.00	3.670	1.00	0	0	[1]
2	0.995	0.512	1.00	3.780	1.00	0	0	[2]
3	0.911	0.099	0.958	1.692	1.00	0	1	[1]
3	0.907	0.136	0.954	6.432	1.00	0	1	[2]
4	0.884	0.904	0.745	12.077	1.00	0	0	[1]
4	0.768	0.897	0.606	11.878	1.00	0	0	[2]
5	0.960	0.052	0.962	2.023	1.00	0	1	[1]
5	0.948	0.133	0.951	1.989	1.00	0	1	[1]*
5	0.963	0.044	0.964	2.531	1.00	0	1	[2]
6	0.931	0.455	0.902	7.556	0.286	3	1	[1]
6	0.907	0.622	0.744	14.004	0.286	3	1	[2]

5.2. Comparing evaluation measures

To illustrate the efficiency of LSBEM, it is compared to the performance evaluation metrics presented in [3]. The selection of [3] over [4] and [5] is due to the fact that its metrics are more comprehensive to the criteria for tracking algorithms. The result of evaluating [1] and [2] for the same six videos using [3] is presented in Table 2.

Comparison between the result of the performance evaluation of [3] and LSBEM shows a close similarity for video 1 and video 2. The values of TRDR and ATDR of the metrics proposed by Black et. al. is almost identical to OBR and LTDR of LSBEM, respectively for video 1 and 2. This expected behavior is because errors due to labeling did not occur in video 1 and video 2. However, in video 5 where the label of one of the object changed, the resulting values of the performance evaluation proposed in [3] could not be used to convincingly outline this error. However, the effect of the error due to the change of labels for the same object is reflected in the metrics used by LSBEM, which is shown by the value of LTDR and NTDR. This is indicated on the second line for Video 5 in Table 1 and 2 marked with asterisks. The proposed evaluation metric LSBEM is more suitable for evaluating object tracking than related work (e.g., [3]) because its values directly reflects the behavior of the tracking systems in regards to how the tracking algorithm tracks each individual objects with reference to how the objects are labeled.

The LSBEM's LTDR and NTDR provides information about how the tracking algorithm accounts for labeled objects in regards to the ground truth. For example, the ratio of LTDR to NTDR of Video 1 (Table 1) indicates that during the life span of the objects being tracked, the ground truth and the tracked object did not correspond in a few frames. Whereas the ratio of LTDR to NTDR of 2 (Table 1) shows that the ground truth object was correctly tracked through the objects lifespan, and they were other objects, although not present in the ground truth but correctly tracked. This characteristic of LSBEM to evaluate both object detection and object tracking is not shared with the popularly used Tracker Detection Rate and False Alarm Rate measures in [5] and [3], which only measures the correspondence in the ground truth objects and the objects detected by the tracking algorithm with no reference on how the objects are tracked in other frames. Thus, LSBEM provides information that allows for detection of errors that affect post-tracking processes. Also, LSBEM uses fewer measures than [3] and [5] to characterize the performance of tracking algorithms, all of which are normalized.

6. CONCLUSION

Evaluating the performance of tracking algorithms is very important for establishing the confidence required for the commercialization. To efficiently evaluate the performance of a tracking algorithm the features and behavior of the tracking system must be analyzed. From our analysis we found that there is great need for the inclusion of object labels in the process of performance evaluation of tracking algorithms, which lead to the proposal of LSBEM. From the results presented the need to include information about object label, size, and location for performance evaluation of tracking algorithms is illustrated. Thus, we prove that LSBEM is able evaluate the performance of tracking algorithm better than previously proposed evaluation metrics.

7. REFERENCES

[1] A. Amer, "Voting-based simultaneous tracking of multiple video objects," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 1448–1462, 2005.

[2] J. Li, T. Yang, Q. Pan, and S. Z. Li, "Real-time multiple objects tracking with occlusion handling in dynamic scenes," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 20–25, 2005.

[3] J. Black, T. Ellis, and P. Rosin, "A novel method for video tracking performance evaluation," *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 125–132, 2003.

[4] S. Muller-Schneiders, T. Jager, H. S. Loos, and W. Niem, "Performance evaluation of a real time video surveillance systems," *IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 137–143, 2005.

[5] F. Bashir and F. Porikli, "Performance evaluation of object detection and tracking systems," *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 7–14, 2006.