International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

AN COMPARATIVE MEASUREMENT EVALUATION OF MULTI OBJECT VIDEO TRACKING

[1]A.R.REVATHI,[2]R.GIRIDHARAN, [3]V.GOKUL KUMAR,[4]B.NAVEEN KUMAR,[5]M.HARISHANKAR

[1]Asst.Prof, [2-5]UG Students ,Department of Information Technology, Valliammai Engineering College-Anna University.

[2-5] UG Student Department of Information Technology, Valliammai Engineering College

Abstract—Computer vision is an active research topic in the area of video processing and surveillance. The processing of the framework can be divided into a low level, middle level and high level processing. On the middle level, the processing is mainly focused with object tracking in frame sequence. In this paper we proposedan comparative evaluation of multi object video tracking results are based on few parameters, like accuracy, cardinality error and ID changes etc. The performance evaluation and measurement based on TPR, FPR, target size variation combine accuracy, cardinality error and evaluate ID changes respectively and also provide comparative measurements. Moreover we discussed the importance of tracking and best part challenges in tracking which gives best object tracking in multi objects.

Key terms- video tracking, computer vision, recall, METE, MCMCDA, etc.

I. INTRODUCTION

Video tracking is research topic with applications in event detection, object tracking, behavior analysis and understanding. In general video surveillance involves various stages like background modeling,foreground detection, tracking and recognition of object movement.Video tracking is the process of following the image in the successive frame to determine its motion of the object[1]. Based on the classification of object tracking, video tracking can be categorized into single target video tracking and multi target video tracking. Single target tracking is the process of tracking the single image in the video and it can be implemented by various methods such as kalman filter, particle filter, adaptive particle filter etc. In this paper we discussed about the Multi target video tracking which refers to sequential estimation of number of images in successive frames to determine the motion of the object using various object representation[1][2]. Multi target video tracking can be implemented by point based assignment and region based assignment [3][6]. The performance of multi target tracking algorithm is to measure the distance between set of ground truth and estimated tracks. Tracking error can be quantified by discrepancy between estimated and ground truth target region[3][4].In order to establish the association between estimated and ground truth track point based and region based assignment are Used[1][3][5][6]. According to the multi target tracking three evaluation parameters are taken into account [1][3]. In that Accuracy is the closeness between estimated and ground truth track. Cardinality error is the difference between number of estimated and ground truth target. ID changes is the incorrect association between estimated and ground truth targets. Accuracy can be calculated based on various accuracy measures true positive, true negative, False positive and false negative[7]. The various evaluation measures are distance based and overlap based measures[1][3][9][10]. Distance based measures are not suitable to evaluate due to changes in target size. Overlap based measures is suitable to detect the variation of sizes in the object. Measures of Effective Video Tracking [1] is totrack the multiple targets by the estimation of accuracy, cardinality error and ID changes. A Metric for Performance Evaluation of Multi-TargetTracking Algorithm [2] is to measure the distance between two sets of tracks: the ground truth tracks and the set of estimated tracks. The concept of performance evaluation [3] which helps to evaluate the performance of video tracking. Performance evaluation of object tracking algorithms [5] which uses or compares six video sequences to test the two trackers to evaluate the motion of objects. Framework for performance evaluation of face, text, vehicle detection and tracking in video: Data, metrics, and protocol [6]track only the region which is surrounded by bounding box (region based assignment).

Tracking is done using point based assignment [1] which tracks the target position of an object.This remaining of the paper is organized as following Section II gives out related works, Section III concentrates on the proposed work, Section IV provides out the measures of Trackers used, Section V delivers the Experimental setup and analysis which deals with various Trackers used, and comparison of various trackers respectively, Section VI. Concludes the work.

CHALLENGES IN OBJECT TRACKING

There are many issues related with object tracking such as background clutter, foreground clutter, partial occlusion, full occlusions, appearance changes, and tracking drift. Other possible challenges of tracking are when the foreground object is similar to that of the background, the tracking system fails and begins tracking the distracter, and objects which exhibit very slow motion could be considered as background. Occlusion is the condition where the object of interest gets hidden by other surrounding or neighbouring objects. Overhead cameras or input video from multiple cameras could be a solution to the occlusion. Non-rigid object motion or foreground clutter is the spot where the object exhibits sudden changes in its speed of motion. This can also be referred as motion blur. In case of outdoor environment, swaying trees contributes to background clutter. Illumination or appearance change is also experienced in case of outdoor environment. Solution to illumination variation is that learning based tracking where object appearance is determined over time and updated. Instead of taking the RGB value of frames, YCbCr value will drastically overcome the effects of illumination changes.

II.RELATED WORK

This related works is going to provide the over -view of object tracking systems and various types of tracking and its evaluation methods proposed by various authors based on set of ground truth and estimated tracks.

Tahir Nawaz [10] proposed three parameter-independent measures for evaluating multi target video tracking by considering target-size variations, combine accuracy and cardinality errors, quantify long-term tracking accuracy at different accuracy levels and evaluate ID changes. Discussing their limitations point based tracking. Fei Yin [6] Performance Evaluation of Object Tracking Algorithms deals with the non-trivial problem of performance evaluation of motion tracking. He proposes a rich set of metrics to assess different aspects of performance of motion tracking and framework allows the identification of specific weaknesses of motion trackers such modules or failures under specific conditions. John Garofolo [3] presented a framework for evaluating object detection and tracking in video(face, text, and vehicle). He proposes to address the challenges of object detection and tracking through a comparison of techniques for evaluation of framework. James Black[5] presents a methodology for evaluating the performance of video surveillance tracking systems. In this he introduces a novel framework for performance evaluation using pseudo-synthetic video, which captures data and store. Tracks are taken to characterize the quality of data. BrankoRistic [1] describes performance evaluation to measure the distance between two sets of tracks: the ground truth tracks and the set of estimated tracks based on consistent metric for performance evaluation of multi-target filters, referred to as the OSPA metric. C J Needham [8] evaluates how well a tracker is able to determine the position of a target object. Few metrics exist for positional tracker evaluation; here the fundamental issues of trajectory comparison are addressed, and metrics are presented which allow the key features to be described The metrics developed are applied to real trajectories for positional tracker evaluation. Xue Mei [9] sparse representation has been applied to visual tracking to find the target with the minimum reconstruction error from the target template subspace we propose an efficient L1 tracker with minimum error bound and occlusion detection which we call Bounded Particle Resampling (BPR)-L1 tracker. First, the minimum error bound is quickly calculated from a linear least squares equation, and serves as a guide for particle resampling in a particle filter framework. Without loss of precision during resampling, most insignificant samples are removed before solving the computationally expensive `1 minimization function. The BPR technique enables us to speed up the L1 tracker without sacrificing accuracy. Second, we perform occlusion detection by investigating the trivial coefficients in the `1 minimization. These coefficients, by design, contain rich information about image corruptions including occlusion. Detected occlusions enhance the template updates to effectively reduce the drifting problem. TahirNawaz[4] present a singles core evaluation measure and a protocol to objectively compare trackers. The proposed measure evaluates tracking accuracy and failure, and combines them for both summative and formative performance assessment. The proposed protocol is composed of a set of trials that evaluate the robustness of trackers on a range of test scenarios representing several real-world conditions. Priti P. Kuralkar[11] presents a novel algorithm for detecting moving objects from a static background scene that contains shadows using color images. Object tracking based on motion estimation and detection, background subtraction and shadow removal.In the approach, morphological operations are used for identifying and removed the shadow. Vishwadeep proposes [12] method to detect object based on background subtraction method. A optimization threshold method is used to obtain behaviour of moving object and tracking. The centroid of object is The experimental results show that the proposed method runs quickly, accurately and fits for the real-time detection. Shao- Yi chien [13] proposed a robust threshold decision algorithm for video object segmentation with a multi background model. Dong Wang Huchuan [14] provides the Online object tracking algorithm with sparse prototypes, which exploits both classic principal component analysis (PCA) algorithms with recent sparse representation schemes for learning effective appearance models. Olga Zoidi [15] proposed an appearance-based representation of the target object, based on local steering kernel descriptors and colour histogram information. Siew Wen Chin [16] presented a region-based active contour model (ACM) with local information using watershed segmentation is proposed for lip contour detection. Huchuan Lu [17] proposed a Tracking algorithm that combines complementary tracking modules with a new object representation model to balance between stability and Adaptively. Wenxi lui [25] proposed the multi target tracking algorithm by combining particle filter and RVO which can estimate the desired velocity of the pedestrian in online manner.

III. STRUCTURE OF THE PROPOSED METHOD

In order to approach the proposed work, we need to collect ***more than one data set*** for comparative analysis .Before going to major steps of the proposed work, Initially from the video, numbers of datasets are taken and converted into frames or images. In this frame conversion is done from the datasets.From the obtained frame the object that is needed to be tracked is obtained.While object detection is done background modeling of dark pixels will be taken place. The shadow of the background image will coincides with the foreground image needed to be tracked. Which may increases the complexity in tracking hence to remove the problems background modeling is done.After then the object that is going to be tracked is obtained, then it scaling map construction is done to indicate the image using bounded box or various other representation. Once the image needed to be tracked is obtained then feature extraction is done with the help of Gabor filter.Gabor filters designed to respond well in a variety of skin and scaling texture conditions.

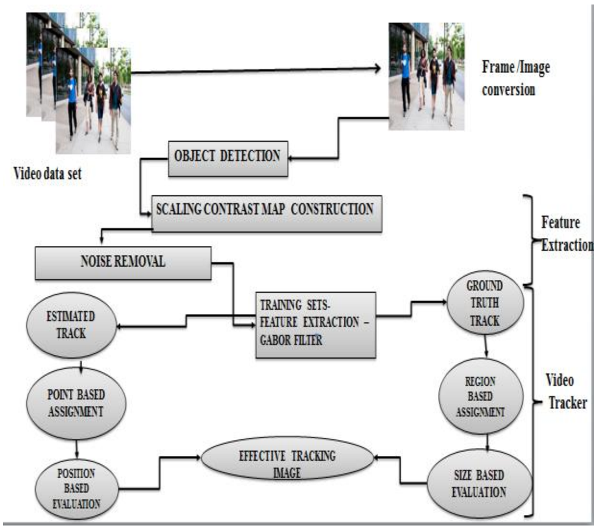

Fig 1. Overview of proposed work

After the training sets are collected estimated and the ground truth tracks are compared to find out the efficiency of the tracking.In estimated track position based evaluation is taken place whereas in ground truth size based evaluation will be taken place.In estimated track point based assignment is done while in ground truth region based assignment is done and finally effectively tracked image is obtained.

IV. a) MULTIPLE EXTENDED-TARGETS TRACKING ERROR ALGORITHM (METE)

Multiple-target tracking algorithm that is autonomous and robust against transmission failures, communication delays and sensor localization error. These important issues are ignored by many tracking algorithm designed for sensor networks. The cardinality error is the difference between the number of estimated and ground truth targets. *ID* changes are the incorrect associations between estimated and ground-truth targets. Then the number of false positive (i.e. its spatial overlap with the ground truth is insufficient) and false negative estimations (i.e. missed estimation having spatial overlap with the ground truth to be zero) is determined by comparing their spatial overlaps with a pre-defined threshold. Moreover, some existing measures are numerically unbounded and not well defined for the worst tracking case.

Multi extended target tracking error algorithm (METE) is formulized as follows:

$$A_K = \min_{\Pi\in\Pi_{\max(v_k - u_k)}} \sum_{i=1}^{\min(v_k - u_k)} (1-0(\overline{A}_{,k}, i, A_k, \Pi(i)) \quad (1)$$

where $O((\overline{A}_k, i, A_{k,}\pi(i)) = | \overline{A}_{K,}i \cap A_{K,}\pi(i) || \overline{A}_{K,}i \cup A_{K,}\pi(i) |$ computes the amount of Spatial overlap between $\overline{A}_{K,}i$ and $A_{K,}\pi(i)$; and $O(\cdot)$.

The Best tracking case is when $A_k = 0$: $O(\cdot) = 1$ for all the associated pairs, and $C_k= 0$ since $u_k = v_k$. This implies $METE_k = 0$.whereas Worst tracking caseis whenAk has its maximum value, i.e. $A_k = u_k = v_k$ when $u_k = v_k$, $A_k = v_k$ when $u_k > v_k$ (the association is performed only for the vk terms) and $A_k = u_k$ when $u_k < v_k$ (the association is performed only for the uk terms). Thus the numerator becomes $A_k+C_k = v_k = u_k : u_k = v_k$ meaning $C_k = 0$; $A_k+C_k = v_k+|u_k-v_k| = u_k : u_k > v_k$; $A_k + C_k = u_k+|u_k-v_k| = v_k : u_k < v_k$. Therefore, $A_k + C_k = \max(v_k, u_k)$, which implies $METE_k = 1$. As the same METE values for two trackers may be caused by different accuracy and cardinality error combinations, it may be useful to analyze these errors separately in order to determine their individual influence in the estimation of METE. To this end, use two error rates such as Accuracy error rate (AER) and Cardinality error rate (CER).

Accuracy Error Rate (AER):

$$\text{AER} = 1/K \textstyle\sum_{k=1}^{k} A \quad (2)$$

Cardinality Error Rate (CER):

$$\text{CER} = 1/K \textstyle\sum_{k=1}^{k} c_k \quad (3)$$

b) MULTIPLE EXTENDED-TARGET LOST-TRACK RATIO(MELT)

Multiple Extended-target Lost-Track ratio (MELT) evaluates tracking accuracy across the sequence in a parameter-independent manner and enables analysis at different levels of accuracy. Given *X'* and *X*, the association is first performed at each frame based on the minimization of the cost $(1-O(\cdot))$ computed for all pairs of estimated and ground truth targets.

Accuracy at track level is evaluated by computing the lost- track ratio (λ_i^τ) for each associated pair of ground-truth track i and estimated track(s)

$$\lambda_i^\tau = \frac{N_i^\tau}{N_i} \quad (4)$$

where N_i^τ is the number of frames with spatial overlap $O(\cdot) \leq \tau : \tau \in R(0,1]$ between the associated pair and Ni is the total number of frames in the ground-truth track i. $\lambda_i^\tau \in [0,1]$; the lower λ_i^τ , the better the performance. We compute the lost-track ratio for a range of a finite number of τ values and obtain $\lambda_i^\tau = \{\lambda_i^\tau\}\tau\in R(0,1]$ such that the total number of sampled τ values is S_τ (required for numerical approximation).To compute λ_i^τ for all V ground-truth tracks to generate the matrix

$$A = \left|\lambda_i^\tau\right| v \times s_\tau \quad (5)$$

where V and S_τ are the number of rows and columns of the matrix.

Tracking performance is quantified by defining the Multiple Extended-target Lost-Track ratio($MELT_\tau$):

$$MELT_\tau = \tfrac{1}{V}\sum_{I=1}^{V} \lambda_i^\tau \quad (6)$$

The computation of $MELT_\tau$ may be useful from an application view point, the performance comparison among trackers can be facilitated by providing the single-score average tracking performance which is generated as

$$MELT = \tfrac{1}{S} \sum_{r\in R_{(0,1)}} MELT_\tau \quad (7)$$

c) NORMALIZED ID CHANGES

Normalized ID Changes (NIDC) measure evaluates the ID changes taking into account the track duration in which they occur. In the case of a comparison of trackers producing tracks of different lengths, the normalization of ID changes is preferable to simply counting the ID changes. Unlike IDC and MOTA, NIDC is parameter independent since its assignment solution used for detecting ID changes. The NIDC value for ground-truth track *i* and $IDC_I^{\max}$ the maximum number of ID changes that can occur for ground truth track *i* (i.e. the length of track *i*) can be formulized as follows:

$$NIDC_i = \frac{|IDC_i|}{IDC_I^{\max}} \quad (8)$$

$NIDC_i$ includes a contribution of ID changes for track *i* that is scaled by IDC*max*, which is proportional to the duration of track *i*. This penalizes the ID changes by the length of the track in the estimation of NIDC, instead of simply relying on counting ID changes. NIDC quantifies the number of ID changes corresponding to all ground-truth tracks of the sequence.

In this work the measurement can be evaluated by five performance parameters and four accuracy parameters are used. The first performance evaluation measures is Distance based measures which is not suitable to evaluate changes due to variation in target size and fast moving objects. The Second parameter is Overlap based measures considers the estimated target size variations and can detect instances of tracking failure. Third parameter is Accuracy which is the closeness between estimated and ground truth states. Fourth parameter is Cardinality error which indicates the differences between the number of estimated and ground truth targets and the fifth parameter is ID changes which indicates the incorrect association between estimated and ground truth targets. The various Accuracy parameters are True Positive, True Negative, False Positive, false negative.

(i)True Positive Rate (TPR) indicates the correct prediction of target movements.

$$TPR = \frac{TP}{(TP + FN)} \quad (9)$$

(ii)True Negative Rate (TNR) indicates the prediction about the target movement succeeds initially, after that it fails.

$$TNR = \frac{TN}{(FP + TN)} \quad (10)$$

(iii)False Positive Rate (FPR) indicates the prediction about the target movement fails initially, after that it succeeds.

$$FPR = \frac{FP}{(TP + FP)} \quad (11)$$

(iv)False Negative Rate (FNR)indicates the correct prediction of target movements.

$$FNR = \frac{FN}{(TN + FN)}. \quad (12)$$

V. EXPERIMENTAL SETUP AND ANALYSIS

We analyze the efficiency of this work is measured by comparing them with performance and accuracy parameters of the recent work of trackers in real world dataset.

a) EXPERIMENTAL SETUP

In this work we use Four real-world datasets, namely *TownCenter*[18], *ETH Bahnhof*[19], *ETH Sunnyday*[19] and *iLidsEasy*[20]. TownCentre, recorded from an overhead static camera, is composed of 4491 frames of size 1920×1080 pixels recorded at 25 fps. The ground truth has 231 head/person-tracks with an average of 16 people per frame. ETH Bahnhof and Sunnyday, recorded from a human-height moving camera, are composed of 999 and 354 frames, respectively, with a frame size of 640 × 480 recorded at 14 fps. The ground truth of Bahnhof has 95 person-tracks with an average of eight people per frame, while that of Sunnyday has 30 person-tracks with an average of five people per frame. iLids Easy is composed of 5220 frames of size 720×576 pixels recorded at Westminster subway station (London, UK) at 25 fps. The ground truth has 17 person-tracks with an average of 1.9 people per frame. These data sets could be summarized in the table I

Table I. summary of dataset

Datasets	Frames	Pixels	Fps	Tracks	APF
Towncenter	4491	1920X1080	25	231	16
Eth bahnhof	999	640X480	14	95	8
Eth sunnyday	354	640X480	14	30	5
Ilidseasy	5220	720X576	25	17	1.9

APF-Average People per frame

b) TRACKERS

In Experimental validation we choose four trackers namely Kanade-Lucas- Tomasi [21] tracker with Markov-Chain Monte-Carlo Data Association [18] (MCMCDA) algorithm, a data association algorithm with the online learned Conditional Random Field Based Tracker [23] (CRFBT) , a Multi-Target Track-Before-Detect (MT-TBD) with a post-processing stage [24], and the Dynamic Programming Non-Maxima Suppression based tracker [25] (DP-NMS). Tracking includes head and person (full-body) tracks from both static and moving cameras. Dynamic Programming Non-Maxima Suppression based tracker (DP-NMS) is tested on TownCentre, ETH Bahnhof and Sunnyday, and iLids Easy sequences for person tracking. MT-TBD is used for head tracking on the TownCentre sequence and for person tracking on the ETH Bahnhof and Sunnyday, and iLids Easy sequences. MCMCDA is used for head tracking on TownCentre and for person tracking on TownCentre and iLids Easy sequences. CRFBT is tested on ETH Bahnhof and Sunnyday sequences for person tracking.

c) COMPARISON OF TRACKERS

Towncenter with head tracking (towcenter-H) and with person tracking (towncenter-p) shows that the evaluation result based on METE and N-MODA is contracted and also Sunnyday also decides these two measures based on the relative ranking of trackers. METE performance is effectively ranked on the basis of accuracy measures true positive, false positive and false negative. Both the tracker such as MT-TDB and DP-NMS shows the same result on the measure MOTP in case of Bahnhof but the MT-TDB is ranked as a high performance tracker than the DP-NMS based on MELT.

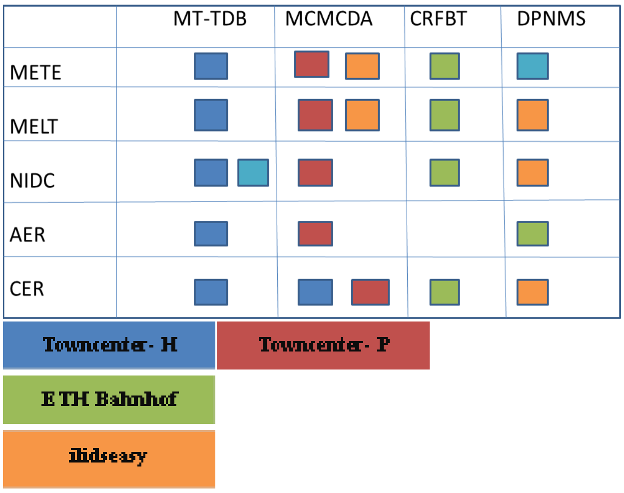

Fig. 2 Comparison of trackers

In order to select the best tracker in case of Sunnyday , MELT and MOTP shows the discrepancy. To conclude, MODA has the dependence over the threshold limits and it has the ability to distinguish results of various trackers whereas MOTP has the inaccurate evaluation over the tracking results due to its threshold dependency.

The Fig. 2 shows evaluation of various trackers MT-TDB, MCMCDA, CRFBT and DP-NMS with the basis of MELT, METE, NIDC, AER and CER on the various real world datasets towncenter, ETH bahnhof, ETH sunnyday and ilidseasy which gives information about best tracker of multi object tracking. In Fig. 2 also shows that MCMCDA is the best tracker while comparing to other trackers while next to MCMCDA, MT-TDB gives out the best tracking result. Fig. 3 shows the plot between multi extended target tracking error relative to the number of frames involved and also shows how the METE values changes in accordance with each frames. Normalised matching error is plotted in accordance with the number of frames that shows in Fig. 4. Generally (Fig 5)CardinalityError(CE) occurs due to overlap of background with the target image, the coincidence of target image with another image, target size variations, etc. This graph plots the cardinalty error rate with reapect to number of frames involved .

VI. CONCLUSION

We measures three measures of METE, MELT, NIDC trackers in multi target which is measured by key factors like accuracy, cardinality error and ID changes. In this paper, we obtained the effective video tracking results by comparing the estimated track with ground truth tracks with the consideration of various paramètres such as accuracy, cardinality error and normalised Id changes and also to obtain the best result the robust algorithm of METE is used which not only considers the position of the target and also considers the region of the target object. We compared the performance of the various trackers like MCMCDA, MT-TDB, CRFBT, DP-NMS, which is compared on the basis of measures like MODA, MOTP etc. and also shows that MCMCDA provides the best result of comparing to its competing trackers.

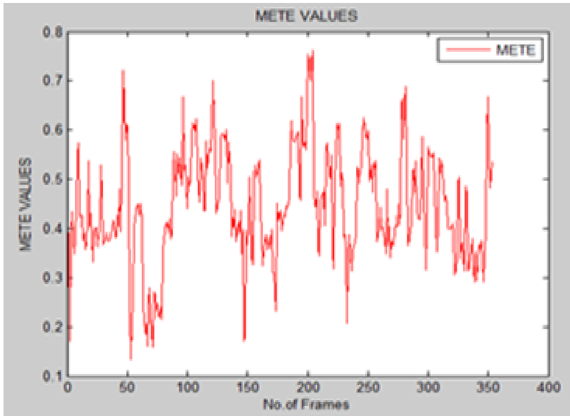

Fig. 3 METE values

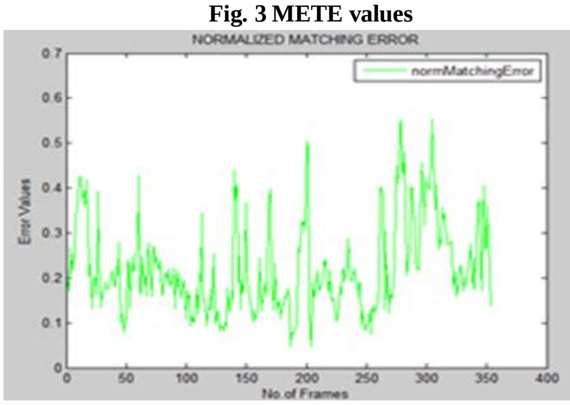

Fig. 4 Normalized matching error

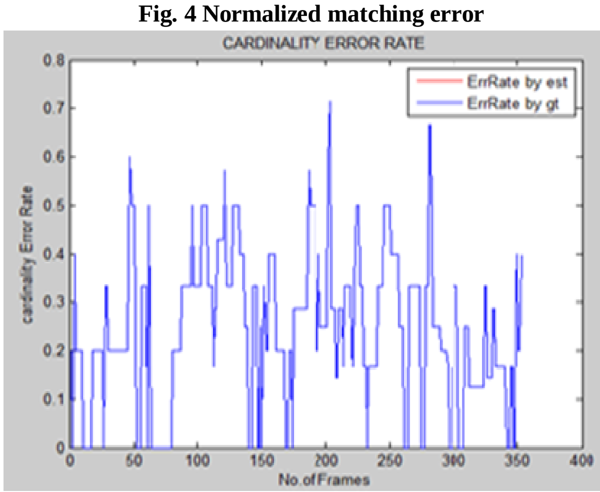

Fig. 5 Cardinality error rate

REFERENCES

[1] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, “A metric for performance evaluation of multi-target tracking algorithms,” IEEE Trans. Signal Process., vol. 59, no. 7, pp. 3452–3457, Jul. 2011.

[2] Manisha Chate, S.Amudha ,Vinaya Gohokar, “ object detection and tracking in video sequences” ACEEE Int. J. on Signal & Image Processing, Vol. 03, No. 01, Jan 2012.

[3] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, et al., “Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol,” IEEE Trans. Pattern Anal. Mach. Int., vol. 31, no. 2, pp. 319–336, Feb. 2009.

[4] T. Nawaz and A. Cavallaro, “A protocol for evaluating video trackers under real-world conditions,” IEEE Trans. Image Process., vol. 22, no. 4, pp. 1354–1361, Apr. 2013.

[5] J. Black, T. Ellis, and P. Rosin, “A novel method for video tracking performance evaluation,” in Proc. WPETS, 2003, pp. 125–132.

[6] F. Yin, D. Makris, and S. A. Velastin, “Performance evaluation of object tracking algorithms,” in Proc. 10th WPETS, 2007, pp. 1–8.

[7] E. Maggio and A. Cavallaro, Video Tracking: Theory and Practice. New York, NY, SUA: Wiley, 2011.

[8] C. Needham and R. Boyle, “Performance evaluation metrics and statis- tics for positional tracker evaluation,” in Proc. ICVS, 2003, pp. 125–132.

[9] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, “Minimum error bounded efficient l1 tracker with occlusion detection,” in Proc. CVPR, 2011, pp. 1257–1264.

[10] Tahir Nawaz, Fabio Poiesi, and Andrea Cavallaro, “Measures of effective video tracking ”, ieee transactions on image processing, vol. 23, no. 1, january 2014.

[11] Priti P. kuralkar, Human object tracking using background subtraction and shadow removal techniques.

[12] Vishwadeep Uttamrao Landge, Object detection and object tracking using background subtraction for surveillance application .

[13] Shao-Yi Chien, Wei-Kai Chan, Yu-Hsiang Tseng, and Hong-Yuh Chen, “Video Object Segmentation and Tracking Framework With Improved Threshold Decision and Diffusion Distance”, IEEE Trans. on Circuits And Systems For Video Technology, Vol. 23, No. 6, June 2013.

[14] Dong Wang Huchuan Lu Ming-Hsuan Yang, “Online Object Tracking with Sparse Prototypes”, IEEE Trans. on Image Processing, Vol. 22, No. 1, January 2013.

[15] Olga Zoidi, Anastasios Tefas, and Ioannis Pitas, ” Visual Object Tracking Based on Local Steering Kernels and Color Histograms”, IEEE Trans. on Circuits And Systems For Video Technology, Vol. 23, No. 5, May 2013.

[16] Siew Wen Chin, Kah Phooi Seng, and Li-Minn Ang, “Lips Contour Detection and Tracking Using Watershed Region-Based Active Contour Model and Modified H∞”,IEEE Trans. on Circuits And Systems For Video Technology, Vol. 22, No. 6, June 2012.

[17] Huchuan Lu, Shipeng Lu, Dong Wang, Shu Wang, and Henry Leung, “Pixel-Wise Spatial Pyramid-Based Hybrid Tracking”, IEEE Trans. on Circuits And Systems For Video Technology, Vol. 22, No. 9, September 2012.

[18] B. Benfold and I. Reid, “Stable multi-target tracking in real-time surveillance video,” in Proc. CVPR, Jun. 2011, pp. 3457–3464.

[19] (2012, Aug.). ETH Bahnhof and Sunnyday Datasets[Online].Available:http://www.vision.ee.ethz.ch/ ˜ aess/iccv2007/

[20] (2012, Oct.). i-Lids Easy Dataset [Online]. Available:http://www.eecs.qmul.ac.uk/~andrea/avss2007 _d.html

[21] C. Tomasi and T. Kanade, “Detection and tracking of point features,” Dept. Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-91-132, Apr. 1991.

[22] B. Yang and R. Nevatia, “An online learned CRF model for multi-target tracking,” in Proc. CVPR, Jun. 2012, pp. 2034–2041.

[23] F. Poiesi, R. Mazzon, and A. Cavallaro, “Multitarget tracking on confidence maps: An application to people tracking,” Comput. Vis. Image Understand., vol. 117, no. 10, pp. 1257–1272, Oct. 2013.

[24] H. Pirsiavash, D. Ramanan, and C. Fowlkes, “Globally-optimal greedy algorithms for tracking a variable number of objects,” in Proc. CVPR, Jun. 2011, pp. 1201–1208.

[25] Wenxi Liu, Antoni B. Chan, Rynson Lau, Dinesh Manocha, “Leveraging Long-Term Predictions and Online Learning in Agent-Based Multiple Person Tracking”, *IEEE Trans. Circuits Syst. Video Techn.*, vol. 25, pp. 399-410, 2015.