

Multiple Object Tracking with Kernelized Correlation Filters in Urban Mixed Traffic

Yuebin Yang*

University of Electronic Science and Technology of China
Chengdu, China
Email: patrick.ybyang@gmail.com

Guillaume-Alexandre Bilodeau

LITIV lab. Dept. of computer & software eng.
Polytechnique Montréal, Canada
Email: gabilodeau@polymtl.ca

Abstract—Recently, the Kernelized Correlation Filters tracker (KCF) achieved competitive performance and robustness in visual object tracking. On the other hand, visual trackers are not typically used in multiple object tracking. In this paper, we investigate how a robust visual tracker like KCF can improve multiple object tracking. Since KCF is a fast tracker, many KCF can be used in parallel and still result in fast tracking. We built a multiple object tracking system based on KCF and background subtraction. Background subtraction is applied to extract moving objects and get their scale and size in combination with KCF outputs, while KCF is used for data association and to handle fragmentation and occlusion problems. As a result, KCF and background subtraction help each other to take tracking decision at every frame. Sometimes KCF outputs are the most trustworthy (e.g. during occlusion), while in some other cases, it is the background subtraction outputs. To validate the effectiveness of our system, the algorithm was tested on four urban traffic videos from a standard dataset. Results show that our method is competitive with state-of-the-art trackers even if we use a much simpler data association step.

Keywords—Multiple object tracking; Correlation filters; Urban scenes; Road users;

I. INTRODUCTION

Multiple object tracking (MOT) is a fundamental task in computer vision with numerous conceptually diverse tracking algorithms being proposed every year. A popular application area is traffic surveillance because of its practical use for reducing traffic jam and for assessing the security of various road configurations. However, the existing intelligent transportation systems (ITS) exhibit a low performance when faced with problems like occlusions, illumination changes, motion blur and other environmental variations [1]. To address these problems, we propose a multiple object tracker based on a robust visual object tracker. Using a visual object tracker allows our method to handle occlusions in a straight-through manner (objects are tracked individually during occlusions) even if the detected objects are obtained as blobs from background subtraction. This contrast with the typical merge-split approach used by recent trackers based

on background subtraction [2], [3]. Furthermore, the use of a robust visual tracker reduces the need for a complex data association method. Our proposed method is online meaning that it takes decision based on past frames only.

In this paper, we combine background subtraction with KCF (Kernelized Correlation Filters) tracker [4] and propose a promising model-free MOT method for road users in urban mixed traffic. Our method is based on background subtraction because urban mixed traffic can include unexpected objects. As such, a trained object detector is difficult to design so that it does not miss objects. Background subtraction provides us with noisy candidate object regions. We analyze and process each candidate region to get higher-quality potential objects. For tracking, we create a KCF tracker for each candidate region using a grayscale and color names appearance model. Sequentially, moving objects are tracked and their states are updated frame by frame. To handle scale variations, object fragmentation, occlusions and lost tracks, we use simple data association between KCF trackers and targets detected by background subtraction. Object states are determined by using the information of both KCF and background subtraction as they can both make errors at different times. Finally, the system is evaluated with four different urban videos involving cars, pedestrians, bikes, and trucks. Results show that our method is competitive even if using simple data association, demonstrating the benefit of using a strong visual object tracker in a MOT framework.

This paper is organized as follows: In section II, we discuss related work, in section III, we present our proposed method, including extracting regions of interest (ROI), background subtraction, foreground analysis, and the tracking of objects. In section IV, we test our method and summarize our results. Finally, in section V, we conclude the paper.

II. RELATED WORK

For tracking road users, there exist basically two approaches depending on the way the road users are detected, which is either by using optical flow, or by using background subtraction. Methods based on pre-trained bounding box detectors are not applied in road user tracking scenarios because it is difficult to design a universal detector that

*This work was conducted while Yuebin Yang was doing a MITACS Globalink internship at Polytechnique Montréal

can detect all possible road users from various viewpoints. Objects may be missed. The use of background subtraction and optical flow is not without limitation either, because although all objects are usually detected, they are often fragmented into two or more parts, or they can be merged.

A. Optical flow-based methods

In this family of methods, objects are detected using optical flow by studying the motion of tracked points in a video. Feature points that are moving together are considered as belonging to the same object. Several methods accomplished this process using the Kanade-Lucas-Tomasi (KLT) tracker [5]. Among others, the trackers of Beymer *et al.* [6], Coifman *et al.* [7], Saunier *et al.* [8] and Aslani and Mahdavi-Nasab [9]. For example, Saunier *et al.* [8] method, named *Traffic Intelligence*, tracks all types of road users at urban intersections by continuously detecting new features and adding them to current feature groups. Then, the right parameters should be selected to segment objects moving at similar speeds, while at the same time not over-segmenting smaller non-rigid objects such as pedestrians. Because objects are identified only by the motion of feature points, nearby road users moving at the same speed may be merged together. Furthermore, the exact area occupied in the frame by the road user is unknown because it depends on the position of sparse feature points. Finally, when an object stops, its features flow interrupts which usually leads to fragmented trajectories.

B. Background subtraction-based methods

This second family of methods rely on background subtraction to detect road users in a video. Background subtraction gives blobs that can correspond to parts of objects, one object, or many objects grouped together. The task is then to distinguish between merging, fragmentation, and splitting of objects. This approach works fairly well under conditions with no or little occlusion. However, under congested traffic conditions or slow speed conditions, vehicles are harder to extract from the background as they may partially occlude with each other and many objects may merge into a single blob. In this case, it is hard to divide them.

Examples of trackers based on background subtraction include the work of Fuentes and Velastin [10], Torabi *et al.* [3], Jun *et al.* [11], Kim *et al.* [12], Mendes *et al.* [13], and Jodoin *et al.* [2].

Fuentes and Velastin [10] proposed a method that performs simple data association via the overlap of foreground blobs between two frames. In addition to matching blobs based on overlap, Torabi *et al.* [3] validated the matches by comparing the histograms of the blobs and by verifying the data association over short time windows using a graph-based approach. These approaches track objects in a merge-split manner as objects are tracked as groups during occlusion.

Jodoin *et al.* [2], [14] proposed *Urban tracker*, which is a tracker that in addition to background subtraction uses feature points to model the appearance of the tracked objects. This helps in solving the data association problem in the case of occlusion. The position of the bounding box of occluded objects can even be predicted based on the visible feature points and their previous positions on the object. A finite state machine is used for handling fragmentation and occlusion. The method follows initially the split-merge paradigm but after solving occlusion, backtracks and segments the occluded objects.

Jun *et al.* [11] used background subtraction to estimate the object properties. A watershed segmentation technique is used to over-segment the vehicles. The over-segmented patches are then merged using the common motion information of tracked feature points. This allows to segment vehicles correctly even in the presence of partial occlusion. Kim *et al.* [12] combines background subtraction and feature tracking approach with a multi-level clustering algorithm based on the Expectation-Maximization (EM) algorithm to handle the various object sizes in the scene. The resulting algorithm tracks various road users such as pedestrians, vehicles and cyclists online and the results can then be manually corrected in a graphical interface. Mendes *et al.* [13] proposed a method that also combines KLT and background subtraction. This approach increases its accuracy by setting *in* and *out* regions to reconfirm object trajectories and by segmenting regions when occlusions are detected.

In our work, we combine KCF and background subtraction and propose a model-free MOT method that can track any types of road users without prior knowledge. The advantage of our method is that it addresses the occlusion problem by tracking object individually inside blobs of merged objects. Backtracking is not necessary, neither is the use of an explicit region-based segmentation.

III. TRACKING METHODOLOGY

Our algorithm combines background subtraction with the KCF tracker. Background subtraction is used for extracting the foreground (candidate road users to track), and the KCF tracker is applied for data association and for tracking objects during occlusions. To improve the overall performance of our method, we added several refinements to solve problems like occlusion and scale adaptation. Figure 1 shows the diagram of the steps of our method.

For each frame, we extract the foreground to obtain candidate object regions (blobs). Then, we analyze the blobs and apply several filters to obtain a final list of candidate object regions COR_i . Next, to update the tracked object states, we compare the COR_i with current tracks CT_j for data association. This is done by updating all the currently active KCF trackers. The KCF trackers KCF_j^{t-1} find the most plausible data association with a COR_i based on proximity and their internal model. Correspondences may

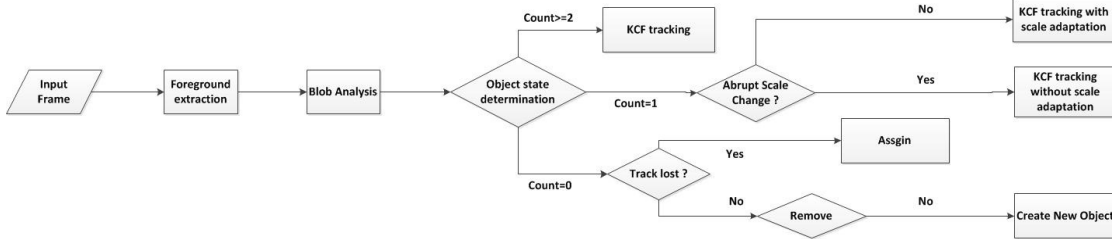


Figure 1: Diagram of the steps of our proposed method

be many to one, one to many, one to one, or an object may be entering and leaving the scene. These different cases are determined by counting the number of KCF trackers (one per CT_j) associated to each COR_i . For objects in occlusion, we update their positions without adapting scale (KCF is more trustworthy, we use the object state provided by KCF tracker). For one to one association, we update the state of the tracked object using the information from background subtraction (Background subtraction allows adjusting the scale of the KCF tracker, and in one to one association, background subtraction is assumed to be often trustworthy). For a new object, we assign a new KCF tracker. Finally, undesired objects will be deleted if not detected by background subtraction during several frames. In the next subsections, we describe our method in detail (refer to Figure 1).

A. Foreground detection

Foreground detection allows us to obtain candidate object regions COR_i . Blobs B_i are extracted at each frame. Foreground detection is used to begin new tracks and to validate the existing ones. If a detected blob B_i is not associated to a KCF trackers, a new tracker will be initialized on that blob. If a KCF tracker is tracking an object and that object is not detected for many frames, this may indicate a failure of the KCF tracker. In this case, tracks are discarded. Any foreground detection method can be used with our proposed method. Since we are using a publicly available dataset with foregrounds detected using LOBSTER [15], our method has been tested using this method.

B. Blob Analysis

At this step, we process extracted blobs B_i with several operations to get improved candidate object regions COR_i . Morphological operations including median filtering, closing and hole filling are applied to each B_i . To minimize the influence of environmental disturbance, regions smaller than a threshold T_r or with inappropriate width to height ratio (e.g. objects that are too thin) are discarded. To partially solve fragmentation problems that may occur, regions that are very close spatially (based on a threshold T_c) are merged together. After these operations, we get the final candidate object regions COR_i . These regions are then used as input for tracking.

C. Object Tracking

The advantages of background subtraction are that: 1) it can detect any objects even the ones that were never observed before, and 2) it can provide scale information about the object to track. However, background subtraction also has limitations in the processing of traffic scenes: 1) under heavy traffic conditions, objects move slowly and may be integrated in the background model, thus some may not be detected correctly, and 2) when partial occlusion occurs, several objects will merge into one large blobs B_i instead of several blobs. These cases can be handled more easily by the KCF tracker. Thus, our method will combine both approaches to handle these various conditions.

KCF [4] is a correlation-based tracker that analyzes frames in the Fourier domain for faster processing. It does not adapt to scale changes but updates its appearance model at every frame. KCF can be used with various appearance models. In this work, we use both grayscale intensities and color naming as suggested by the authors of [16].

1) *Finding the object states*: Before deciding if at frame t a COR_i^t will be tracked mostly with the KCF tracker or with background subtraction, we need to determine in which condition it is currently observed. The COR_i^t may be entering the scene, leaving the scene, isolated from other objects, or in occlusion. To determine the states of all COR_i^t , the active KCF trackers at frame $t-1$, KCF_j^{t-1} , are applied to the original RGB color frame t . Then the resulting tracker outputs TO_j^t are tested for overlap with each COR_i^t and vice versa:

- If a COR_i^t overlaps with only one TO_j^t , it means that the object is isolated and currently tracked (*state: tracked*).
- If a COR_i^t overlaps with more than one TO_j^t , it means that the object is tracked but under occlusion (*state: occluded*).
- If a COR_i^t does not overlap with any TO_j^t , it means that it is an object not currently tracked (*state: new object*).
- If a TO_j^t does not overlap with any COR_i^t , it means that it is an object not currently visible (*state: invisible or object that has left*).

The overlap is evaluated using

$$Overlap(x, y) = \frac{x \cap y}{x \cup y}, \quad (1)$$

where x and y are two bounding boxes.

2) *State: Tracked*: If a COR_i^t overlaps with one and only one TO_j^t , we assume that the object is isolated and tracked correctly (but not necessarily precisely). Since the KCF tracker does not adapt to scale change, we add the COR_i^t bounding box to the CT_j^{t-1} associated to KCF_j^t most of the time. This way the scale can be adapted progressively during tracking based on the blob COR_i^t from background subtraction. In this case, we re-initialize KCF_j^t with the information for COR_i^t to continue the tracking.

We use TO_j^t only if COR_i^t suddenly shrinks because of fragmentation of the tracked target. Indeed, background subtraction cannot always handle well abrupt environmental changes which cause over-segmentation. Therefore, when the KCF tracker bounding box is much larger than the background subtraction bounding box (as defined by thresholds T_{ol} and T_{oh}), we assume that KCF is more reliable. Formally,

$$CT_j^t = \begin{cases} CT_j^{t-1} \cup TO_j^t, & \text{if } T_{ol} \leq \frac{A(TO_j^t)}{A(COR_i^t)} \leq T_{oh} \\ CT_j^{t-1} \cup COR_i^t, & \text{otherwise} \end{cases}, \quad (2)$$

where $A()$ is the area of a bounding box. The upper bound threshold T_{oh} is used to allow adaptation of the scale when the object is leaving the scene. In this case, the object shrinks but it is not caused by over-segmentation, so the use of COR_i^t should be considered. Note that when an object is fragmented, KCF_j^t will be associated with the larger fragment. Therefore, the shrinking is conditioned by the larger fragment, and as a result, normally the shrinking is limited. If it is too large, most of the time it is explained by an object that is leaving the video frame. This is why T_{oh} is used.

Finally, even if COR_i^t is significantly larger than TO_j^t , we still used the background subtraction blob as shadows are not problematic in our application and COR_i^t localization accuracy is better than TO_j^t from KCF.

3) *State: Occluded*: if more than one TO_j^t overlap with a COR_i^t , we assume that several previously isolated objects are now in occlusion. Background subtraction fails to distinguish objects as they are merged into one blob. In this case, we handle the occlusion problem in a straight-through manner and we use the bounding boxes outputted by the KCF trackers. That is, we add the TO_j^t to the CT_j^{t-1} related to COR_i^t with

$$CT_j^t = CT_j^{t-1} \cup TO_j^t. \quad (3)$$

Background subtraction cannot handle over-segmentation problems in some conditions. As it can be seen from Figure

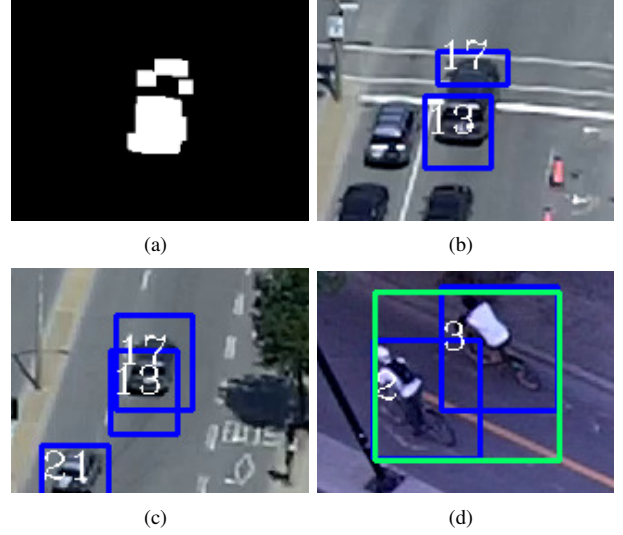


Figure 2: Example of redundant KCF trackers. (a) Extracted regions, (b) Object starts moving, (c) Redundant trackers, (d) Parallel moving objects.

2(a, b, and c), when one object starts moving, redundant KCF trackers may be created on its segmentation as the object may be temporarily over-segmented. This is not a real case of occlusion. As time goes by, trackers tracking the same object may partially overlap with each other (As shown in Figure 2c). We proposed a method to address this problem. If the area of the blob COR_i^t is smaller than the sum of the bounding boxes TO_m^t and TO_n^t of these two KCF trackers for eight consecutive frames, it is very likely to be two trackers tracking the same object. In this case, we delete the shorter-lived KCF tracker and we update the other one with COR_i^t . If not, we assume that two objects are really occluded and do not take any action (see Figure 2d). That is,

$$\text{if } A(TO_m^t) + A(TO_n^t) > A(COR_i^t), \quad \text{Delete } KCF_n^t \\ \text{otherwise, no action}, \quad (4)$$

where $A()$ corresponds to the area of a bounding box, and KCF_n^t is the most recent of the two KCF trackers tracking the same object.

4) *State: New object*: Because the KCF tracker sometimes loses track of objects when occlusion occurs, we first need to identify whether COR_i^t is not overlapping with any TO_j^t because of lost track or because of a possible new object. A problem that sometimes occurs is that after an occlusion two KCF trackers are on a single object, while the other object that was in occlusion is associated with no tracker. This is because of the continuous update process of KCF that can make the model drift during occlusions. In

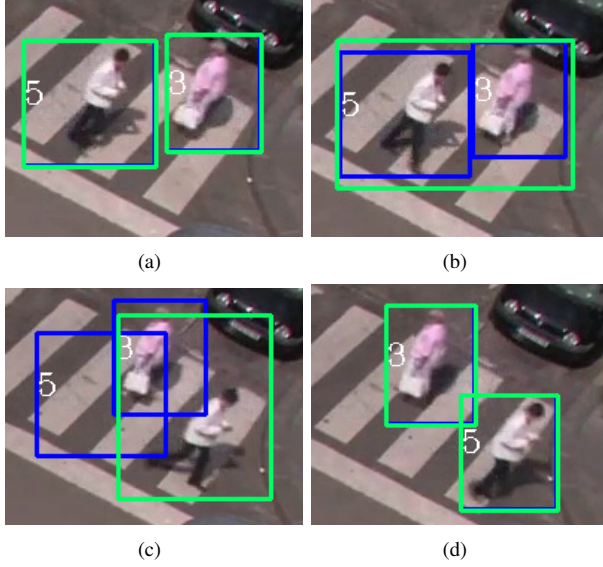


Figure 3: An example of handling tracker drift problem. (a) Before occlusion, (b) Group during occlusion, (c) Two trackers on the same object, none on the other one, (d) Relabeling after split.

that case, a tracker should be re-assigned to the other object.

To address this issue, we keep track of groups of objects. As it can be seen from Figure 3, when occlusion occurs, we label occluding objects as being inside a specific group. As soon as a member splits from the group without being tracked, we search for redundant KCF trackers among other group members. When more than one KCF trackers are found within a same group member, we re-assign the tracker that is matching less.

If it is indeed a new object, a new track is initialized with

$$CT_j^t = COR_i^t. \quad (5)$$

5) *State: Invisible or exited object*: When a TO_j^t does not overlap with any COR_i^t , we considered that the object is *invisible*. If the object has exited the scene it will not reappear, but if it was hidden, it should reappear eventually. To distinguish these two cases, we use a rule based on the number of consecutive frames for which the object is *invisible*. If the object is invisible for more than eight consecutive frames, it will be removed from active tracks. Based on objects average size and moving speed, eight frames equals generally to the diagonal of an object bounding box. To minimize environmental disturbances that cause false detection blobs, such as camera shake, objects that exist for less than six frames will be discarded.

To improve tracking performance, when parts of an object track are missing because the object was at time *invisible*, we restore the missing parts by using interpolation from neighboring frames when the object was successfully tracked.

Our method is summarized by algorithm 1.

Algorithm 1 Multiple object tracking with KCF

```

1: Input: video
2: Output: trajectories + bounding boxes
3: procedure TRACKING
4:   for each video frame do
5:     Extract foreground
6:     for Each extracted region do    ▷ Section III-B
7:       Apply morphological operations
8:       Merge small blobs
9:       Delete unlikely regions
10:    end for
11:    for Each KCF tracker do
12:      Find best matching blob (section III-C1)
13:    end for
14:    Track objects based on current state
15:    if State == Tracked then
16:      Adapt scale and update KCF tracker (section
III-C2)
17:    else if State == Occluded then
18:      Track in straight-through manner (section
III-C3)
19:    else if State == New object then
20:      Create a new KCF tracker (section III-C4)
21:    else if State == Invisible or exited object then
22:      Delete or save trajectory (section III-C5)
23:    end if
24:  end for
25: end procedure

```

IV. EXPERIMENTS

To verify the efficiency of our proposed method, we tested our algorithm on four challenging video sequences from a publicly available dataset [2]. Sample frames of these four videos are shown in figure 4. These videos include cars, cyclists, pedestrians, trucks and buses. The dataset provides tracks that are annotated for all objects that are sufficiently large. We evaluate our method (named MKCF, for multiple KCF) by comparing it with three other trackers: Urban Tracker (UT) [2], Traffic Intelligence (TI) [8] and the tracker of [13] (Mendes et al.) that we implemented based on the information provided in their paper. Result shows that our algorithm is competitive and promising.

Our method was implemented in C++ and can be downloaded from <https://github.com/iyybpatrick/MKCF>.

A. Parameters and performance metrics

To test our method, we varied two parameters for each video. Blob size T_r is a threshold for filtering regions of small size and the centroid distance T_c is for merging segmented regions in the foreground (see section III-B). The thresholds we adopt for each video are listed in table I. T_{ol}

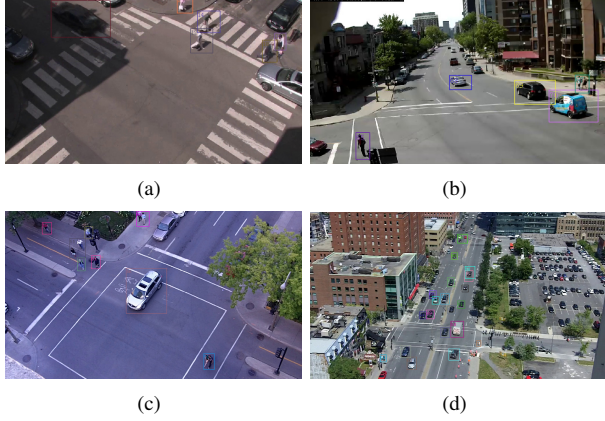


Figure 4: Samples results of our method on the four videos of the Urban Tracker dataset [2]. (a) Rouen, (b) Sherbrooke, (c) St-Marc, (d) Rene-Levesque

Video	Parameters	
	T_r	T_c
Sherbrooke	23	44
Rouen	41	63
St-Marc	35	55
Rene-Levesque	20	24

Table I: Parameters for each video

and T_{oh} are fixed to 1.4 and 1.8, respectively. The other tested methods also used different parameter sets for each video.

As for the evaluation part, we use the tools provided with the Urban Tracker dataset to calculate CLEAR MOT metrics [17]. MOTA is for calculating the multiple object tracking accuracy, which is evaluated based on miss rate, false positive and ID changes. MOTP is for calculating the multiple object tracking precision. It measures the average precision of instantaneous object matches. UT, MKCF and Mendes et al. methods all used the same background subtraction results provided the dataset.

B. Results and analysis

Quantitative results on the Urban Tracker dataset are provided in table II. For the Sherbrooke video, our method is competitive with UT and is the second best for the MOTA. In this video, there is a traffic light at the center of the frame. Cars and pedestrians from different directions stop alternatively. Our method often considers two or more objects as one because objects are merged when they first appear in the video. This problem is hard to address because background subtraction fails to segment occluding objects. Moreover, again because of background subtraction, redundant KCF trackers may be created on segmentation of one car because when a car starts moving after a long stop, sometimes it is fragmented into many blobs (see Figure 2a and b). It is difficult to distinguish redundant KCF trackers from many

KCF trackers tracking objects in occlusion (see Figure 2c and d). Therefore, redundant trackers cannot be discarded because we are not sure whether they are segments of one object or several isolated objects moving in the same direction. A more complex scheme would be required to handle this case.

For the Rouen video, our method also ranks second behind UT. For the St-Marc video, although the score of cars is much lower than that of other videos, our proposed method outperforms UT globally and in tracking the pedestrians. Results are lower for cars because more than one KCF trackers are assigned to several segments of one object when the object starts moving, which is the same problem discussed for the Sherbrooke video. This problem affects mostly cars.

For the Rene-Levesque video, our method exhibits the best MOTA for cars and cyclists. However, considering all the objects in the scene, our method performs second because pedestrians in this video are very small. Setting a smaller T_r would improve the performance for tracking pedestrians, but the overall performance will decrease because many environment disturbance cannot be eliminated.

Globally, our method is competitive with the state-of-the-art tracker UT despite not using a complex data association scheme. We just select between background subtraction blobs and KCF bounding boxes. This shows that a state-of-the-art tracker like KCF can help multiple object tracking as demonstrated with performances close to UT. With a more complex data association scheme, the fragmentation problem limitations when initializing a new KCF tracker could be addressed. This problem occurs essentially when an object stops for a long time in the scene. It does not occur for active KCF trackers. Thus, for objects that do not stop for a long time, the use of the KCF tracker is beneficial for handling temporary fragmentation.

Finally, we also evaluated the speed of our proposed method. Computation times are provided in table III. Reported processing times were obtained on a computer with an intel Core i5-6267U running at 2.9Ghz with 8 Gigabytes of RAM.

V. CONCLUSION

In this paper, we presented a multiple object tracker that combines a visual object tracker with background subtraction. Background subtraction is applied to extract moving objects and get their scale and size in combination with KCF outputs, while KCF is used for data association and to handle fragmentation and occlusion problems. As a result, KCF and background subtraction help each other to take tracking decision at every frame. This mixed strategy enables us to get competitive results and reduce errors under complex environment. The advantage of our method is that it addresses the occlusion problem by tracking object individually inside blobs of merged objects. Backtracking is

Video	Type	MKCF (Ours)		UT		TI		Mendes et al.	
		MOTA	MOTP	MOTA	MOTP	MOTA	MOTP	MOTA	MOTP
Sherbrooke	Cars	0.789	11.82	0.887	<i>10.59</i>	0.825	7.42	0.707	13.21
	Pedestrians	<i>0.671</i>	7.63	0.705	6.61	0.014	11.98	0.601	7.97
	All Objects	<i>0.763</i>	10.55	0.787	<i>8.64</i>	0.3841	7.54	0.695	9.95
Rouen	Cars	0.813	14.86	<i>0.896</i>	9.73	0.185	66.69	0.918	<i>11.41</i>
	Pedestrians	<i>0.804</i>	<i>14.81</i>	0.830	13.77	0.647	20.04	0.672	14.88
	Cyclists	<i>0.890</i>	14.67	0.927	14.13	0.869	13.11	0.881	<i>13.67</i>
	All Objects	<i>0.813</i>	14.86	0.844	13.19	0.589	24.20	0.718	<i>14.29</i>
St-Marc	Cars	0.590	12.54	0.889	10.90	-0.178	38.99	<i>0.713</i>	<i>11.44</i>
	Pedestrians	0.834	7.35	<i>0.730</i>	5.05	0.693	10.44	0.505	7.09
	Cyclists	<i>0.975</i>	6.33	0.989	6.39	0.895	7.46	0.935	6.30
	All Objects	0.825	7.93	<i>0.764</i>	5.99	0.602	14.58	0.560	7.71
Rene-Levesque	Cars	0.855	2.72	<i>0.796</i>	<i>3.04</i>	0.547	5.23	0.163	7.09
	Cyclists	0.291	2.77	<i>0.232</i>	2.20	0.232	3.14	0.216	2.26
	All Objects	<i>0.572</i>	2.74	0.723	2.98	0.503	5.10	0.402	2.95

Table II: Results on the Urban Tracker dataset. MOTP values are in pixels. MOTA should be high, while MOTP should be low for the best performances. **Boldface**: best result, *Italic*: second best.

Video	Resolution	Number of frames	Processing time (s)	FPS
Sherbrooke	800x600	1001	27.7	36.2
Rouen	1024x576	600	53.1	11.3
St-Marc	1280x720	1000	98.2	10.2
Rene-Levesque	1280x720	1000	53.8	18.6

Table III: Processing times of MKCF. FPS: Frames per second

not necessary, neither is the use of an explicit region-based segmentation.

Future work is to design methods to handle stopping objects. Because of background subtraction, they are progressively integrated into the background model. So, when they start moving again they are considered as one or many new objects. Furthermore, we should better distinguish exiting objects from stopping objects.

REFERENCES

- [1] A. Lessard, F. Belisle, G.-A. Bilodeau, and N. Saunier, "The countingapp, or how to count vehicles in 500 hours of video," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.
- [2] J. P. Jodoin, G. A. Bilodeau, and N. Saunier, "Tracking all road users at multimodal urban traffic intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–11, 2016.
- [3] A. Torabi and G. A. Bilodeau, "A multiple hypothesis tracking method with fragmentation handling," in *Computer and Robot Vision, 2009. CRV '09. Canadian Conference on*, 2009, pp. 8–15.
- [4] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, March 2015.
- [5] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, 1994, pp. 593–600.
- [6] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, "A real-time computer vision system for measuring traffic parameters," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 495–501.
- [7] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 6, pp. 271–288, 1998.
- [8] N. Saunier and T. Sayed, "A feature-based tracking algorithm for vehicles in intersections," in *The 3rd Canadian Conference on Computer and Robot Vision, 2006*, 2006, pp. 59–59.
- [9] S. Aslani and H. Mahdavi-Nasab, "Optical flow based moving object detection and tracking for traffic surveillance," *International Journal of Electrical, Robotics, Electronics and Communications Engineering*, vol. 7, no. 9, pp. 773–777, 2013.
- [10] L. M. Fuentes and S. A. Velastin, "People tracking in surveillance applications," *Image and Vision Computing*, vol. 24, no. 11, pp. 1165–1171, 2006.
- [11] G. Jun, J. K. Aggarwal, and M. Gokmen, "Tracking and segmentation of highway vehicles in cluttered and crowded scenes," in *Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision*, ser. WACV '08. Washington, DC, USA: IEEE Computer Society, 2008, p. 16.
- [12] Z. Kim, "Real time object tracking based on dynamic feature grouping with background subtraction," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, 2008, pp. 1–8.

- [13] J. C. Mendes, A. G. C. Bianchi, and A. R. P. Júnior, "Vehicle tracking and origin-destination counting system for urban environment." in *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP 2015)*, 2015.
- [14] J. Jodoin, G. Bilodeau, and N. Saunier, "Urban tracker: Multiple object tracking in urban mixed traffic," in *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, March 24-26, 2014*, 2014, pp. 885–892. [Online]. Available: <http://dx.doi.org/10.1109/WACV.2014.6836010>
- [15] P. St-Charles and G. Bilodeau, "Improving background subtraction using local binary similarity patterns," in *IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, March 24-26, 2014*, 2014, pp. 509–515. [Online]. Available: <http://dx.doi.org/10.1109/WACV.2014.6836059>
- [16] M. Danelljan, F. S. Khan, M. Felsberg, and J. v. d. Weijer, "Adaptive color attributes for real-time visual tracking," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1090–1097.
- [17] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *J. Image Video Process.*, vol. 2008, p. 1:11:10, jan 2008.