A System for Visual Analysis of Objects Behavior in Surveillance Videos

Cibele Mara Fonseca
Faculty of Computing
Federal University of Uberlandia
Uberlandia - MG - Brazil
Email: cibeleemara@gmail.com

Jose Gustavo S. Paiva
Faculty of Computing
Federal University of Uberlandia
Uberlandia - MG - Brazil
Email: gustavo@ufu.br

***Abstract*—Closed-circuit television (CCTV) surveillance systems are employed in different scenarios to prevent a variety of threats, producing a large volume of video footage. Several surveillance tasks consist of detecting/tracking moving objects in the scene to analyze their behavior and comprehend their role in events that occur in the video. Such analysis is unfeasible if manually performed, due to the large volume of long duration videos, as well as due to intrinsic human limitations, which may compromise the perception of multiple strategic events. Most of smart surveillance approaches designed for moving objects analysis focus only on the detection/tracking process, providing a limited comprehension of objects behavior, and rely on automatic procedures with no/few user interaction, which may hamper the comprehension of the produced results. Visual analytics techniques may be useful to highlight behavior patterns, improving the comprehension of how the objects contribute to the occurrence of observed events in the video. In this work, we propose a video surveillance visual analysis system for identification/exploration of objects behavior and their relationship with events occurrence. We introduce the *Appearance Bars* layout to perform a temporal analysis of each object presence in the scene, highlighting the involved dynamics and spatial distribution, as well as its interaction with other objects. Coordinated with other support layouts, these bars represent multiple aspects of the objects behavior during video extent. We demonstrate the utility of our system in surveillance scenarios that shows different aspects of objects behavior, which we relate to events that occur in the videos.**

I. Introduction

The worldwide increasing on the security concern led to the popularization of Closed-circuit television (CCTV) systems [1], producing large volumes of recording footage. This material holds a rich informational content, whose analysis is useful for comprehend all the phenomena captured in scene, and may help solving crimes and thefts, understanding people movement, identifying suspicious behavior, among other tasks. However, the manual analysis of these videos is an unfeasible task due to the large volume of videos, as well as their long duration. In addition, due to intrinsic human limitations, the occurrence of multiple strategic events can be unnoticed by the security agents, hampering the analysis process.

Several surveillance video analysis tasks focus on detecting/tracking moving objects in the scene to study their behavior [2]. As these objects are the elements directly involved in the occurrence of strategic events, such analysis may provide a better comprehension of these events. Aspects related to the objects behavior include time presence (when a object is in the scene, when it is not), trajectories (positions occupied by the object in specific time instants), movement patterns and relationships among them, among others.

Several computational systems exist to analyze surveillance videos contents in a variety of scenarios [3], but most of them focus only on the objects detection and tracking [4]. Some applications focus on automatic behavior analysis [5], [6], but provide only basic user interaction, which may limit the comprehension of the results produced by these approaches. Visual Analytics strategies have been used to improve the video contents comprehension, providing effective views of different aspects of the videos [7], such as objects trajectories [8], events identification and summarization [9], among others, but to the extent of our knowledge, no approach focus on the objects behavior in the scene, as well as on the relationship and interaction among them. We believe such analysis may help users in identifying and comprehending the occurrence of strategic events in the video.

We propose in this work a computational system for visual analysis of surveillance video, with focus on the identification and exploration of objects behavior and their relationship with events occurrence. Our approach employs the result of automatic and/or manual detection/tracking procedures to create an effective representation of several aspects of objects behavior, such as their presence in scene, relationships/interaction among them, movement and scene occupation. The system uses three coordinated interactive layouts: *Appearance Bars View*, *Brush View* and *Frame View*. The main one, *Appearance Bars View*, depicts the dynamics and spatial distribution of each object during its presence in scene, revealing detailed information about the moments in which it is in the scene, interacts with other objects, occupies specific regions in the scene, as well as its average speed variation. Our proposed layouts support the identification of objects behavior and allows the comprehension of how these behavior influence on the relevant events occurred in surveillance videos. The contributions of this work are listed as follows:

- A surveillance video visual analysis system with focus on summarizing/exploring object behavior in the scene, alone and/or interacting with other objects, and how these

behavior relates to the occurrence of events in the video;
- The system validation through a series of case studies considering various surveillance scenarios, including different movement patterns and types of events.

The following sections describe related work, detail our approach and the proposed computational system, discuss the results of the case studies and present our conclusions.

II. Related Work

Smart surveillance systems employ automatic video analysis techniques aiming to identify important aspects in surveillance videos [3], supporting several surveillance tasks. One important task is the analysis of the objects in scene, whose actions are highly related to strategic events occurring in the video. A crucial step of such analysis is the detection/tracking of these objects using a variety of computer vision approaches [4], whose results are used by Machine Learning techniques for automatic behavior analysis, including activity recognition [10], abnormal behavior detection [6], crowd analysis [5] and trajectory analysis [11]. These applications however focus only on basic aspects of objects detection/tracking, such as their identification, clustering and classification. Moreover, these systems often implement automatic procedures, and provide few or no user intervention in the process, which may limit the comprehension of more sophisticated action patterns. Our proposal aims at providing an effective representation of objects behavior to the user, so he/she is able to better comprehend aspects related to their actions and relationships among them.

Several visual analytics methods exist to effectively communicate strategic information from surveillance videos and to assist security agents in analysis tasks [7]. Mendes et al. [9] present a methodology for video summarization with focus on event identification, employing a point-placement visualization technique to highlight events spatial aspects and a Temporal Self-similarity Maps to explore the temporal aspects. Zhang et al. [12] employ coordinated views that provide, for a set of target objects selected by the user, a timeline depicting the frequency of occurrence of these objects, an object recognition view and a frame representation to contextualize these objects in the video. For representing trajectories, Trajectolizer [13] draws trajectory group patterns over the scene background in order to provide group trajectory dynamics over time. Meghdadi and Irani [8] propose a layout to visualize trajectories using a single action shot image that combines multiple frames, also plotting the trajectories in a space-time cube that displays an overall timeline view of all the movements.

Although the presented approaches suit their purposes, most of them, even considering time information, focus mostly on spatial aspects of the objects behavior, as well as the movement patterns through the scene. We instead propose a visual strategy that focus on the temporal dynamics of the objects behavior, specially on the relationship among them during their presence in scene.

III. System Description

Our system performs a post-event surveillance analysis, using previously generated videos, and employs the result of object identification/tracking methods to provide the analysis of the identified objects behavior during the video extent. Addressing the quality of these methods is beyond the scope of this work, and we consider the use of proper methods for such tasks. Based on the nature of this analysis and on existent works from the literature, we outlined a set of requirements that our proposed system must fulfil:
R1: provide an effective view of the surveillance video within a reasonable period of time that is affordable to the users;
R2: provide the analysis of the dynamics related to the objects presence in scene;
R3: provide a detailed view of the meetings among objects;
R4: provide the analysis of the objects speed distribution when moving through scene;
R5: provide the select regions of interest in the scene for analysis.

Fig. 1 illustrates our proposed analysis workflow. We first employ an identification/tracking approach on a previously generated video, and from its outputs we generate the layouts. We also extract all video frames, which will be used for interaction purposes. The user then interacts with the produced layout, executing a variety of basic exploration tasks, such as timeline zoom/pan, objects selection, as well as more sophisticated interactions detailed in this section.

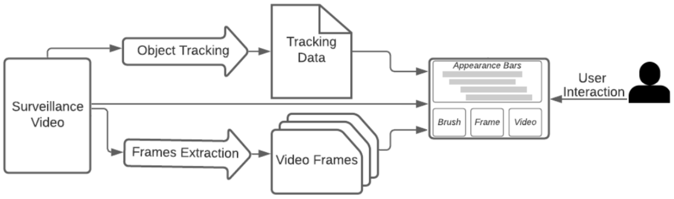


Fig. 1. Our surveillance video visual analysis system workflow.

The system interface is shown in Fig. 2, and provides the following coordinated views: *Appearance Bars View* (A), *Brush View* (B), *Frame View* (C) and *Video Player View* (D).

The *Appearance Bars View* (A) presents all the identified objects and their presence over the entire video extent, highlighting their behavior and the relationship with other objects. The horizontal axis, horizontally scaled to the screen width, depicts the video duration and the vertical axis lists all the identified objects. A bar or a set of bars is associated to each object, and the bar vertical borders indicate the instants in which the associated object entered/left the scene or is occluded. When hovering the appearance bars, users can check the object label, objects participating in a meeting at a specific instant, among other information. The *Brush View* (B) allows the user to select a region of interest from the scene background, concentrating the analysis in objects which crossed that region. In the *Frame View* (C), users can investigate a particular instant of the video. The bounding

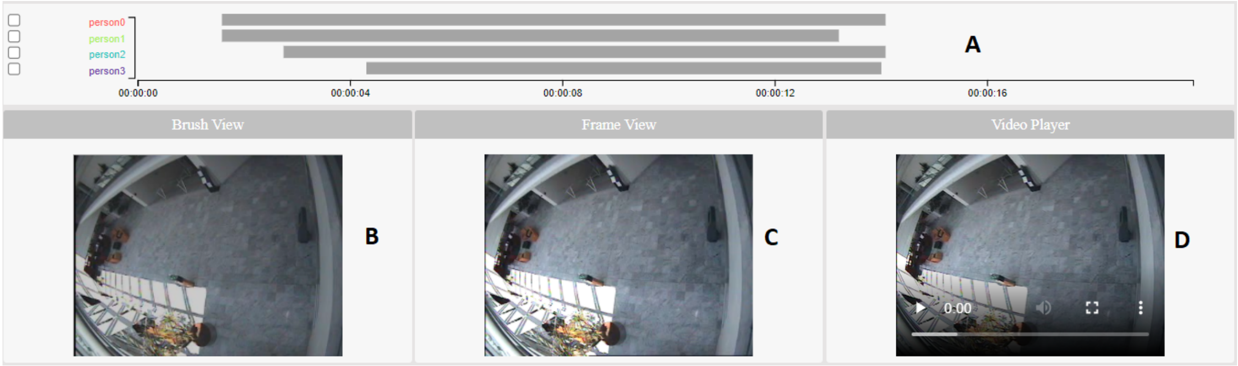


Fig. 2. Overview of the system interface. *Appearance Bars View* (A) present all identified objects and details about their presence in scene, focusing in all moments in which they perform actions and in which they interact with other objects. *Brush View* (B) allows to select a region of the scene background for concentrate the analysis in objects which crossed that region. *Frame View* (C) shows selected video frames with all objects bounding boxes highlighted, allowing to investigate their actions in particular instants. *Video Player View* (D) contains a traditional video player that allows users to watch the video.

boxes of the identified objects are highlighted, allowing for a deeper investigation of specific actions involving them. The *Video Player View* (D) implements a traditional interaction tool used for watch and navigate in videos, allowing the user to play/pause the video execution, advance/rewind frames and set the video in full screen or in picture-to-picture.

In order to enhance the exploration of the proposed layout and to highlight strategic behavior patterns, the system provides a set of interaction tools, described as follows.

A. *Meetings*

Meeting among objects represent an important aspect to comprehend the relationship among them during their presence in the scene. In our proposal, a meeting between two objects is defined as the occurrence of an intersection between their correspondent bounding boxes for a minimum consecutive time interval. An intersection is detected when at least one pixel of both bounding boxes coincide at the same frame.

In order to distinguish between significant meetings and objects just passing by each other, the system allows the user to configure a minimum meeting time parameter. The idea of this parameter is allow users to configure the adequate perspective of the meetings, enhancing the analysis process.

The system provides two meeting analysis tools. The *Show All Meetings* tool provides a general view of all objects meetings. All the meetings moments of each object are highlighted in a red layer over their appearance bars, in the portions which represent these moments, as shown in Fig. 5. By hovering an object appearance bar, one can see, in that instant, the number of objects it met and a list containing all these objects. Users may also see which objects met a specific object in a specific instant, by clicking in the object appearance bar. These meetings are then highlighted by black markers on the appearance bars of the objects which participated in this meeting. The *Frame View* shows the correspondent instant frame, and the identified objects bounding boxes are highlighted to allow their identification. The resulting layout is showed in Fig. 6a.

The *Show Meetings* tool allows viewing all meetings considering one or multiple objects selected in the *Appearance Bars View*. When a single object is selected, all the moments in which it participated in a meeting are mapped to layers over its appearance bar, in portions representing these moments. The layers color intensity are proportional to the number of object it met. The moments in which the other objects met the selected object are mapped to a dark gray layer over their appearance bars, in the portions which represent these moments. Yellow marks are used to highlight the instants in which the meetings changed somehow, either by removing or adding new objects. The resulting layout is shown in Fig. 7.

B. *Sortings*

The default appearance bars ordering is the first appearance time. However, the system allows users to sort the appearance bars, in descending order and in a top-down, according to a variety of aspects, as follows:

- *Scene Permanence:* The frames in which an object is detected are counted, and the objects are sorted according to these counts;
- *Number of Meetings:* All the objects meetings are calculated and counted. Meeting changes, in terms of addition/removal of new participant objects are considered as new meetings. The objects are then sorted according to these counts;
- *Distinct Objects Meetings:* All the objects non-recurring meetings are calculated and counted, and the objects are then sorted according to these counts;
- *Time in Meetings:* The frames in which an object participates in meetings are counted, and the objects are sorted according to these counts.

C. *Speed*

The system provides the analysis of the objects movement speed variation during their presence in the scene. Users may define time intervals (in seconds) for which an average speed is calculated, providing an analysis in distinct time resolutions.

The resulting average speed for each time interval is then mapped to the correspondent time portion of the appearance bar, whose color intensity is proportional to the average speed value, as shown in Fig. 3.

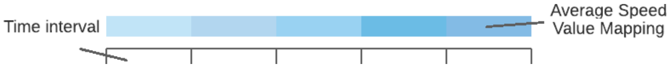


Fig. 3. Example of average speed layer, mapping each average speed value to a color intensity in the bar.

The average speed in each time interval is calculated considering the Euclidean distance (in pixels) between the bounding boxes centers positions in consecutive frames of the interval. Considering an object k, whose bounding box center is C_k, a user defined time interval as t, the average speed S in the interval t is calculated as follows:

$$S = \frac{\sum_{i=f_1}^{f_{|F|-1}} d_{(i,i+1)}}{t}, \tag{1}$$

$F = \{f_1, f_2, ..., f_{|F|}\}$: frames set in t.
$d_{(i,j)}$: Euclidean distance between C_k in frames i and j.

The result value is a pixel/sec. measure. Although the resulting value presents no correspondence with employed real world speed units, the idea here is only to visually highlight the speed variation of an object, as well as to compare speeds from different objects. If the object presence last time interval is shorter than the defined time interval, the average speed is calculated considering only this remaining time interval. An example of resulting layout is shown in Fig. 9.

D. *Scene Region Filtering*

The system allows users to select a rectangular region of interest in the video background and concentrate the analysis in the objects which crossed this region only. The selection is performed in the *Brush View*, and the results are shown in the *Appearance Bars View*. The moments in which the filtered objects crossed the selected region are mapped to a dark gray layer in the appearance bars, in portions representing these moments. The user is then able to identify which objects in which moments crossed a specific region captured by the surveillance camera. The resulting layout is shown in Fig. 10.

IV. Experimental Results

In this section we present the results of applying our proposed visual system to two surveillance scenarios. We first explore the general view of the layouts to investigate how the identified objects behave during their presence in scene. We then refine the analysis exploring the relationships between the objects and temporal/spatial aspects of objects movements/actions. We also analyze how previously known events are shown in layout, as well as how the layout represents different event categories. Finally, we identify which of the requirements presented in Sec. III our proposal fulfill.

The objects labels shown in the layout are defined by the detector, and do not necessarily represent what the object really is, thus we always refer them as objects. However, it is important to highlight that an accurate object type detector can enhance the layout capabilities, which becomes even more intuitive/informative, as it allows the expert to make more accurate inferences about the relationship between the objects in the scene, as well as their behavior during the video.

A. ***MeetCrowd***

The Context Aware Vision using Image-based Active Recognition (CAVIAR[1]) repository consists of video clips representing several surveillance scenarios, from which we selected the *Meet_Crowd.mpg* video, named here as **MeetCrowd**. This video was recorded in the entrance lobby of the INRIA Labs at Grenoble, France, and is composed of 497 frames distributed in 19 secs. By watching this video, we manually identified its main events, which are described in Table I.

TABLE I
MEETCROWD MAIN EVENTS DESCRIPTION.

Event	Frames	Description
1	0-39	The lobby is empty.
2	40-68	Two people enter the scene.
3	69-110	Two other people enter the scene.
4	111-329	Four people cross the lobby together.
5	330-352	Four people leave the scene.
6	353-497	The lobby is empty again.

The layout produced from **MeetCrowd** is shown in Fig. 4, and was generated in a short period of time, 2 millisec. **(R1)**. The *Appearance Bars View* displays 4 identified objects, and their appearance bars quickly allows the identification of when they entered/left the scene (begin/end of each bar). The appearance bars do not fill the entire timeline, as there were no identified objects in the scene in the beginning/end of the video (Events 1 and 6). It is also possible to notice the order in which people entered and left the scene and the portion time in which all of them are simultaneously in the scene (between 4s and 13s). ***Person0*** and ***person1*** were the first objects to enter the scene, and they entered at the same time (Event 2). The last objects to leave the scene were ***person0*** and ***person2***, practically at the same time. All the objects were simultaneously on the scene in a certain moment and their presence time are relatively similar. The layout is able to highlight moments in which all activities occurred, allowing the surveillance agent to quickly identify which time portions are interesting for analysis **(R2)**.

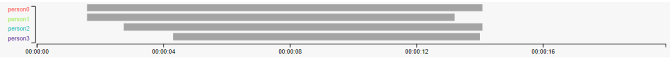


Fig. 4. **MeetCrowd** *Appearance Bars View* layout.

Fig. 5 shows the result of highlighting all meetings in *Appearance Bars View*. One notices that all objects met another

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

object in at least one specific moment. It is also possible to notice that all of them entered the scene meeting at least one object, suggesting that multiple objects entered the scene together or an object entered the scene and immediately met another objects. It is also possible to notice that ***person0*** and ***person1*** were the last objects participating in a meeting. When ***person2*** entered the scene, the two objects already in the scene were participating in meetings, and it is not possible to identify which of them it met. However, as we know these objects were together, we can conclude that ***person0***, ***person1*** and ***person2*** participated in this meeting. When ***person3*** entered the scene however, although one notices that it met one/some objects, nothing can be inferred about which ones it exactly met, because at this time multiple separate meetings involving all the objects in the scene were in progress. The layout also shows that ***person0*** and ***person1*** spent most of their scene time participating in meetings, while ***person3*** was the one with less meeting moments. By highlighting when objects meet over time **(R3)**, the layout offers a quick guidance to the surveillance agent about in which objects, in which moments, he/she must concentrate the analysis.

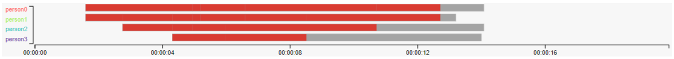
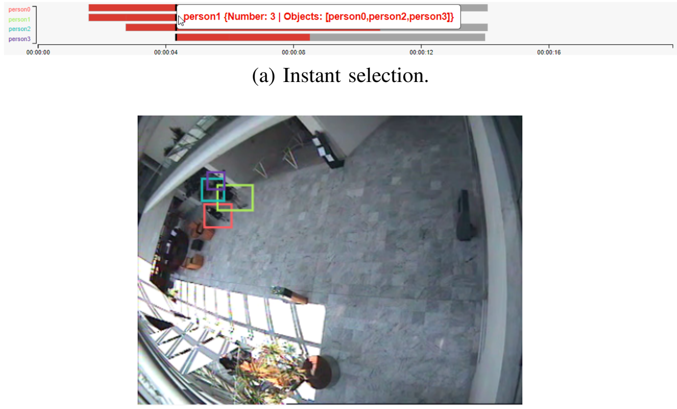


Fig. 5. Meetings in **MeetCrowd** highlighted in the *Appearance Bars View*.

Fig. 6a highlights the instant when ***person1*** met with all objects on scene. The black mark at the beginning of ***person3*** bar indicates that ***person1*** met it as soon as it entered the scene. The frame corresponding to this instant is shown in Fig. 6b, showing ***person3*** bounding box (purple) intersecting with the ***person1*** and ***person2*** ones (green and blue, respectively), and it is possible to notice that ***person1*** bounding box intersects with all others in the scene, which suggests that all objects were meeting at this moment.



(a) Instant selection.

(b) Correspondent frame.

Fig. 6. Selection in *Appearance Bars View* of the instant in which all the objects met ***person1***, highlighted by a black mark. The correspondent frame, with objects bounding boxes highlighted, shows that ***person1*** started this meeting as soon as he entered the scene.

When filtering ***person1*** meetings in *Appearance Bars View*, it is possible to notice all the moments in which ***person1*** participated in a meeting, as well as all the moments when the other objects met ***person1***, as shown in Fig. 7. The instant in which ***person1*** met the other three objects is highlighted by the highest red intensity in its bar. The *Appearance Bars View* shows that ***person1***, together with ***person0***, met ***person2*** and later ***person3***. Although one notices, by watching the video, that ***person0*** and ***person1*** walked together during all their video presence, the *Appearance Bars View* shows some portions in the bar in which there is no meeting between these objects. These "gaps" can be produced by the method used for object tracking or even when the objects distance slightly increases for a moment, which impacts the bounding boxes generation and consequently the meeting detection.

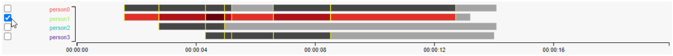


Fig. 7. ***Person1*** meetings in *Appearance Bars View*. The time portions corresponding to each meeting are highlighted, as well as their distribution over time (red intensities). ***Person1*** meets all the objects and in a specific moment it meets all of them simultaneously (highest color intensity in ***person1*** bar). The bars gray portions indicate the other participants of this meeting.

Although it is possible to notice, by watching the video, that all four people walked together during most of the time, Fig. 8 shows that no simultaneous meeting between the four objects was observed. The reason is that there were no moments in which all bounding boxes simultaneously intersected with each other. However, each bounding box intersected at most two other bounding boxes simultaneously, and the meeting involving all the objects is thus indirectly indicated.

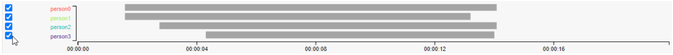


Fig. 8. Selection of all identified objects. The absence of red portions in the bars indicate that no simultaneous meetings occur among the four objects (the corresponding bounding boxes do not simultaneously intersect to each other).

Fig. 9 shows the objects average speed when moving in the scene, considering 1 sec. intervals. The blue shade variation in the bars segments suggests an acceleration in the objects speed in most of their presence in scene, and a small deceleration before they leave. In general, no sudden speed variation is observed, and the speed increasing is roughly homogeneous, suggesting a group walking pattern (Event 4). Only ***person3*** presents a movement pattern that diverges from the others, specially after 10 sec. When watching the video, one notices that ***person3*** went to the same direction of the other people to leave the scene, but as he/she was a little far from the group, he/she made a bigger curve, then increasing its speed to come closer to the group again. The layout highlights movement behavior patterns based on objects speed distribution **(R4)** allowing to identify actions such as people and vehicles running, sudden stops, parked vehicles or people stopped for a long time, among other actions.

Fig. 10 shows the selection of two regions in the *Brush View* and the respective *Appearance Bars View* results. The first selection (Figs. 10a and 10c) highlights that only ***person3***

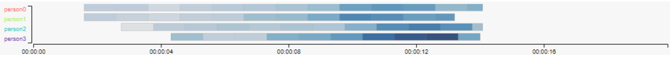
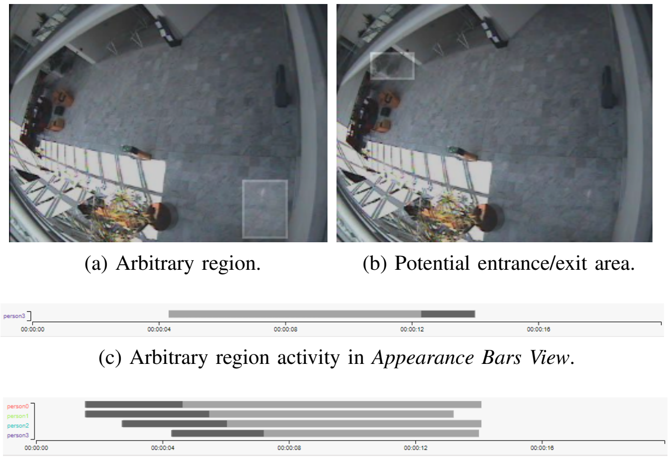


Fig. 9. Objects average speed in **MeetCrowd**, showing how ***person3*** speed increases in the last portions of its bar.

passed through the region. The bar portion in dark gray, which coincides with the last portion of ***person3*** bar, suggests that it was inside this region alone when leaving the scene. The second selection (Figs. 10b and 10d) shows that all objects in the video occupied this space at their trajectories beginnings, suggesting that they entered the scene by crossing this region (Events 2 and 3). The brush can be used to monitor static objects such as store cashs, ATMs and safe boxes, as well as places where permanence is forbidden, risk areas and entrances/exits. The layout quickly highlights identified objects crossing a scene region **(R5)**, no matter how fast they are, which allows users to notice important quick events that could be missed just by watching the video.



(a) Arbitrary region.

(b) Potential entrance/exit area.

(c) Arbitrary region activity in *Appearance Bars View*.

(d) Potential entrance/exit area activity in *Appearance Bars View*.

Fig. 10. Selection of regions of interest in **MeetCrowd**. One notices that only ***person3*** crossed the selected region shown in Fig. 10a, while the one shown in Fig. 10b was used as an entrance place by all of the them.

B. *ParkingLot*

VIRAT[2] [14] is surveillance videos repository representing interesting scenarios for analysis. An annotation file is provided with each video, depicting the bounding boxes of a set of identified objects, for each frame. From this repository we selected the *VIRAT_S_000002.mp4* video, named here as **ParkingLot**. This video was recorded in a parking lot in USA, and is composed of 9075 frames distributed in 5 mins. and 2 secs. In order to improve the annotations and reflect the employment of a highly accurate object detector, we manually annotated two additional objects (objects 7 and 9), and adjusted some inaccurate identifications. By watching this video, we manually identified its main events, which are described in Table II.

TABLE II
PARKINGLOT MAIN EVENTS DESCRIPTION.

Event	Frames	Description
1	0-2457	A group of three people (**group1**) walks through the parking lot and stops near a facility.
2	2067-4735	A car enters the scene and parks close to group1.
3	2397-9074	Another group of two people (**group2**) enters the parking lot by the upper part of the scene.
4	2457-3265	A person from group1 gesture to the group2.
5	2517-3655	The driver gets out the car and walks around it.
6	3266-9075	A person leaves group1 and join group2. Both groups walk to different positions and stop at the bottom of the scene.
7	3386-4885	A person with a hand truck dolly enters the parking lot by the upper right part of the scene, walks to the car, get a box from the car trunk, and leaves the parking lot by the upper right part of the scene.
8	4345-5394	The driver enters the car and leaves the parking lot by the upper part of the scene.

Fig. 11 shows the **ParkingLot** *Appearance Bars View*, which was generated in a short period of time, 2 millisec. **(R1)**. It presents 10 identified objects, as well as when each of them entered/left the scene. One notices that there are no bars of objects labeled as "car" occupying the entire timeline, suggesting that the parking spaces were empty for at least one moment during the video extent. The bars of objects labeled as "People" occupy the scene in most of the time, and 3 of them occupy the entire timeline, suggesting that at least three people were always in the parking lot. ***Person0***, ***person1*** and ***person2*** were in the scene during all the video extent, but ***person5*** and ***person6*** entered the scene after 1 min. and 28 sec., remaining until the end of the video.

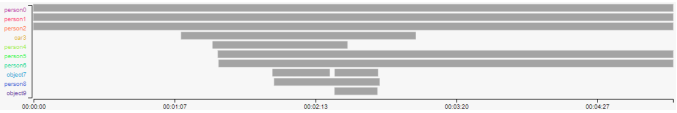


Fig. 11. **ParkingLot** *Appearance Bars View* layout.

Fig. 12 shows the result of highlighting all meetings in *Appearance Bars View*. One notices that all objects met another object at least once. When the video starts, ***person0***, ***person1*** and ***person3*** were already meeting. As they were the only objects in the scene, one can conclude they were meeting each other. ***Person6*** and ***object9*** met other objects during all their presence in scene, while ***person2***, ***person5***, ***object7*** and ***person8*** met someone during most of their presence time. According to the *Time in Meetings* sorting (Fig. 13), ***person2***, ***person6*** and ***person5*** were the objects that spent more time in meetings.

Fig. 14 highlights ***person2*** meetings in the *Appearance Bars View*. One can notice that ***person2*** starts the video meeting ***person0*** and ***person1***, and then it meets ***person5***, remaining together until the end of the video. It also meets ***person6*** during a short period of time, while in meeting with ***person5***.

Fig. 15 shows **ParkingLot** frames illustrating some ***person2*** related actions. One notices that it participated first in group1

[2] https://viratdata.org/

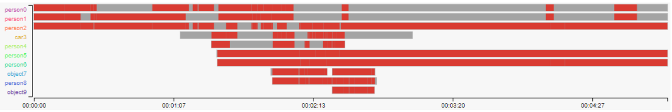


Fig. 12. Meetings in **ParkingLot** highlighted in the *Appearance Bars View*.

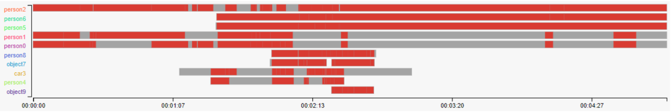


Fig. 13. Objects in **ParkingLot** ordered according to the time they spent in meetings. ***Person2***, ***person6*** and ***person5*** were the ones with highest meeting times.

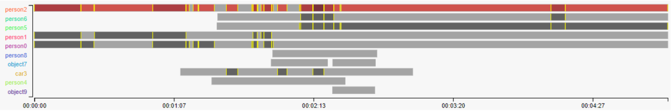
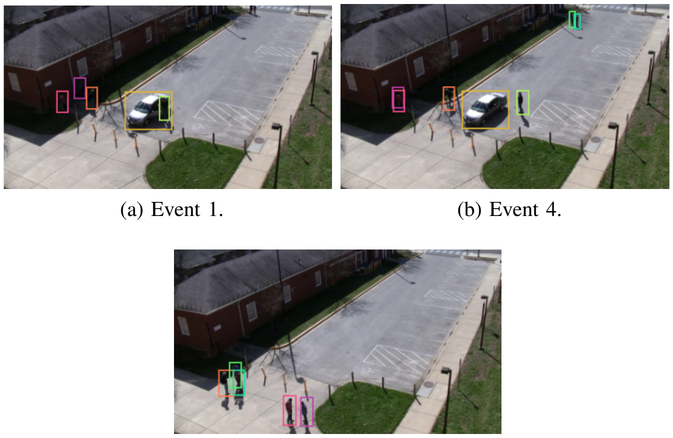


Fig. 14. ***Person2*** meetings in *Appearance Bars View*. ***Person2*** first meets ***person0*** and ***person1***, and then meets two other objects (***person5*** and ***person6***).

(Fig. 15a, orange bounding box - Event 1), moved away to gesture to group2 (Fig. 15b - Event 4), and then participated in group2 (Fig. 15c - Event 6). The layout illustrates these events segmenting ***person2*** appearance bar into one initial long segment, a set of small segments in the middle, and another long segment at the end of its appearance bar. ***Person2*** also met ***car3*** three times, twice alone and once with ***person5*** and ***person6***. In the first meeting, he/she gestured to group2 positioned next to the car (Fig. 15b). He/she then walked alone to meet group2, and walked with ***person5*** and ***person6*** to the bottom left part of the scene (Fig. 15c). The meetings between ***person2*** and ***car3*** are captured considering that the bounding boxes must intersect for at least 1 sec. to be considered as a meeting.


(a) Event 1.

(b) Event 4.

(c) Event 6.

Fig. 15. Frames illustrating events related to ***person2*** (orange bounding box).

Fig. 16 highlights ***object7*** meetings. It appears in the scene already meeting ***person8***, and this meeting lasts practically all their presence in scene. By watching the video, its possible to notice that ***object7*** was a hand truck dolly carried by ***person8***, which, together with ***person4***, occluded it for a few seconds, producing a gap in the appearance bar. ***Object9*** met ***object7*** during all its presence in scene, and this meeting started after ***object7*** stopped a meeting with ***car3*** and ***person4***. By watching the video, one notices that ***object9*** was a box collected by ***person8*** from ***car3***. ***Person8*** then walked away from ***car3*** and ***person4*** carrying this box in ***object7*** (Event 7).

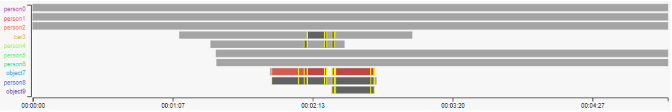


Fig. 16. ***Object7*** meetings in *Appearance Bars View*. ***Person7*** meets ***person8*** during almost all its presence in scene, indicating they were together. As ***object9*** also meets ***object7*** during all its presence in scene, it suggests that ***person8*** was with these two objects.

The layout shown in Fig. 17 highlights multiple interactions of several objects with a single object (***car3***), in which the *Distinct Objects Meetings* sorting is applied. One notices that ***car3*** is the object that meets more distinct objects. The sorting tools provide ways to facilitate the analysis of strategic situations such as a vehicle theft or suspicious meetings among people. ***Person4*** starts and finishes its presence in scene meeting ***car3***, suggesting that he/she was inside the car, left it (Event 5) and then entered again (Event 8). There is a moment in which ***car3*** meets most of the objects simultaneously (***person2***, ***object7***, ***person8***, ***person5*** and ***person6***), which is highlighted in the layout by the darkest red portion of ***car3*** appearance bar.

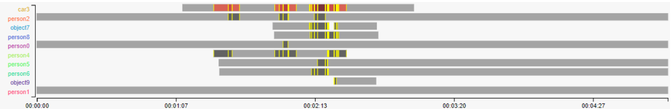
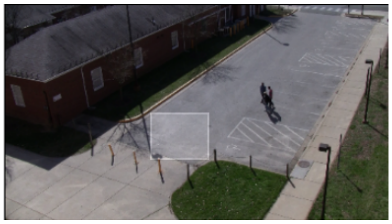


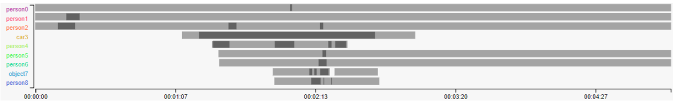
Fig. 17. ***Car3*** meetings in **ParkingLot** *Appearance Bars View*. ***Car3*** is the object that meets more distinct objects in the video and ***person4*** meets ***car3*** several times, including at the beginning and at the end of its bar, suggesting a sequence of actions that represent Events 5 and 8.

Fig. 18a shows the selection of a region in the *Brush View*. The result in the *Appearance Bars View* is shown in Fig. 18b. One notices that almost all the identified objects crossed the selected region at some moment. It is also possible to notice that ***car3*** was positioned in the selected region in most of its presence in scene, suggesting that it was parked there (Event 2). ***Person4*** occupies the region at the beginning/end of its bar, indicating it entered/left the scene by this region.

Fig. 19 shows the objects average speed when moving in the scene, considering 1 sec. intervals. The blue shade variation in the bars segments suggests a small acceleration/deceleration of the objects speed in most of their presence in the video. ***Car3***



(a) Arbitrary region.



(b) Arbitrary region activity in *Appearance Bars View*

Fig. 18. Selection of an arbitrary region in **ParkingLot**. Almost all the identified objects crossed the selected region, and ***car3*** was positioned in the selected region in most of its presence in scene, suggesting that it was parked there (Event 2).

however shows two acceleration peaks, which, by watching the video, represent the moments in which the car entered/left the parking lot. Both ***person8*** and ***object7*** show a movement pattern similar to ***car3***, illustrating a general pattern which represents objects that enter the scene, move to a specific position, stop for a moment and then move again to leave the scene. By watching the video, one notices that ***car3*** and ***person8*** moved toward a scene region, stopped for a while and then left the scene. Moreover, ***person0***, ***person1*** and ***person2*** presents an homogeneous speed decreasing at the beginning of their bars, suggesting a group movement pattern (Event 1). These pattern is confirmed in the video, when ***person0***, ***person1*** and ***person2*** walked together and then stopped next to a facility (Fig. 15a).

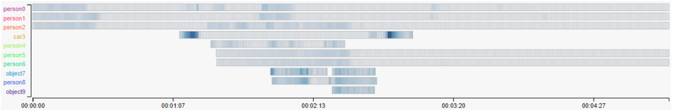


Fig. 19. Objects average speed in **ParkingLot**. The objects present a small acceleration/deceleration in most of their presence in scene, except for ***car3***, which shows an acceleration peak.

V. Conclusion

This paper presented a surveillance video visual analysis system with focus on the exploration of objects behavior in surveillance scenarios. Our system employs three coordinated layouts which highlight the dynamics and distribution of each object presence in scene. A set of interaction tools provided means to explore several objects behavior aspects, specially interactions among them. Our case studies demonstrated the ability of the system in highlighting several objects actions, and in distinguishing between different types of interaction among them, allowing to relate these actions to the occurrence of relevant events.

To faithfully represent the objects behavior, our approach depends on the accuracy of the chosen detection/tracking technique. However, multiple manual/automatic strategies can be combined to improve this accuracy. Moreover, although the layout generation process is not computationally costly, a large number of objects/events results in a large number of bars, and patterns representing small actions may be omitted, requiring the use of time and space filters. We intend to investigate strategies to improve the scalability of our strategy for these scenarios. Finally, the objects distance from the camera impacts on the meetings detection, as objects distant from each other may have intersecting bounding boxes depending on the perspective, and on the speed calculation, as distant objects may present lower speed variation than closer ones. We intend to improve these procedures to take this distance into account.

Future work include a user study with surveillance experts to evaluate the system ability to communicate the object behavior in the video and the adaptation to real time surveillance scenarios, expanding its application scenarios.

Acknowledgment

The authors are grateful to CAPES and CNPq for the support in the execution of this project.

References

[1] H. Kruegle, *CCTV Surveillance: Video practices and technology*. Elsevier, 2011.

[2] K. A. Joshi and D. G. Thakore, "A survey on moving object detection and tracking in video surveillance system," *Int. J. of Soft Comput. and Eng.*, vol. 2, no. 3, pp. 44–48, 2012.

[3] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior, and Y. Tian, "Smart surveillance: applications, technologies and implications," in *Proc. Joint 4th Int. Conf. on Inf., Commun. and Signal Process. and 4th Pacific Rim Conf. on Multim.*, vol. 2. IEEE, 2003, pp. 1133–1138.

[4] M. Elhoseny, "Multi-object detection and tracking (modt) machine learning model for real-time video surveillance systems," *Circuits, Syst., and Signal Process.*, vol. 39, no. 2, pp. 611–630, 2020.

[5] G. Sreenu and M. S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *J. Big Data*, vol. 6, no. 1, pp. 1–27, 2019.

[6] A. B. Mabrouk and E. Zagrouba, "Abnormal behavior recognition for intelligent video surveillance systems: A review," *Expert Syst. with Appl.*, vol. 91, pp. 480–491, 2018.

[7] W. Liang and H. W. M. T. Niu, "A survey of visual analysis of human motion [j]," *Chinese J. of Comput.*, vol. 3, pp. 225–237, 2002.

[8] A. H. Meghdadi and P. Irani, "Interactive exploration of surveillance video through action shot summarization and trajectory visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2119–2128, 2013.

[9] G. Mendes, J. G. S. Paiva, and W. R. Schwartz, "Point-placement techniques and temporal self-similarity maps for visual analysis of surveillance videos," in *2019 23rd Int. Conf. Inf. Vis. (IV)*. IEEE, 2019, pp. 127–132.

[10] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Front. in Robot. and AI*, vol. 2, p. 28, 2015.

[11] S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, "Trajectory-based surveillance analysis: A survey," *IEEE Trans. Circuits and Syst. for Video Technol.*, vol. 29, no. 7, pp. 1985–1997, 2018.

[12] Z. Zhang, R. Zuo, R. Guo, Y. Li, T. Zhou, H. Xue, C. Ma, and H. Wang, "Multi-scale visualization based on sketch interaction for massive surveillance video data," *Pers. and Ubiq. Comput.*, pp. 1–11, 2019.

[13] A. Sawas, A. Abuolaim, M. Afifi, and M. Papagelis, "Trajectolizer: Interactive analysis and exploration of trajectory group dynamics," in *2018 19th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2018, pp. 286–287.

[14] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. T. Lee, S. Mukherjee, J. Aggarwal, H. Lee, L. Davis *et al.*, "A large-scale benchmark dataset for event recognition in surveillance video," in *CVPR 2011*. IEEE, 2011, pp. 3153–3160.