Multi-Object Tracking in Dynamic Scenes By Integrating Statistical and Cognitive Approaches

Saira Saleem Pathan, Omer Rashid, Ayoub Al-Hamadi, and Bernd Michaelis
Institute for Electronics, Signal Processing and Communications (IESK)
Otto-von-Guericke-University Magdeburg, Germany

Abstract: **In this paper, we have addressed a quite researched problem in vision for tracking objects in realistic scenarios containing complex situations. Our framework comprises of four phases: object detection and feature extraction, tracking event detection, integrated statistical and cognitive modules, and object tracker. The objects are detected using fused background subtraction approach along with feature computation. Next, the tracking events are inferred by finding spatial occupancy of moving objects. Third module is the key to proposed approach and the motivation is to tackle the tracking problem by axiomatizing and reasoning human-tracking abilities with associated weights. Each object contains a unique identity and a data structure of cognitive and statistical attributes whilst satisfying the global constraints of continuity during motion. Consequently, the results are linked with Kalman filter based tracker to estimate the trajectories of moving objects. We show that combining cognitive and statistical information gives a straightforward way to interpret and disambiguate the uncertainties occurred due to conflicted situations in tracking. The performance of the proposed approach is demonstrated on a set of videos representing various challenges. Besides, quantitative evaluation with annotated ground truth is also presented.**

Keywords: Object tracking, Statistical modeling, Cognitive processing, Kalman filter, Applications.

I. Introduction

Behavior understanding is one of the important research domains of computer vision where tracking is a primary element. Modeling the detected trajectories of objects along with their cognitive behaviors over longer intervals of time reveal paths to a diverse range of applications such as: surveillance system, traffic monitoring system, human-computer interaction, etc. Practically, tracking is a difficult problem due to the direct and indirect influences of real-time complications in the scene. For example, direct interferences refer to partial and full occlusion, changing object proximity, complex motion, object fragmentation, and variation in orientation whereas indirect interferences include scene illumination changes, camera motion and on-field obstacles.

Fig. 1 shows the example of direct interferences during object movement in the scene where the occlusion among objects is observed very frequently. There are various types of occlusions such as: 1) object-to-object occlusion and 2) object-to-scene occlusion. In this research, we focus on object-to-object occlusion problem during tracking.

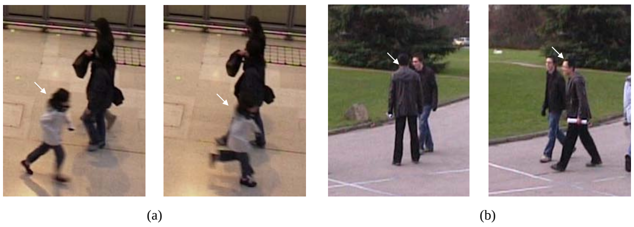


Figure. 1: shows the various examples of direct interference (pointed by arrows). a) indicates the motion variation of object and full occlusion among objects; b) indicates situation where orientation of the object is changed significantly.

In this paper, we propose a novel algorithm for real-time object tracking and behavior understanding by employing the concepts of human cognitive behaviors with statistical measures for vision system. We argue that the domain specific cognitive models (i.e. conceptualization of human perception abilities) with statistical techniques can provide a subtle way for tracking in the conflicted situation. The contributions of the proposed research can be revealed in three ways. First, the uncertainties of data association are disambiguated which are observed during occlusion using cognitive explanation and inference of objects behavior in a world domain, second the motion-behavior of objects during the entire tracking is interpreted and third, the capabilities of Kalman filter based tracker is extended so that it can behave normally under non-linear situations.

The paper is organized as follows: section II discusses and reviews the relevant literature. Section III describes the proposed approach whereas section IV and section V cover the object detection and tracking event detection, section VI explains the statistical and cognitive model of the proposed approach and section VII presents the Kalman filter based tracker. Section VIII shows the experimental results. Finally, section IX sketches the concluding remarks and future directions of the work.

II. Related Work

The tracking algorithms usually follow a modular scheme which either perform Detection prior to Tracking (DpT) or Tracking prior to Detection (TpD) in a flexible architecture. In the DpT approach, objects of interest are first detected at every instance of time and tracking of the objects is performed, only. In contrast, in TpD approach, a hypothesis is built about the object location in the generated state space which is then evaluated by the computed set of features in an image. Moreover, in recent years, object identity management is also gaining attention which incorporates the concepts of data association through similarity measurements. The objective is to assign unique identities to each object and to manage these identities over time. Later, elementary trackers (i.e., Kalman filter or Mean shift filter) are used to track these objects.

In DpT approaches, a wide range of literature has been published to handle the fundamental limitations of data association approaches [1]. For instance, Cox and Hingorani [2] have presented an efficient variant of MHT approach in which the k-best hypotheses are determined in polynomial time using Murty's approach. With the similar motivation, Isard and MacCormick [3] proposed Bayesian multiple block tracking system. In their approach, a multi-blob likelihood function assigns the direct comparable likelihoods to hypotheses containing different number of objects. Similarly, Smith et al. [4] proposed a Bayesian framework for the fully automatic tracking of a variable number of interacting targets. They have employed a joint multi-object state-space formulation and a trans-dimensional Markov Chain Monte Carlo particle filter to recursively estimate the multi-object configuration and efficiently search the state-space. However, the actual blobs may contain multiple categories of objects, such as shadows, reflection regions, and blobs due to camera motion parallax. More recently, Ryoo and Aggarwal [5] presented a paradigm for tracking objects under severe occlusion named as observe-and-explain. The system chooses the hypothesis path with the highest probability which enables the tracking of even fully occluded objects with the cost of higher computational effort.

A different way to address the object tracking issues is through object identity recognition. In the literature, not much work has been reported for object identity recognition in which a specific individual detected at certain time instance is matched with the previous observations. A framework is presented by Guo et al. [6] for vehicle matching in aerial views but their main focus is blob extraction and alignment rather than the recognition. Gheissari et al. [7] presented a two layer method for human identification. In the first layer, a graph based spatio-temporal segmentation is applied to group the object pixels that belong to the similar cloth. The second layer used the decomposable triangulated graphs to segment and link different parts of the human body. Even though, human recognition is not the direct focus of the literature, but some seminal advances in human detection have been reported that can be indirectly associated. For instance, Dalal and Triggs [8] trained SVM classifier using features of Histograms of Oriented Gradients for human detection and localization. However, these methods are highly dependent on image details for extracting the features, such as faces or body parts, and therefore, can only be applied to high-quality ground images. Both, tracking and object identity recognition are closely related problems, since solving the tracking implicitly accomplishes the task of identity recognition. Similarly solving identification over consecutive frames is actually one of the fundamental tasks of object tracking, therefore based on this concept, we have developed our framework.

On contrary, both the propositional and imaginistic components are essential for reasoning the spatial concept along with the coordinated use of heterogeneous representation and inferential process [9]. In essence, it is empowered by the conceptualization of human perception and inference ways. Therefore, it guides in interpretation of the ambiguous discrete data adequately. A view-based approach is proposed by Sherrah and Gong [10] in a highly constrained environment with Bayesian framework and explicit probabilistic reasoning to handle the plausible interpretation of incomplete data due to occlusion. Similarly, Bennett et al. [11] proposed a technique by fusing the logical reasoning explicitly in tracking module. However, in their work, uncertainties due to occlusion are disambiguated after classification through long-term reasoning unlike our proposed work which is online and more focused to exploit the logics with likelihood as suggested by Halpern et al. [12].

Discussion: The research in computer vision directly addresses the issues in real scenarios which are not anticipated earlier by the research community. The ample of research in vision underpinned the statistical techniques but it is evident that the statistical approaches have their limitations (e.g. exponential increment in search space or false hypothesis generation during occlusion). Particularly, prediction of coherent description of statistical data without considering domain specific logical constraints can manifest errors. Our analysis at this stage is more biased towards the cognitive approaches considering efficacious performance even if the discrete data is incomplete and ambiguous. Thus, combining these two approaches complement each other and this is the motivation behind the proposed approach.

III. Proposed Framework

The proposed framework has four main modules: i) object detection and feature extraction, ii) tracking event detection, iii) quantitative and qualitative, and iv) tracking system. Fig. 2 illustrates the relationship between each phase.

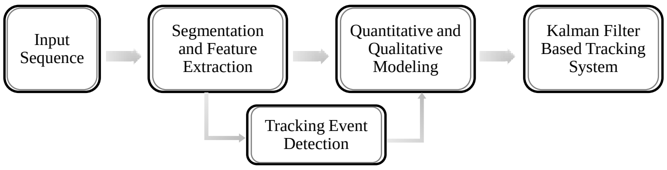

Input Sequence
Segmentation and Feature Extraction
Quantitative and Qualitative Modeling
Kalman Filter Based Tracking System
Tracking Event Detection

Figure. 2: The proposed framework.

Object Detection and Feature Extraction: First, the object detection is performed by employing the integrated background subtraction approach, and the visual features are computed. The detected objects and corresponding features at each time instance.

Tracking Event Detection: There are many events observed during tracking process, such as occlusion, split, new entry, and exit. In this module, we have detected these events, and the respective logical functions are triggered based on these events.

Statistical & Cognitive Approaches: In this phase, we have proposed two algorithms (i.e. statistical and cognitive) along with their integration. At the conceptual level, the overall goal is to track and understand the individual and collective behaviors of the objects. In dynamic scenes, identification of objects using typical statistical matching algorithms normally give very poor results. For example, occlusion is observed very frequently in which objects overlap each other partially or completely while moving across the scene. It is evident that treating object tracking as the cognitive problem and use it with the typical statistical algorithm can improve the overall tracking performance under non-feasible situations. At the practical level, all the detected objects are considered as a node and constituted an undirected graph. Moreover, these objects are described by the unique identity (i.e. from Identity Pool) and a data structure which comprises of quantitative (i.e., visual characteristics) and qualitative (i.e. logical characteristics) information at each time instance as illustrated in Fig. 4. First, the matching weights of the objects are computed. Second, the axioms are developed by employing the principles of human-perception for the tracking process. These axioms assign the behavioral-states to the detected objects where the decision of object state is associated with the matching weights. Essentially, these two approaches function together and the state of an object is refined and updated by incorporating the reasoning functions satisfying the fundamental constraints of continuity of objects during tracking.

Tracking System: We have developed a Kalman filter-based tracking system where every tracker is associated with an object and estimates its state over time.

IV. Object Detection and Feature Extraction

The first task in tracking is to detect objects in given video sequence which serves as an input for further high level process. For this purpose, we have suggested integrated background subtraction segmentation approach to detect objects in the test sequences as shown in Fig. 6. The next objective is to compute features that depict the unique representation for each object. A number of practical situations are taken into account, and it is found that multiple features can influence substantially in performance of object matching. In the sequel, how to combine the features appropriately is the major concern of feature fusion approaches. So, we have exploited the two approaches of object's color characteristics in our proposed feature fusion approach in section VI. In the first, we have computed the ellipse around the detected object and build an elliptical histogram based on color of the object. In the second approach, we have exploited CSC approach [13] which segments object into color segments referred as color-patches. Besides, we have also taken into account object's geometrical features, such as area of the object and their bounding region. So, the object features set (f) is written as:

$$f = \{\epsilon_h, \zeta_p, area, bb\}, \tag{1}$$

$$\zeta_p = \left\{c_n^{k:id}; n = 1, .., N\right\}, \tag{2}$$

$$c_n^{k:id} = (c_{area}, c_{\overline{ncolor_{rgb}}}, c_{bb}), \tag{3}$$

where ϵ_h is the elliptical color histogram, $area$ defines the area, and bb presents the bounding box of the object. In the following, we have explained how these color features for the detected objects are computed. ζ_p shows the color-patches of object, $c_n^{k:id}$ is the color-patch with a set of attributes[1], such as c_{area} is the area, $c_{\overline{ncolor_{rgb}}}$ is the mean color, and c_{bb} is the bounding region of the color-patches.

V. Tracking Event Detection

We have categorized these events into four types:1) new, 2) exit, 3) occlusion, and 4) split. The detections of these events are based on mapping the spatial occupancy of objects at time k with the objects at time $k-1$. So, ideally each object should contain only one object which shares its spatial space in the next frame where it is assumed that the objects are moving smoothly. Based on this assumption, a "spatial mapping matrix" of the detected objects at k and $k-1$ is built. Later, the criteria for these events are defined which are deduced based on the spatial mapping matrix values.

A detected object is a segmented region of input image from a sequence and is defined as a 2D function $I(k) = i(x, y)$ where i indicates the intensity and (x, y) contains the spatial information. We consider the spatial information (x, y) of each detected region at k and $k-1$, for creating the spatial mapping matrix. The number of columns in the matrix is equal to number of the detected objects at k. In contrast, number of rows in the matrix is equal to number of detected objects at $k-1$. Now, the main concept is to map the spatial correspondence of objects where the objects that do not share spatial correspondence are excluded.

Let us assume that the detected objects at k are $I(k) = \{o_i; i = 1, \ldots, m\}$ and objects at $k-1$ are $I(k-1) = \{o_j; j = 1, \ldots, n\}$. The spatial correspondence is computed by measuring the spatial occupancy which is the ratio of the spatial region of an object at k mapped over the spatial region of an object at $k-1$. The area of the spatial region of object at k is represented by $I(k)_{s_a}$ and the area of spatial region of object at $k-1$ is presented by $I(k-1)_{s_a}$. The ratio between the mapped spatial regions is defined as the percentage of spatial mapping S_r which determines the relative spatial occupancy of an object at k and is computed as follows:

$$S_r = \left(\frac{I(k-1)_{s_a}}{I(k)_{s_a}}\right) \times 100; \tag{4}$$

Based on above spatial relationship quantity, we have developed a set of criterion for each of the tracking event. Moreover, we have demonstrated a test case in Fig. 3 to provide an insight about each of following criteria for detecting the respective events as given in Table. 2 after conducting empirical studies.

VI. Statistical & Cognitive Approaches

In this section, we have suggested the statistical and cognitive algorithms which are driven by object detection and feature extraction approaches. In statistical approach, Bayes inference is employed to measure the posteriori probability of the objects which is referred as likelihood of objects. In the cognitive model, an explicit novel qualitative approach is

[1] It is possible to measure other attributes, for example color-patch histogram or Eigen vectors. But currently, we are only taken the above attributes into account.

Table 1: Moving object cognitive states with description

State	Description
normal (n)	an object continues its motion with consistent visual properties by governing the laws of motion.
new (ne) and exit (e)	when an object enters the scene, it is registered in the memory of an observer and when object under observation leaves the scene, respectively.
occluded (oc) and overlaper (ov)	any object, when disappears due to intersection by another object. Intuitively, the observer infers ”occludee” and ”overlaper” states depending upon the contribution of each objects in occlusion.
reappear (re)	the occluded object reappears and continues its motion when occlusion is over.

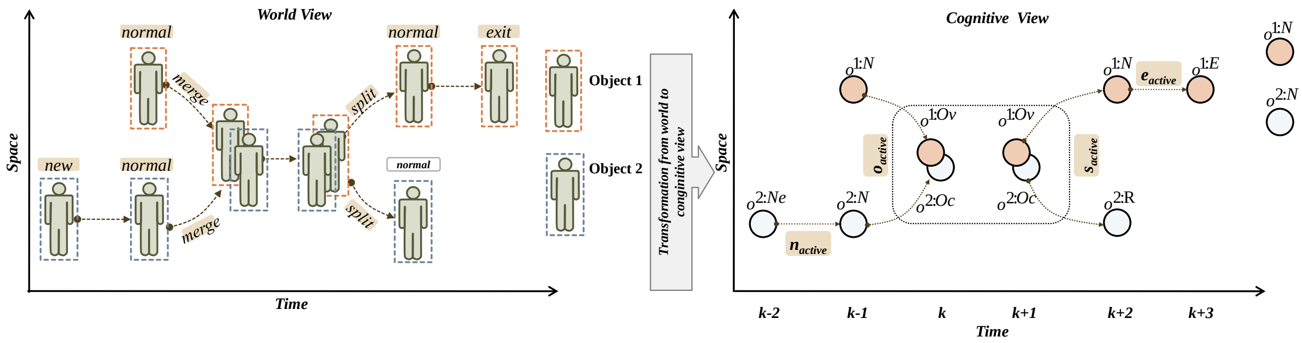


Figure. 4: From the real scene to the concept, and the representation of detected object in world view and cognitive view are presented.

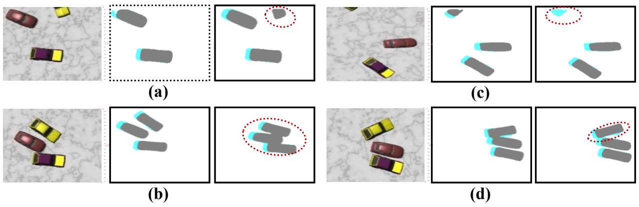


Figure. 3: presents the concept of tracking event detection such as a) new, b) exit, c) occlusion, and d) split where gray pixels represent the existence of object at k and the cyan pixels indicate mapping of pixel at $k-1$.

suggested to handle the ambiguities due to data cluttering in object matching process. The matching weights are put on the tracking axioms to interpret and deduce the appropriate behavioral states of the objects by satisfying the fundamental constraints of continuity during tracking (i.e., frame-by-frame). Essentially, this mechanism refines and handles the conflicted situations to disambiguate the object's behavioral states during tracking. Later, these objects are linked with tracking system where each object is associated with its respective tracker to estimate the trajectories in the scene.
Given a video sequence which is composed of K frames, it is assumed that each detected object is a continuous function of time in the scene until it leaves permanently. The detected object at frame k is:

$$I(k) = \left[o_i^{id}; i = 1, \ldots, n; id > 0\right]; \qquad (5)$$

where $I(k)$ is the image frame at k time instance, o_i^{id} are the i detected objects with unique identities id as shown in Fig. 4. Each detected object contains a unique identity id, its quantitative characteristics Q_q which are processed by statistical approach to measure the matching weight w_m, and set of behavioral states Q_l which are inferred with cognitive approach. So, the object is defined as:

$$o_k^{id} = \{id, w_m, Q_q, Q_l\}; \qquad (6)$$

Each individual object from the above structure is a tuple $\langle id, w_m, Q_q, Q_l \rangle$ where,
id : is the unique identity of object.
k : is the frame number.
Q_q : represents the visual characteristics of the object.

$$Q_q = (f); \qquad (7)$$

the above expression can be rewritten as:

$$Q_q = (\epsilon_h, \zeta_p, area, bb); \qquad (8)$$

Q_l : represents the object behavioral states.

$$Q_l = (n, oc, ov, r, ne, e); \qquad (9)$$

Table 2: criterion for tracking conflict detection

Event: true or false	S_r
new = true	$< 5\%$
occlusion = true	$<= 70\% and > 40\%$
split = true	$< 60\% \ and \ > 30\%$
exit = true	$< 10\%$

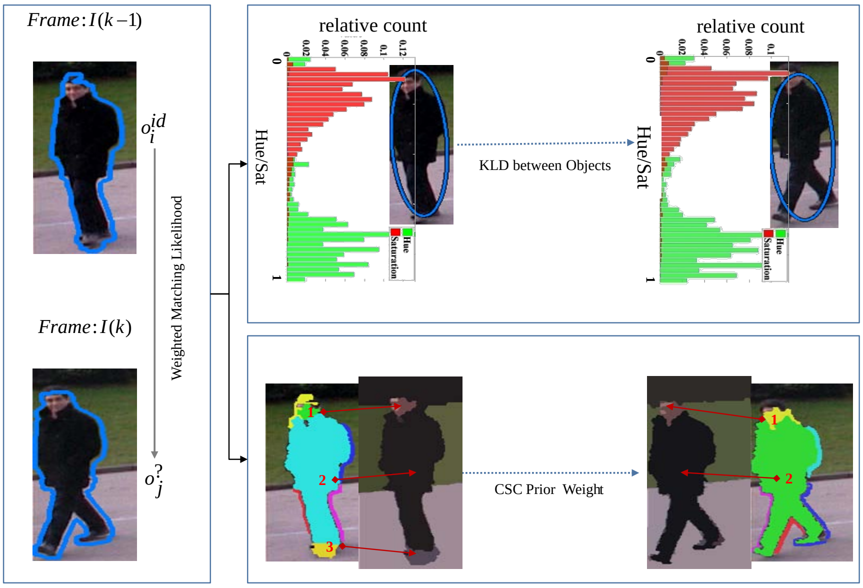


Figure. 5: shows conceptually the BMW approach. The first level defines the main formulations whereas the second and third level show the computation of BMW approach components.

A. *Statistical Approach-Bayesian Matching Weight (BMW)*

A multi-layered statistical algorithm based on Bayesian inference is proposed by formulating the detected object as undirected graph as shown in Fig. 4. The main objective is to compute the matching weights of object in an efficient manner during object tacking in both ideal and partially cluttered situations. To achieve this task, we have aggregated two different techniques in Bayesian inference for computing the matching weights (i.e., posterior probability, unless specified) of an object as shown in Fig. 5. First, the elliptical histogram is approximated around the detected objects at $k-1$ and the predicted objects at k. Second, a novel idea is introduced by exploiting the CSC approach to compute the prior weights of the objects.

Inferencing about an object's possible occurrence at $I(k)$ is made by incorporating the prior probability measured from the possibilities $I(k-1) = \left[o_i^{id}; i=1,\dots,n; id>0\right]$, and the likelihood evidence of the observed data $I(k) = \left[o_j^{'}; j=1,\dots,m\right]$. The posterior probability of object at $I(k)$ corresponding to objects at $I(k-1)$ is computed as:

$$P(o_k^{'} = w_m) = \underbrace{argmax}_{o_j^{'}} \prod_i^n P(o_j^{'}|o_i^{id})P(o_i^{id}); \tag{10}$$

where $P(o_j^{'}|o_i^{id})$ is the likelihood between objects at k and $k-1$ which can be interpreted as $D \in [0,1]$, the distance between objects at $I(k-1)$ and $I(k)$. Similarly, $P(o_i^{id})$ is the objects prior probability at $I(k-1)$ which can be defined as prior weight $w(o_i^{id})_c$ assigned to each object at $I(k-1)$. The above Equation 10, can be re-formalized as:

$$P(o_j^{'} = w_m) \propto \underbrace{argmax}_{o_j^{'}} \prod_i^n (exp(-D(o_j^{'}, o_i^{id})) \times w(o_i^{id})_c); \tag{11}$$

In the following section, we have explained the methodology of measuring the distance D and prior weight w_c.

1) *KL-Divergence Between Objects*

Every object is represented as an elliptical histogram ϵ_h, therefore, to measure the distance, we use KL-Divergence between objects in the corresponding frames as shown in Fig. 5 . The KL-Divergence quantifies the proximity of two distributions which has the intimate relationship with likelihood theory. Let the object $o_j^{'}$ is detected at $I(k)$ and object o_i^{id} is detected at $I(k-1)$, the corresponding distributions are defined as $o_j^{'}(\epsilon_h)$ and $o_i^{id}(\epsilon_h)$. The KL-Divergence is defined as:

$$D(o_j^{'}(\epsilon_h), o_i^{id}(\epsilon_h)) = 1 - \sum_{n=1}^{bins} (o_j^{'}(\epsilon_h(n))) ln \frac{o_j^{'}(\epsilon_h(n))}{o_i^{id}(\epsilon_h(n))}; \tag{12}$$

2) *CSC-Based Object Prior-Weight*

The prior weight is computed based on how much information an object carries at $I(k-1)$ about the newly observed object at $I(k)$ as shown in Fig. 5. In other words, given the set of color-patches representing an object, we have measured the similarity as prior weight and find the most similar color-patches of the objects satisfying the search space criterion. The computed prior weights of color-patches of objects are averaged to measure the final prior weight. We have employed euclidean distance to measure the similarity among the color-patches of the objects. Let object $o_j^{'}$ is detected at $I(k)$ contains color-patches $\zeta_p = c_m^{k:'}$, and the object o_i^{id} is detected at $I(k-1)$ contains color-patches $\zeta_p = c_n^{k-1:id}$. So, the prior weight indicates that how much an object detected at $I(k)$ contains the contents of objects detected at $I(k-1)$ and is computed as follows:

$$w(o_i^{id})_c = 1 - \sqrt{\frac{(o_j^{'}(c_1^{k:'}) - o_i^{id}(c_1^{k-1:id}))^2 + \ \dots \ + (o_j^{'}(c_n^{k:'}) - o_i^{id}(c_m^{k-1:id}))^2}{N}}, \tag{13}$$

$$n = \{1,\dots,N\}, \ \ m = \{1,\dots,M\};$$

where $c_n^{k-1:id} = c_{\overline{ncolor_{rgb}}}$ contains mean normalized RGB color of N color-patches, and $w(o_i^{id})_c$ defines the prior weight of the detected objects based on previous observations (i.e., object detected at $I(k-1)$) as shown in Fig. 5.

The Equations 11, 12, and 13 allow us to compute efficiently the posterior probability of the observations (i.e., $o_j^{'}$) at time k which we interpret as the matching weight (w_m) of our objects. In the ideal situation, the matching probabilities reflect the reliable relationships but under the conflicted situations, these inferences become uncertain. Therefore, it is inevitable to look for the alternative ways and to continue the inferring mechanism correctly under conflicted situations.

B. Cognitive Modeling

In this section, we formulate the tracking axioms (i.e. abstract qualitative reasoning expression) for each of the described cognitive states of the moving object during tracking (see Table. 1). These expressions work in conjunction with statistical model to handle the conflicted situations. Each state inference mechanism follows the rules of cognitive processing which is mapped over the tracking problem. In n-dimensional space, the detected objects at k time instance are mapped on the 2D plane as shown in Fig. 4 where their attributes are defined in Equation 5. The Q_l demonstrates the behavioral states during tracking which are derived by the developed axioms. The Q_l contains six behavioral states, including normal (n), occluded (oc), overlaper (ov), reappear (r), exit (e), and new (ne) whereas to infer each state, a specific axiom is designed by incorporating the human perception abilities.

$Max()$**:** computes the maximum weight of the detected object at k with possible explanations observed at previous frame.

$Min()$**:** computes the minimum weight of the detected object at k with possible explanations observed at previous frame.

$S\_S()$**:** this function checks the existence of object in the predicted region.

$Assign\_Id()$**:** assigns the identity to corresponding object.

$Deactive\_Id()$**:** de-activates the identity when the object is no more in the scene.

$Make\_Child()$**:** creates a child parent (occluded object as child and overlaper as parent) relationship when occlusion is observed.

Normal State Axiom: We assume that each object moves governing the laws of continuous motion with consistent visual attributes. In the following axiom, $Max()$ function returns the associated weight of an object with the list of objects given the object presence in the predicted $S\_S()$ region. Only normal state of the object is activated in Q_l whereas the other states are set to false.

$$normal(o_j^{id}, id) = \left\{ Max_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \wedge \ S\_S_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \right\}$$
$$Q_l = \{n \rightarrow T, oc \rightarrow F, ov \rightarrow F, r \rightarrow F, e \rightarrow F, ne \rightarrow F\}$$
$$Assign\_Id(o_j^{id}) = \left\{ o_i^{id} \right\}$$

New State Axiom: The correspondence of new object is examined with all the observations when $new == true$. The $Min()$ function returns that the new entered object possesses minimum association weight with all possibilities and assigned a new id. Besides, the new object does not fall in the predicted region whereas Q_l is updated by activating new and normal state of object.

$$new(o_j^{'}) = \left\{ Min_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \wedge \neg\ S\_S_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \right\}$$
$$Q_l = \{n \rightarrow T, oc \rightarrow F, ov \rightarrow F, r \rightarrow F, e \rightarrow F, ne \rightarrow T\}$$
$$Assign\_Id(o_j^{id}) = \{id_{new}\}$$

Exit State Axiom: This axiom is called when the exit event is triggered. It is assumed that the object has passed through scene. Function $max()$ returns the association of object from the list of objects. Besides, the registered object must fall in specific exit region $S\_S()$ and Q_l is updated:

$$exit(o_j^{id}, id) = \left\{ Min_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \wedge \neg\ S\_S_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \right\}$$
$$Q_l = \{n \rightarrow F, oc \rightarrow F, ov \rightarrow F, r \rightarrow F, e \rightarrow T, ne \rightarrow F\}$$
$$Deative\_Id(o_j^{id})_{o_j^{'} \in I(k)} = \left\{ o_j^{id} \right\}$$

Overlaper State Axiom: When $occlusion == true$, overlaper state axiom is activated and finds the participation of each objects in occlusion. Both the objects must fall into the conflicted region $S\_S()$ whereas $Max()$ returns weight for overlaper. The $Make\_Child()$ function creates a parent-child relationship and overlaper becomes the parent of occludee. After occlusion, child adopts the visual features of its parent which is updated frame-by-frame using depth first search strategy. Besides, Q_l of the corresponding object is updated.

$$overlaper(o_j^{id}, id) = \left\{ Max_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \wedge \ S\_S_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \right\}$$
$$Q_l = \{n \rightarrow T, oc \rightarrow F, ov \rightarrow T, r \rightarrow F, e \rightarrow F, ne \rightarrow F\}$$
$$Assign\_Id(o_j^{id}) = \left\{ o_i^{id} \right\}$$
$$Make\_Child(o^{id}_{j+1}{}_{if(\exists(o^{id}_{j+1} \in I(k))Q_l = \{oc=T\})},$$
$$o_j^{id}{}_{if(\exists(o_j^{id} \in I(k))Q_l = \{ov=T\})})$$

Occludee State Axiom: The following tracking axiom determines state for the occludee object. The weight of occluded object is computed through statistical modeling which must be less than the weight of the overlaper as it is assume that the occluded object is being hidden by overlaper and therefore lost its visual context. Consequently, object becomes child of its occludee state of object is activated.

$$occluded(o_j^{id}, id) = \left\{ Min_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \wedge \ S\_S_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{id}) \right\}$$
$$Q_l = \{n \rightarrow T, oc \rightarrow T, ov \rightarrow F, r \rightarrow F, e \rightarrow F, ne \rightarrow F\}$$
$$Assign\_Id(o_j^{id}) = \left\{ o_i^{id} \right\}$$

Reappear State Axiom: The formulation determines reappeared state of the object when $split$ event is active. The reappeared object's relation is computed with the list of occluded objects. The relationship of child-parent is ended and the Q_l is updated.

$$reappear(o_j^{id}, id) = \left\{ Max_{o_j^{'} \in I(k)}(o_j^{'}, o_i^{*id}) \right\}$$
$$Q_l = \{n \rightarrow T, oc \rightarrow F, ov \rightarrow F, r \rightarrow T, e \rightarrow F, ne \rightarrow F\}$$
$$Assign\_Id(o_j^{id}) = \left\{ o_i^{*id} \right\}$$

Each detected object at $I(k)$ is assigned a unique id with respective behavioral states. Now, the next task is to perform object localization at each time instance. In the following, we have developed a Kalman filter based tracking system to estimate the object trajectories over time.

VII. Tracking Module

Mathematically, Kalman filter is an estimator that predicts and corrects states of a wide range of linear processes [14]. In our tracking system, each Kalman filter is defined in terms of its process (i.e., x_k), measurement model (i.e., z_k), and available information about the model's initial conditions which are governed by linear stochastic difference and measurement equation respectively:

$$x_k = \begin{bmatrix} x \\ y \\ dx/dk \\ dy/dk \end{bmatrix}, \quad z_k = \begin{bmatrix} x \\ y \\ dx/dk \\ dy/dk \end{bmatrix}, \tag{14}$$

$$x_k = Ax_{k-1} + w_{k-1}, \tag{15}$$

$$z_k = Hx_k + v_k, \tag{16}$$

The matrix A in the difference Equation 15 relates the state at the previous time $k-1$ to the state at the current time k, in the absence of either a driving function or process noise w_{k-1}. H in the measurement Equation 16 relates the state x_k to the measurement z_k. In practice, both A and H matrices can change with each time step, but here we assume that it is constant.

After the initialization, in the next frames, the normal state updation continues until any other tracking events (i.e. occlusion, split, new, or exit) are detected. When the objects are occluded, Kalman filter of the occluded object follows the states and measurement information of the corresponding overlaper object. In contrast, during the split, the Kalman filter resumes the tracking by taking into account the parameters of its own object to perform estimations. In this manner, we are able to perform object localization with classical Kalman filter in a linear and non-linear situations.

VIII. Experimental Results

Several experiments are conducted on state of the art video sequences. The initial dataset selected for the development of ideas is taken from IESK, OvG University called IESK dataset. The videos are filmed in the vicinity of campus to capture real attitude of objects using a single static camera.

Fig. 7 shows key frames from important instances of the sequences. The results are visualized by trajectory of the object and labeling of object identity with respective behavior. Moreover, Tracking and Behavior Information Interface (TnBII) Panel describes the overall information that includes: object identities with behaviors, object pace, orientation, and tracking events for each corresponding frame. **Frame 39** shows the initial situation of the tracking and behavior understanding. All the detected objects are assigned unique ids **1**, **2**, and **3** with trajectories indicating their tracks. It is observed that yellow truck and red cars shared similar spatial region. Therefore, it is not possible to distinguish between them and the same identity is assigned. **Frame 49** shows a new object which is entered in the scene from the back view and is assigned a unique id **3**. This object continues its path till the end while the events of occlusion and split are observed multiple times, and the behavior of the object is updated accordingly. **Frame 103** shows the object with id **3** left the scene. Its respective tracker terminates the estimation task, and the identity is released so that it will be assigned to other objects. It is notable from the trajectory of the object that how the object keeps its track during motion. The **TnBII** panel demonstrates the updated information about the object behavior, its pace, and orientation along with the respective tracking events.

The second dataset which is used for testing is taken from PETS2006. The PETS series of workshops make available public datasets for tracking and behavior understanding tasks. Fig. 8 demonstrates the tracking results on PETS2006. **Frame 3** shows that every detected object is identified by their unique identities and associated behavioral states from our integrated modules of the framework where the corresponding trajectory demonstrates the outcome of tracker. **Frame 44** shows the occlusion situation when two objects with ids **0** and **1** interact with each other. The object with id **0** is assigned overlaper behavioral state whereas the object with id **1** is assigned the occluded behavioral state. The tracker of occluded object tracker is unable to advance the localization. **Frame 59** indicates that the split event is observed and the occluded object is reappeared from the occlusion phase. The tracker re-estimates its path based on its own visual characteristics and physical location. The **TnBII** panel demonstrates the respective behavior of objects during normal, occluded, and reappear situations. It is also noticeable that object with id **4** keeps its motion and state quite suspiciously in the scene.

A. Evaluation

We have evaluated our tracking and behavior understanding framework on the basis of generating the correct identities and behaviors corresponding to objects during tracking. For such evaluation, the first essential requirement is the ground truth. For this purpose, we have manually assigned the identities to objects and interpret their respective behaviors. Finally, the performance is evaluated by computing precision and recall measures. In the context of identities (i.e., analogous to tracks) and behaviors, precision and recall measures are defined as follows:

$$precision = \frac{Number\ of\ correct\ identities\ or\ behaviors}{Number\ of\ established\ identities\ or\ behaviors}, \tag{17}$$

$$recall = \frac{Number\ of\ correct\ identities\ or\ behaviors}{Number\ of\ actual\ identities\ or\ behaviors}, \tag{18}$$

where *actual identities or behavior* denotes the identities or behaviors available in the ground truth. Moreover, the evaluation is performed based on their ability to detect tracking events: 1) deal with entry and exit of objects, 2) handles occlusion event, and 3) handles the split event when objects are reappeared from occlusion.

Table 3 presents the quantitative performance of the proposed framework on test sequences. It is important to observe that the precision and recall of object identity recognition and normal behaviors are more prominent in performance. The precision and recall of the exit and new event are interrelated, because, if the object is wrongly classified as exit behavior then in the next instance, the algorithm will treat that object as a new. So, the recall is better but precision is degraded. Table 4 presents the performance values for the tracking event detection algorithm which is developed along with the tracking and behavior understanding framework. In fact, the de-

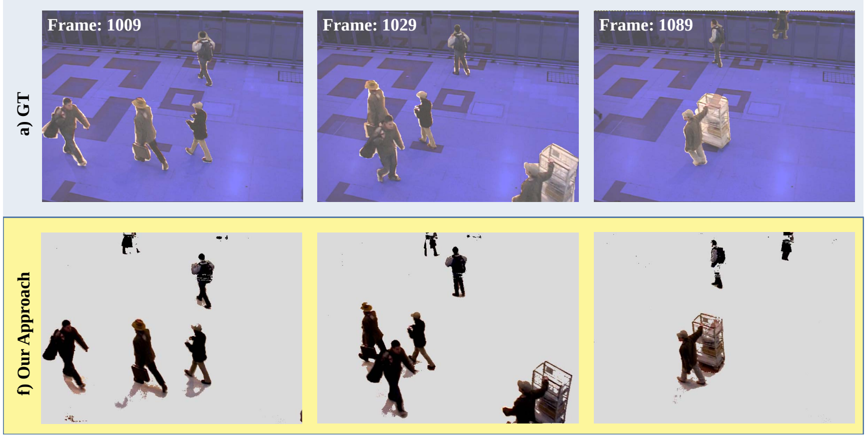
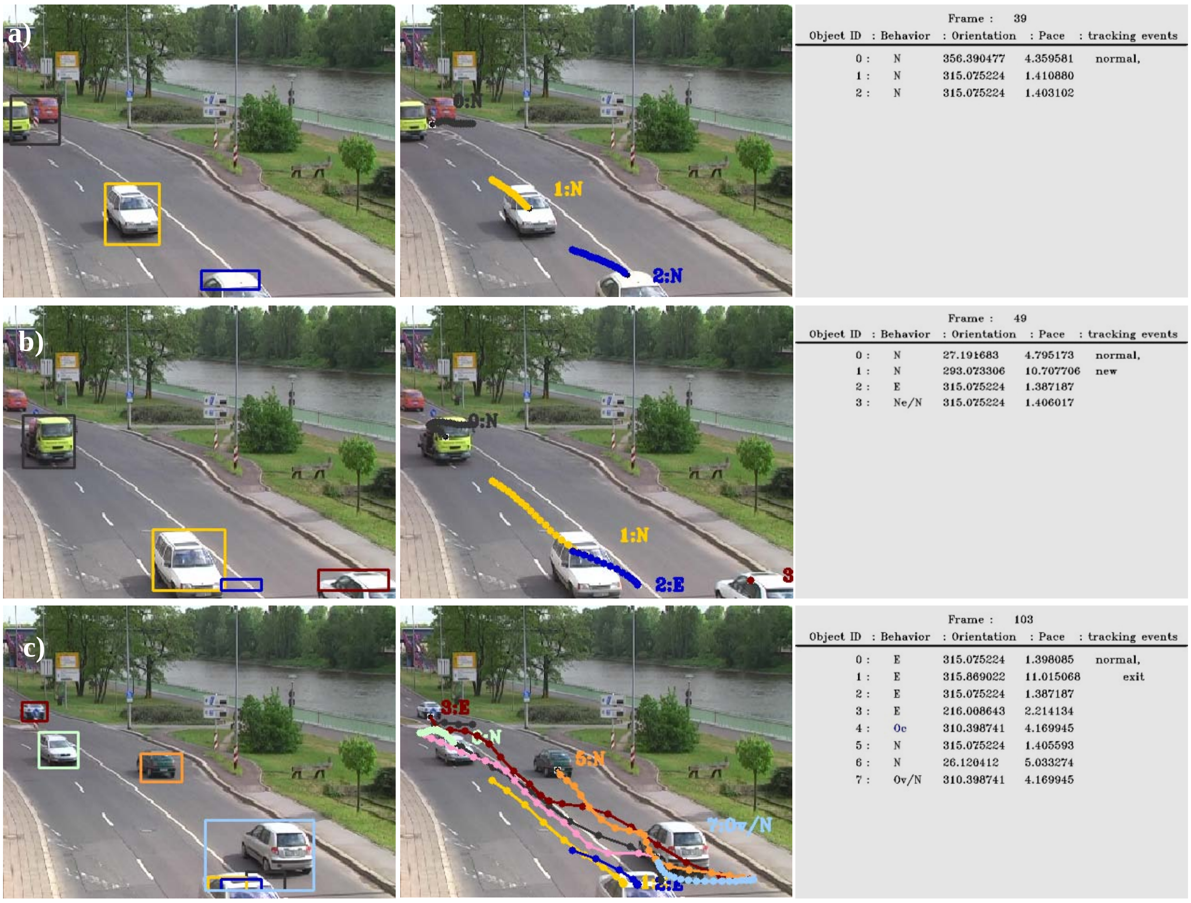


Figure. 6: shows the segmentation result from our segmentation algorithm in test sequence.



```
                    Frame :    39
Object ID  : Behavior  : Orientation  : Pace   : tracking events
       0 :     N       356.390477   4.359581    normal,
       1 :     N       315.075224   1.410880
       2 :     N       315.075224   1.403102
```

```
                    Frame :    49
Object ID  : Behavior  : Orientation  : Pace   : tracking events
       0 :     N       27.191683    4.795173    normal,
       1 :     N       293.073306   10.707706   new
       2 :     E       315.075224   1.387187
       3 :     Ne/N    315.075224   1.406017
```

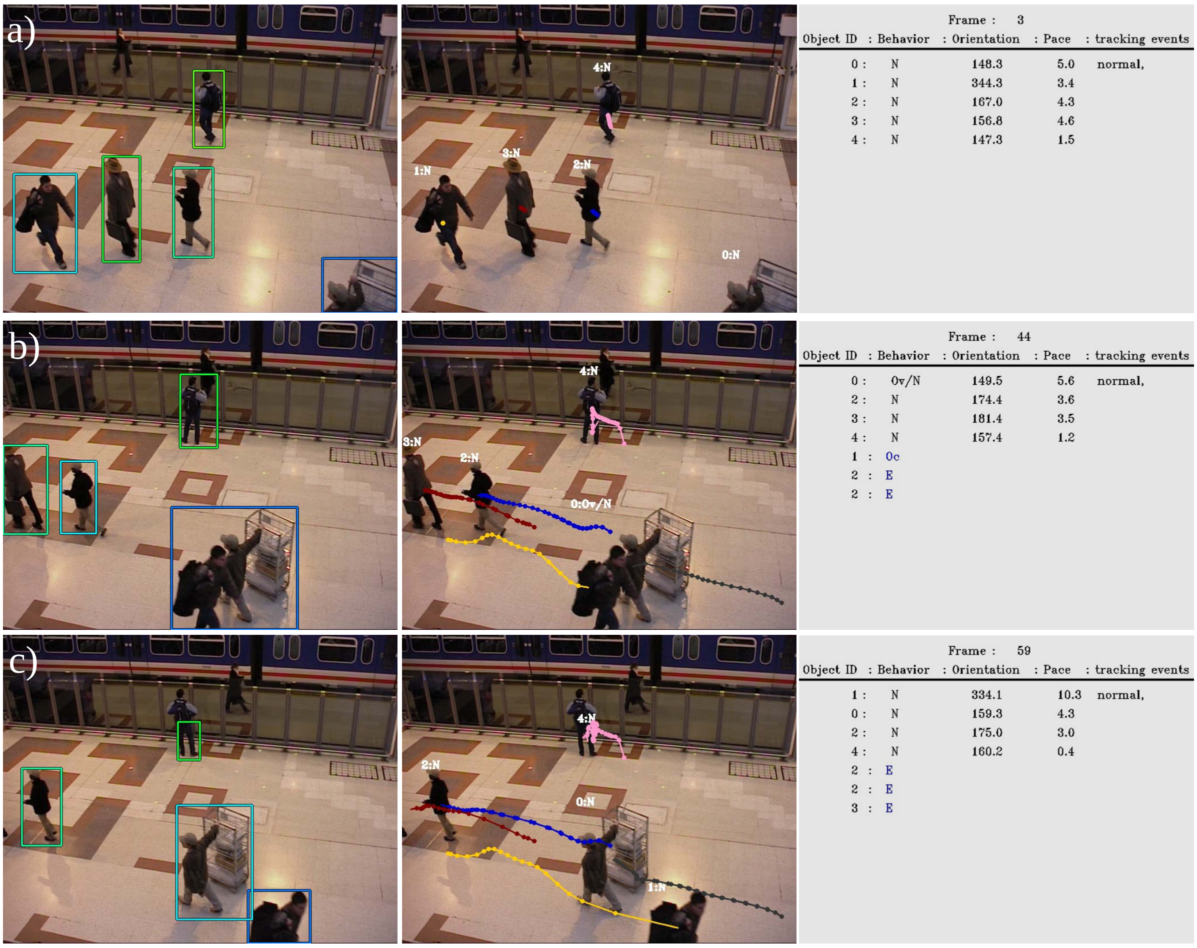
```
                    Frame :    103
Object ID  : Behavior  : Orientation  : Pace   : tracking events
       0 :     E       315.075224   1.398085    normal,
       1 :     E       315.869022   11.015068       exit
       2 :     E       315.075224   1.387187
       3 :     E       216.008643   2.214134
       4 :     Oc      310.398741   4.169945
       5 :     N       315.075224   1.405593
       6 :     N       26.120412    5.033274
       7 :     Ov/N    310.398741   4.169945
```

Figure. 7: shows the results of tracking object in test sequences. Each object is described with a label depicting unique identity and the motion states (i.e. **id:motion stats = 1:N**) whereas the motion trajectory shows the tracking path. We have used a unique color for each object to keep distinction when the conflicts are observed.

tection of accurate events will lead to significant improvements in the results of Table 3, in some way. For instance, the specific axiom is called when a particular tracking event is activated. However, the mechanism of assigning identities, their management, and behavior inferencing is achieved by incorporating tracking axioms which control this type of discrepancy up to some extent in integrated statistical and cognitive modeling approaches. Moreover, it is observed that the recall values are dominant over precision because of many factors. For instance, mis-detections of objects may result in satisfying the condition for the exit events which consequently activates the new events in the next frames. However, the developed constraints for object states prevent these effects in some situation, but at the same time it affects the precision of the normal events.

IX. Conclusion and Future Work

We considered the issues of generic and practical importance in tracking and behavior understanding during surveillance.



```
                 Frame :    3
Object ID  : Behavior  : Orientation  : Pace  : tracking events

       0 :     N           148.3        5.0   normal,
       1 :     N           344.3        3.4
       2 :     N           167.0        4.3
       3 :     N           156.8        4.6
       4 :     N           147.3        1.5
```

```
                 Frame :    44
Object ID  : Behavior  : Orientation  : Pace  : tracking events

       0 :    Ov/N         149.5        5.6   normal,
       2 :     N           174.4        3.6
       3 :     N           181.4        3.5
       4 :     N           157.4        1.2
       1  :  Oc
       2  :  E
       2  :  E
```

```
                 Frame :    59
Object ID  : Behavior  : Orientation  : Pace  : tracking events

       1 :     N           334.1       10.3   normal,
       0 :     N           159.3        4.3
       2 :     N           175.0        3.0
       4 :     N           160.2        0.4
       2  :  E
       2  :  E
       3  :  E
```

Figure. 8: shows the results of tracking object in test sequences. Each object is described with a label depicting unique identity and the motion states (i.e. **id:motion stats = 1:N**) whereas the motion trajectory shows the tracking path. We have used a unique color for each object to keep distinction when the conflicts are observed.

The contention is to combine the domain specific logical hypothesis with statistical model whereas satisfying the global continuity constraint for tracking in real-time. We have performed a qualitative analysis with the annotated ground truth to verify the performance of our proposed approach. Future research will be more focused on investigating the object specific actions and event interpretation in dynamic scenes.

Acknowledgments

This work is supported by Transregional Collaborative Research Center SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” funded by the German Foundation (DFG).

References

[1] Dee, H.M., Velastin, S.A.: How close are we to solving the problem of automated visual surveillance?: A review of real-world surveillance, scientific progress and evaluative mechanisms. Machine Vision Application **19** (2008) 329–343

[2] Cox, I.J., Hingorani, S.L.: An efficient implementation of reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on **18** (2002) 138–150

[3] Isard, M., Maccormick, J.: Bramble: a bayesian multiple-blob tracker. In: IEEE International Conference on Computer Vision. (2001) 34–41

[4] Smith, K., Gatica-Perez, D., Odobez, J.M.: Using particles to track varying numbers of interacting people. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR ’05, IEEE Computer Society (2005) 962–969

[5] Ryoo, M.S., Aggarwal, J.K.: Observe-and-explain: A new approach for multiple hypotheses tracking of humans and objects. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2008)

[6] Guo, Y., Hsu, S., Sawhney, H.S., Kumar, R., Shan, Y.: Robust object matching for persistent tracking with heterogeneous features. IEEE Transaction Pattern Analysis Machine. Intellgence **29** (2007) 824–839

[7] Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR ’06, IEEE Computer Society (2006) 1528–1535

[8] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR’05 - Volume 1, Washington, DC, USA, IEEE Computer Society (2005) 886–893

Table 3: Precision and Recall of Tracking and Behavior Understanding Framework

Dataset	identities		normal		overlaper		occluded		reappear		new		exit	
	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.
IESK	0.88	0.89	0.94	0.90	0.91	0.91	0.81	0.85	0.85	0.9	0.83	0.79	0.65	0.83
PETS 2006	0.86	0.94	0.98	0.85	0.90	0.95	0.85	0.99	0.73	0.99	1	1	0.69	0.9

Table 4: Precision and Recall of Tracking Event Detection

Dataset	normal		occlusion		split		new		exit	
	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.	pre.	rec.
IESK	0.81	0.93	0.85	0.91	0.85	0.79	0.81	0.78	0.81	0.76
PETS2006	0.75	0.91	0.88	0.86	0.81	0.81	0.82	0.791	0.76	0.77

[9] Glasgow, J., Karan, B., N. Narayanan, E.: Diagrammatic Reasoning, Cambridge Mass. AAAI Press/The MIT Press (1995)

[10] Sherrah, J., Gong, S.: Resolving visual uncertainty and occlusion through probabilistic reasoning. In: In Proceedings of the British Machine Vision Conference. (2000) 252–261

[11] Bennett, B., Magee, D., Cohn, A.G., Hogg, D.: Enhanced tracking and recognition of moving objects by reasoning about spatio-temporal continuity. Image Vision Computer **26** (2008) 67–81

[12] Halpern, J.Y.: An analysis of first-order logics of probability. Artificial Intelligence **46** (1990) 311–350

[13] Priese, L., Rehrmann, V.: A fast hybrid color segmentation method. In: Proceedings Mustererkennung, DAGM Symposium, Springer Verlag (1993) 297–304

[14] Welch, G., Bishop, G.: An introduction to the kalman filter. Technical report, University of North Carolina at Chapel Hill. (1995)

A. Biography

Saira Saleem Pathan is a PhD student in Technical Computer Science Group at the Otto-von-Guericke-University, Magdeburg, Germany. She received her Bachelors degree in Computer Systems Engineering from Mehran University of Engineering and Technology, Jamshoro, Pakistan. Further, she has completed her Master degree from Institute of Information Technology, Mehran University of Engineering and Technology Jamshoro, Pakistan. Her current research is focused on motion analysis and pattern recognition.

Omer Rashid is a PhD student in Otto-von-Guericke University, Magdeburg, Germany. He received his Bachelors degree in Computer Engineering from University of Engineering and Technology, Lahore, Pakistan. Further, he has completed his Master degree from Otto-von-Guericke University, Magdeburg, Germany. His current research is focused on human computer interaction, image processing and pattern recognition.

Ayoub K. Al-Hamadi was born in Yemen in 1970. He received his Masters Degree in Electrical Engineering in 1997 and his Ph.D. in Technical Computer Science at the Otto von Guericke University, Magdeburg, Germany in 2001. Since 2002 he has been Junior-Research-Group-Leader at the Institute for Electronics, Signal Processing and Communications at University of Magdeburg. In 2008 he became Professor of Neuro-Information Technology. His research work concentrates on the field of image processing, pattern recognition and artificial neural networks. Professor Al-Hamadi is the author of more than 120 articles in peer-reviewed international journals and conferences.

Bernd Michaelis was born in Magdeburg, Germany in 1947. He received a Masters Degree in Electronic Engineering from the TH Magdeburg in 1971 and his first Ph.D. in 1974. Between 1974 and 1980 he worked at the TH Magdeburg and was granted a second doctoral degree in 1980. In 1993, he became Professor of Technical Computer Science at Otto-von-Guericke-University, Magdeburg, Germany. His research work concentrates on the field of image processing, artificial neural networks, pattern recognition, processor architectures, and microcomputers. Professor Michaelis is the author of more than 200 articles.