

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

## Empirical Analysis of Video Streams Driven based Human Activity Recognition System using Bayesian Foraging Strategy

Ms. Vrushali Rajshekhar Bhimanavaru<sup>1</sup>, Mrs. Swati Joshi<sup>2</sup>

Department of Computer Science and Engineering, Sinhgad College of Engineering, Kegaon, Solapur, India

Head of Department, Department of Computer Science and Engineering, Sinhgad College of Engineering, Kegaon,  
Solapur, India

**ABSTRACT:** In modern world, each and every activity of people can be trapped by cameras and other video recording devices. This kind of scenario is helpful to many activities such as identifying the behavior of individuals as well as this kind of system is useful to monitor person hidden and for security purposes. This kind of activities are collectively called as Video processing, many video processing and manipulation techniques are available in market and all are having certain issues in working such as: (i) Limited area coverage in perceptual view, (ii) Single camera can focus only one portion, (iii) Need to use separate cameras for trapping separate behavior of individuals and (iv) Accuracy is not much concentrated. A new solution is required to solve these issues and perform with more perfect feature extraction as well as automatic behavior analysis based on face and body recognition. These kind of activities such as: activity analysis and movement tracking are handled by means of gaze shift tech and Bayesian foraging theory. With this system, the individual action is captured by two or more cameras and extracting the features by means of Region-of-Interest [ROI] selection. The motion tracking and Activity analysis schemes are performed by using Bayesian Methodology. In this proposed approach four major activities are taken care such as: Reading Books, Picking Objects, Taking over Phone and Hand Waving.

**KEYWORDS:** Gaze Shift Technique, Bayesian Model, Activity Analysis, Multi-Camera Monitoring, ROI.

### I. INTRODUCTION

The volume of data collected by current networks of cameras for video surveillance clearly overburdens the monitoring ability of human viewers to stay focused on a task. Further, much of the data that can be collected from multiple video streams is uneventful. Thus, the need for the discovery and the selection of activities occurring within and across videos for collating information most relevant to the given task has fostered the field of multi-stream summarize. At the heart of multi-stream summarization there is a “choose and leave” problem that moment to moment an ideal or optimal observer (say, a software agent) must solve: choose the most informative stream; detect, if any, interesting activities occurring within the current stream; leave the handled stream for the next “best” stream. We provide a different perspective to such “choose and leave” problem based on a principled framework that unifies overt visual attention behavior and optimal foraging.

The framework we propose is just one, but a novel, way of formulating the multi-stream summarization problem and solution, for a discussion). In a nutshell, we consider the foraging landscape of multiple streams, each video stream being a foraging patch, and the ideal observer playing the role of the visual forager. According to Optimal Foraging Theory (OFT), a forager that feeds on patchily distributed preys or resources spends its time traveling between patches or searching and handling food within patches. While searching, it gradually depletes the food; hence, the benefit of staying in the patch is likely to gradually diminish with time. Moment to moment, striving to maximize

# International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

its foraging efficiency and energy intake, the forager should make decisions: Which is the best patch to search? Which prey, if any, should be chased within the patch? When to leave the current patch for a richer one? Here visual foraging corresponds to the time-varying overt deployment of visual attention achieved through oculo-motor actions, namely, gaze shifts. Tantamount to the forager, the observer is pressed to maximize his information intake over time under a given task, by moment-to-moment sampling the most informative subsets of video streams. Such monitoring decisions have to be made by relying upon automatic interpretation of scenes for detecting actions and activities.

To be consistent with the terminology proposed in the literature, an action refers to a sequence of movements executed by a single object ( e.g., “human walking” or “vehicle turning right”). An activity contains a number of sequential actions, most likely involving multiple objects that interact or co-exist in a shared common space monitored by single or multiple cameras (e.g., “passengers walking on a train platform and sitting down on a bench”). The ultimate goal of activity modelling is to understand behavior, i.e. the meaning of activity in the shape of a semantic description. Clearly, action/ activity/behavior analysis entails the capability of spotting objects that are of interest for the given surveillance task. Thus, in the work presented here the visual objects of interest occurring in video streams are the preys to be chased and handled by the visual forager.

## II. SURVEILLANCE FEATURES AND DECISION ANALYSIS

Decisions at the finer level of a single stream concern which object is to be chosen Video surveillance is a type of surveillance that is found in different kinds of areas, such as public buildings, metro stations and military areas. It has proven to be an effective tool to monitor large areas with limited resources. Because cameras and surveillance systems are continuously developed, these systems become increasingly cost efficient, allowing for larger systems as well. Traditionally, multiple static security cameras are positioned throughout the area, and attached to monitor screens. These screens are monitored by security personnel in order to detect suspicious behavior. Examples of such situations are when a person enters a restricted area, when a person suddenly starts running or when a person has been following another person or a period of time. In case of a suspicious situation the security personnel is alarmed, and relevant information is displayed to allow the personnel to take action if required. Depending on the seriousness of the situation, further action may be taken in order to resolve it.

Though computers may not be as good at computer vision and reasoning as human operators, they provide different advantages. For instance, computers are capable of working 24 hours a day, 7 hours a week. By using an automated intelligent multi-camera video surveillance system the security personnel could be supported in their work, allowing for less security flaws in monitoring the area. By using computers to support the security personnel in their task to watch the monitor screens, some of these problems may be prevented. Computers, unlike humans, are capable of working continuously, without losing focus. Moreover, computers are well scalable to allow for monitoring a large number of security cameras simultaneously. Additionally, computers can communicate constantly, allowing each computer to have a complete and up-to-date view of the monitored area.

The goal of this project was to develop a system that is able to assist the security personnel in observing the NLDA area and alarming the control room in case of a suspicious situation. In order to achieve the project's goals, the prototype system utilizes the static security cameras located in the area. Each camera is linked to a client software application that is able to detect moving objects from the camera images. This client computes different features, such as the locations of the detected objects and whether they are human or not. Because suspicious behavior may occur over multiple cameras, these extracted features are then transmitted to a central server. The central server application gathers, combines and maintains the information from the different security cameras and then reasons about the situation.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

### III. EMOTION ANALYSIS AND MONITORING

Emotion is a feeling evoked by environmental stimuli or by internal body states, which modulates human behavior in terms of actions in the environment or changes in the internal states of the body. The idea of emotional expressions was first recognized by Darwin, who discovered that the natural selection process is not only applicable to anatomic patterns but also can be applied to animal behavior and mind. Emotion is a vital part of human behavior, which is derived from a Latin word which means “move”. It depends on the influences of the external factors such as interaction with people and environmental conditions. In interactive cognitive environments, the evolution of smart spaces is defined by the ambient intelligence paradigms. Emotions take their shape according to the conditions of the environment around and take part in response according to that situation. Emotions are brain processes which evolve to regulate behaviors under different encountered situations. The decision making about these different encountered situations is governed by the evaluation dimension added by emotions, which plays a vital role in it. Emotions are produced as a result of complex (conscious or unconscious) cognitive processes.

#### Past Research Summary

The existing framework is based on the collection of the local features from a crowd and using the Damasio findings to make these observations in causal relationships based on dynamic Bayesian networks. These causal relationships are then being clustered using SOM to define events.

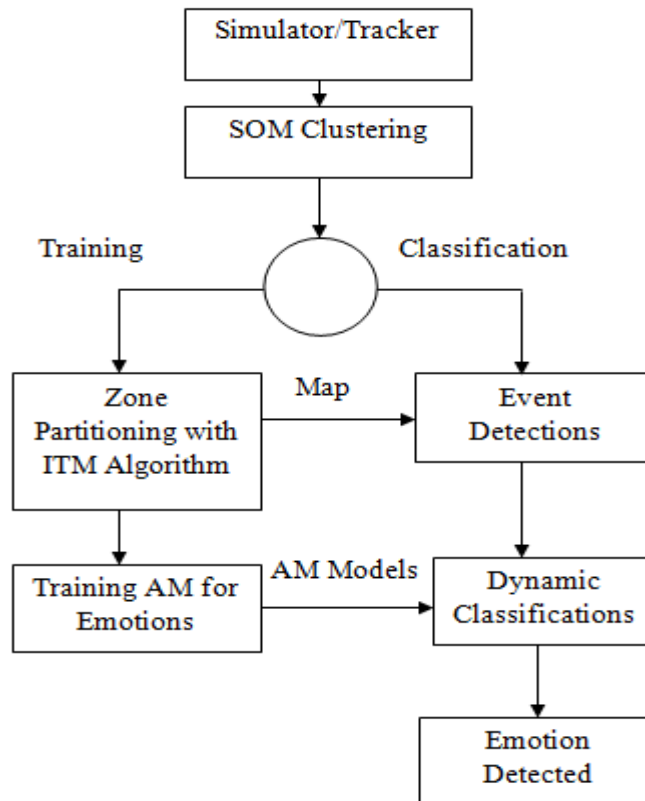


Fig.1 Flow Chart Design

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

The Ortony, Clore and Collins (OCC) theory are used to define causes of emotions for learning and classification and for calibrating the model. Russell et al. circumflex model of affect is being used for calibrating the emotion on valance scale. The above figure-1 shows the flowchart of proposed algorithm. The emotion modeling and classification in a crowd is based on the OCC theories, in which emotions are defined as balanced reaction to the events, agents and objects. The developed theoretical approach of OCC is the assumption that emotions develop as a consequence of certain cognitive processes and their interpretations. Therefore, OCC model involves the cognitive triggering of emotions due to events, agents and objects.

Each cognitive trigger has its causes due to some local variable such as urgency for an event or for other liking or dis-liking of an event, liking or dis-liking of an agent's action and the attractiveness for an object. Each emotion evolves due to unique fusion of each of these variables, which constitutes the cognitive structure of such an emotion. These elicitations in our model are being modeled by the local variables from sensors. The interaction association among people is modeled on the basis of neurophysiologist. Damasio explains the self-conscious is influenced by the conscious interactions with the environment. The brain manipulates the internal state of humans and its relationship with external factors that includes agents, objects and events, defined as proto self and core self. These relationships are stored as visual interpretations the brain's relationships with objects and feelings emerged due to each kind of interaction with the objects. These are stored in the form of images and these images are sensory perceptions of brain interactions. These imagery interaction memories are dependent on visual sensory inputs as well as other sensors such as auditory, somatosensory etc. The images are managed by a pattern of neural networks; the neural network involved in interpretations of interactions and activities constitutes the neural patterns with the help of sensory inputs.

The proto self and core self, which are the visual interpretations/ representations constituted by neural network patterns can be represented as first order neural networks. The causal interaction patterns is between organism and object generate core conscious. Moreover, second order neural patterns that are generated due to the causal relationships among different entities and its effects from inputs from visual perception representations, which modifies itself into proto self. Damasio also claimed that the core conscious is continuously evolving due to interaction with external perceptions. The neuro-physiological observations mentioned above are used to model interactions among the people. The interactions are causal relationships inspired from the Damasio theory. The causal relationships among the people are the basis of the features of our model.

## V. PROPOSED SYSTEM MODEL

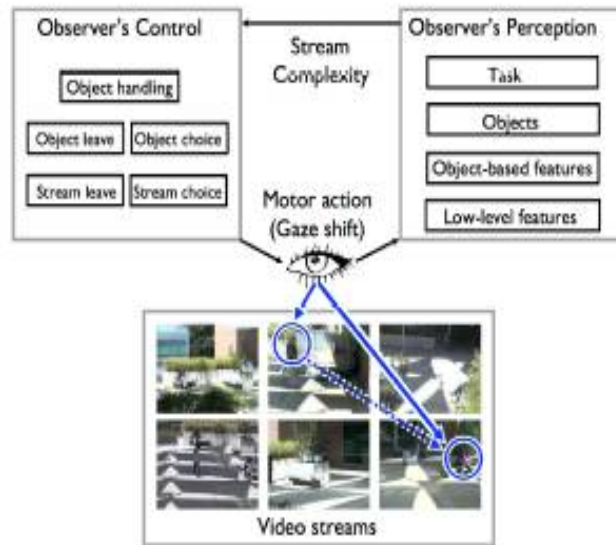
We propose a probabilistic framework to model observer's attentive behavior as the behavior of a forager with the help of Person Identification and activity tracking process. The forager, moment to moment, focuses its attention on the most informative stream/camera, detects interesting objects or activities, or switches to a more profitable stream. The approach proposed here is suitable to be exploited for multi-stream video summarization. Meanwhile, it can serve as a preliminary step for more sophisticated video surveillance, e.g. activity and behavior analysis. Experimental results achieved on the UCR Video-web Activities Dataset, a publicly available dataset, are presented to illustrate the utility of the proposed technique. An input to our system is a visual landscape of  $K$  video streams, each stream being a sequence of time parameterized frames  $\{I(k)(1), I(k)(2), \dots, I(k)(t), \dots\}$ , where  $t$  is the time parameter and  $k \in [1, \dots, K]$ . Denote  $D$  the spatial support of  $I(k)$ , and  $r(k) \in D$  the coordinates of a point in such domain. By relying on the perception/action cycle outlined in Fig. 1, at any point  $t$  in time, we designate the current gazed frame  $I(k)(t)$  of stream  $k$  as the relevant video frame to be selected and included in the final output summarization.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019



**Fig.2 Multi-Video Stream Monitoring**

To such end, each video stream is the equivalent of a foraging patch (cfr. Table I) and objects of interest (preys) occur within the stream. In OFT terms, it is assumed that: the landscape is stochastic; the forager has sensing capabilities and it can gain information on patch quality and available preys as it forages. Thus, the model is conceived in a probabilistic framework. Use the following random variables (RVs):

- ✚ **T**: a RV with  $|T|$  values corresponding to the task pursued by the observer.
- ✚ **O**: a multinomial RV with  $|O|$  values corresponding to objects known by the observer

As a case study, we deal with actions and activities involving people. Thus, the given task  $T$  corresponds to “pay attention to people within the scene”. To this purpose, the classes of objects of interest for the observer are represented by faces and human bodies, i.e.,  $O = \{\text{face, body}\}$ . Since we assume unitary time for between-stream shifts, in the following we will drop the  $k$  index and simply use  $rF$  to denote the center of the FoA within the frame without ambiguity. Relocations occur because of decisions taken by the observer upon his own perceptual inferences. In turn, moment to moment, perceptual inferences are conditioned on the observer’s current FoA set by the gaze shift action.

## VI. LITERATURE SURVEY

In the year of 2014, the authors "S. Haykin and J. M. Fuster" proposed a paper titled "On cognitive dynamic systems: Cognitive neuroscience and engineering learning from each other", in that they described such as: Cognitive dynamic systems provide a broadly defined platform, whereby engineering learns from cognitive neuroscience, and by the same token, cognitive neuroscience learns from engineering. The first part of the paper is of a tutorial nature, addressing recent advances in cognitive perception and cognitive control, which are the dual of each other. The study of cognitive perception, viewed from the perspective of Bayesian inference, starts with sparse coding, well known in neuroscience. However, sparse coding could become ill-posed, particularly when the signal-to-noise ratio is low. In such situations, stability is a necessary requirement, which can only be satisfied if there is sufficient information in the observables. To satisfy this requirement, the sparse-coding algorithm is augmented by the addition of information filtering (i.e., a special case of Bayesian filtering). Accordingly, the performance of sparse coding is improved under the influence of perceptual attention. This improvement enhances the cognitive perceptor to separate relevant information from irrelevant information. Next, moving into cognitive control, viewed from the perspective of Bellman's dynamic programming, two ideas are exploited: entropic state of the preceptor, and the definition of reward as an

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

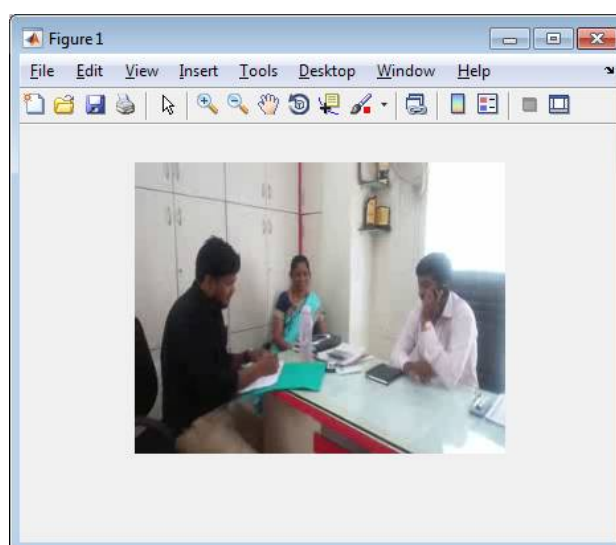
Vol. 8, Issue 5, May 2019

invertible function of two entropic states, namely, the current state and its immediate past value. The net result of building on these two ideas is a modified form of Bellman's dynamic programming, and, therefore, a new reinforcement learning algorithm, which not only outperforms traditional reinforcement learning algorithms, but also offers some highly desirable properties. Among them is a linear law of computational complexity, which is the best that it could be. The second part of the paper addresses two challenging problems: first, how to mediate between cognitive control and cognitive perception and, second, how to formulate a procedure for risk control.

In the year of 2014, the authors "Morye, C. Ding, A. Roy-Chowdhury, and J. Farrell" proposed a paper titled "Distributed constrained optimization for bayesian opportunistic visual sensing", in that they described such as: a control mechanism to obtain opportunistic high resolution facial imagery, via distributed constrained optimization of the pan, tilt, zoom (PTZ) parameters for each camera in a sensor network. The objective function of the optimization problem quantifies the per camera per target image quality. The tracking constraints, which are a lower bound on the information about the estimated position for each target, define the feasible PTZ parameter space. Each camera alters its own PTZ settings. All cameras use information broadcast by neighboring cameras such that the PTZ parameters of all cameras are simultaneously optimized relative to the global objective. At certain times of opportunity, due to the configuration of the targets relative to the cameras, and the fact that each camera may track many targets, the camera network may be able to reconfigure itself to achieve the tracking specification for all targets with remaining degrees of freedom that can be used to obtain high-res facial images from desirable aspect angles for certain targets. The challenge is to define algorithms to automatically find these time instants, the appropriate imaging camera, and the appropriate parameter settings for all cameras to capitalize on these opportunities. The solution proposed herein involves a Bayesian formulation in a game theoretic setting. The Bayesian formulation automatically trades off objective maximization versus the risk of losing track of any target. This paper presents the theoretical solution along with the simulation results.

## VII. EXPERIMENTAL RESULTS

The following figure illustrates the frame estimation scenario of the proposed system and shows the input video sample frame in this.



**Fig.3 Input Video Sample Frame**



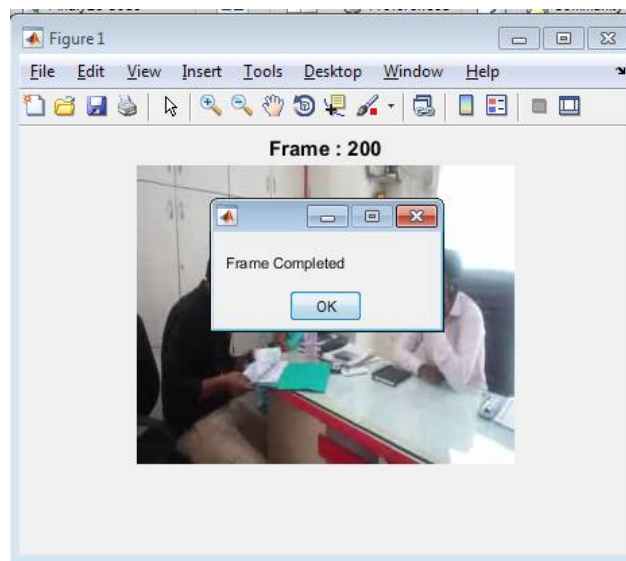
## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

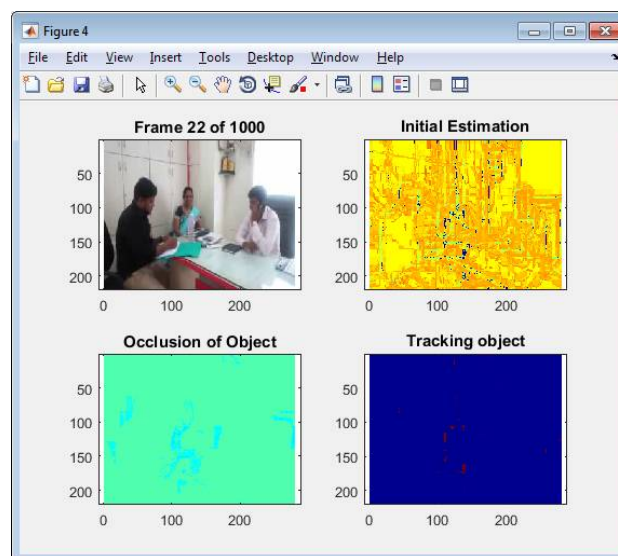
Vol. 8, Issue 5, May 2019

The following figure illustrates the frame estimation process completion scenario and estimate the number of frames and portrait it on the top of the figure.



**Fig.4 Frame Estimations**

The following figure illustrates the human activity and object tracking nature of the proposed system.



**Fig.5 Human Activity and Object Tracking**

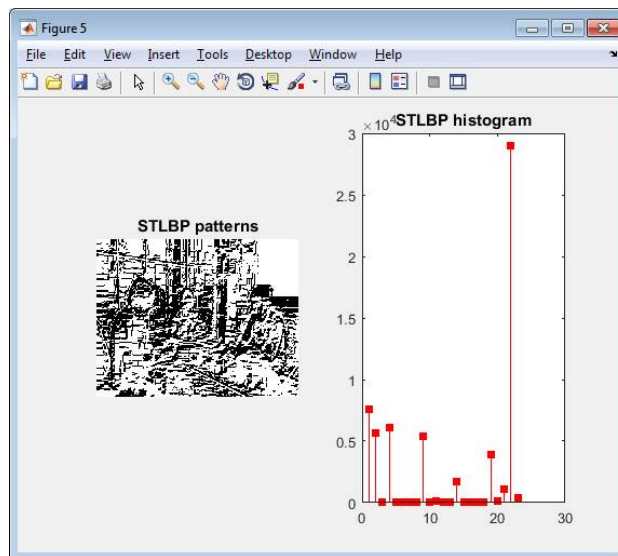
# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

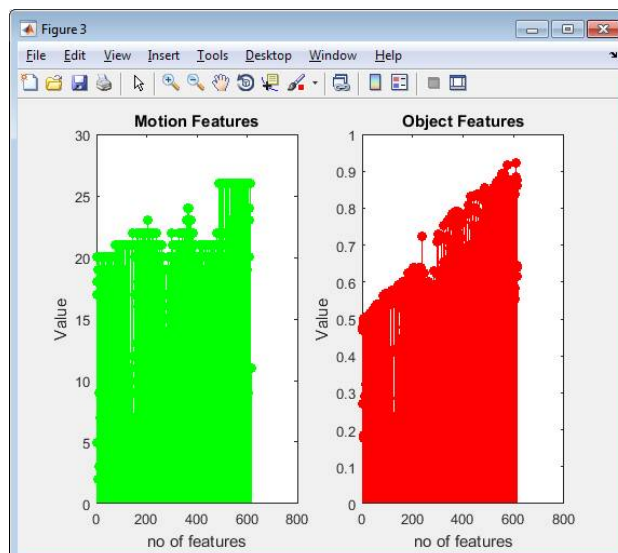
Vol. 8, Issue 5, May 2019

The following figure illustrates the STLBP Pattern and STLBP Histogram view of the proposed system.



**Fig.6 STLBP Pattern and Histogram**

The following figure illustrates the Motion Features and Object Features view of the proposed system.



**Fig.7 Motion and Object Features**



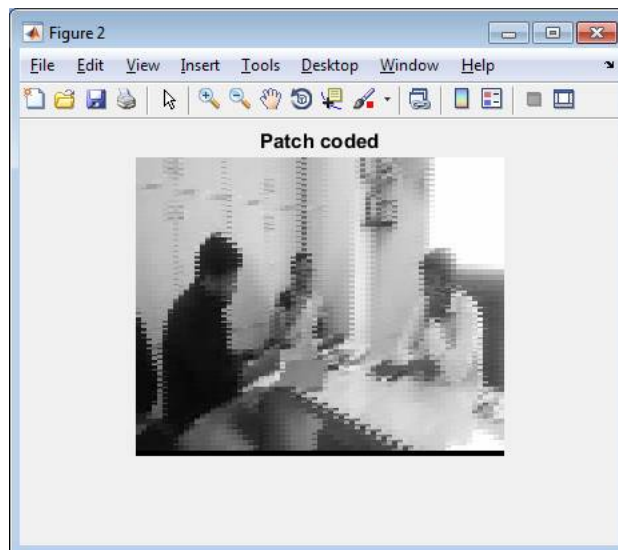
# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

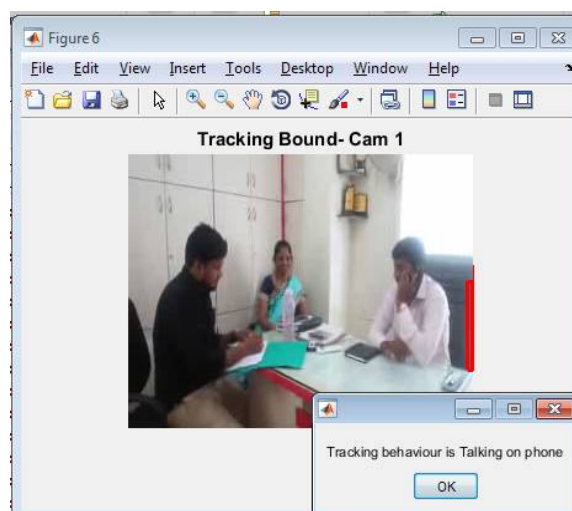
Vol. 8, Issue 5, May 2019

The following figure illustrates the patch coded view of the proposed system.



**Fig.8 Motion and Object Features**

The following figure shows the message popup of identified behavior of proposed system design.



**Fig.9 Identified Behavior**

## VIII. CONCLUSION

Attentive modeling of actions, activities and behaviors is a hot field in computer vision and results obtained up to now could be integrated within our framework with moderate effort. As it has been shown in the experimental analyses, the model offers a probabilistic framework in which it is easy to accommodate a variety of available state-of-the-art attention-based algorithms. Further, note that the Bayesian strategy basically relies on the configurationally complexity  $C(k)(t)$ , which, in turn, is based on spatial entropy. The latter is a mesoscopic quantity that summarizes and

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

can be derived from a variety of "atomic" visual measures. However, from a strict engineering perspective much depends on the specific field of application that is to be addressed. An interesting issue is the applicability of the approach to the case of online multi-camera systems. This case compels to take into account the architectural complexities of the network. The latter can be factored in terms of distribution, mobility and degree of motion of the sensors. As to the distribution issue, the pre-attentive control loop is suitable to be considered for a straightforward fully decentralized implementation, while the attentive loop could be designed at different levels of distribution. Interestingly enough, the ecological Bayesian handoff mechanism is suitable to embed resource-aware conditions, e.g., energy consumption, that is also considered in actual animal foraging. For what concerns the degree of motion of the sensors, clearly, the visual attention rationale that is behind our model calls for considering smart camera networks embedding PTZ cameras that are able to dynamically modify their FOV. In this case, for what concerns single camera activities, techniques developed in the active vision field are apt to be embedded in within-frame analysis either at the pre-attentive or the attentive stage. However, at some point multi-camera activity analysis requires fusing information from multiple camera views. The data fusion problem has not been explicitly considered in this paper. Yet observational data may be combined, or fused, at a variety of levels, again depending on the architecture devised for a specific application.

## REFERENCES

- [1] D. W. Stephens, Foraging theory. Princeton University Press, 1986.
- [2] T. Xiang and S. Gong, "Beyond tracking: Modelling activity and understanding behaviour," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 21–51, 2006.
- [3] T. T. Hills, "Animal foraging and the evolution of goal-directed cognition," *Cognitive Science*, vol. 30, no. 1, pp. 3–41, 2006.
- [4] J. M. Wolfe, "When is it time to move to the next raspberry bush? foraging rules in human visual search," *J. Vis.*, vol. 13, no. 3, p. 10, 2013.
- [5] E. L. Charnov, "Optimal foraging, the marginal value theorem," *Theoretical population biology*, vol. 9, no. 2, pp. 129–136, 1976.
- [6] A. Schütz, D. Braun, and K. Gegenfurtner, "Eye movements and perception: A selective review," *Journal of Vision*, vol. 11, no. 5, 2011.
- [7] J. M. Fuster, "Upper processing stages of the perception–action cycle," *Trends in cognitive sciences*, vol. 8, no. 4, pp. 143–145, 2004.
- [8] "Cortex and memory: emergence of a new paradigm," *J. of Cognitive Neuroscience*, vol. 21, no. 11, pp. 2047–2072, 2009.
- [9] S. Haykin and J. M. Fuster, "On cognitive dynamic systems: Cognitive neuroscience and engineering learning from each other," *Proc. IEEE*, vol. 102, no. 4, pp. 608–628, 2014.
- [10] S. Chiappino, P. Morerio, L. Marcenaro, and C. S. Regazzoni, "Bioinspired relevant interaction modelling in cognitive crowd management," *J. of Ambient Intell. and Humanized Computing*, pp. 1–22, 2014.
- [11] S. Chiappino, L. Marcenaro, P. Morerio, and C. Regazzoni, "Event based switched dynamic bayesian networks for autonomous cognitive crowd monitoring," in *Wide Area Surveillance*. Springer, 2014, pp. 93–122.
- [12] A. Dore, A. F. Cattoni, and C. S. Regazzoni, "Interaction modeling and prediction in smart spaces: a bio-inspired approach based on autobiographical memory," *IEEE Trans. Syst., Man, Cybern. A*, vol. 40, no. 6, pp. 1191–1205, 2010.
- [13] S. Chiappino, L. Marcenaro, and C. Regazzoni, "Selective attention automatic focus for cognitive crowd monitoring," in *10th Int. Conf. Advanced Video Signal Based Surveillance*, Aug 2013, pp. 13–18.
- [14] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, 2013.
- [15] C. De Leo and B. S. Manjunath, "Multicamera video summarization and anomaly detection from activity motifs," *ACM Trans. on Sensor Networks*, vol. 10, no. 2, p. 27, 2014.
- [16] S.-H. Ou, C.-H. Lee, V. Somayazulu, Y.-K. Chen, and S.-Y. Chien, "Online multi-view video summarization for wireless video sensor network," *IEEE J. Select. Topics Signal Processing*, vol. 9, no. 1, pp. 165–179, Feb 2015.
- [17] Y. Cong, J. Yuan, and J. Luo, "Towards scalable summarization of consumer videos via sparse dictionary selection," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 66–75, 2012.
- [18] R. Tron and R. Vidal, "Distributed computer vision algorithms," *IEEE Signal Processing Mag.*, vol. 28, no. 3, pp. 32–45, 2011.
- [19] C. Micheloni, B. Rinner, and G. Foresti, "Video analysis in pan-tilt-zoom camera networks," *IEEE Signal Processing Mag.*, vol. 27, no. 5, pp. 78–90, Sept 2010.
- [20] A. Kamal, C. Ding, A. Morye, J. Farrell, and A. Roy-Chowdhury, "An overview of distributed tracking and control in camera networks," in *Wide Area Surveillance, ser. Augmented Vision and Reality*, V. K. Asari, Ed. Springer Berlin Heidelberg, 2014, vol. 6, pp. 207–234.