# A SYNERGISTIC MODEL FOR REPRESENTING AND INTERPRETING  HUMAN ACTIVITIES AND EVENTS FROM  VIDEO

NIKOLAOS BOURBAKIS [1], JIM R. GATTIKER [2], GEORGE. BEBIS [3]
[1]Wright State University, ITRI, Dayton OH, 45435,
[3]Univ. of Nevada-Reno, CS Dept. Reno, NV 89557,
[2]Los Alamos-national Labs, Los Alamos, NM 87544

This paper describes a new and innovative and approach for representing, recognizing and interpreting human activity from video, contributing to an automated system capable of recognition of complex human behaviors. This technology is directly applicable for monitoring public safety and law enforcement, and capturing of activities is crucial for supporting virtual collaborations between citizens, and between citizens and government. Digital video stores of terabytes are now common, and will continue to increase until they dominate stored data. The government of the future will have to manage, organize, recall, and interpret information from this resource. This paper addresses one important facet of this.

The approach presented here is a model based on the hierarchical synergy of three other models: the Local/Global (L-G) graph, the Stochastic Petri Net (SPN) graph and a neural network (NN) model. The application focus is the description of activity of actors in a video (or multi-sensor) scene, from the snapshot state description through higher levels of organization into events. The concept of importance is the distinction and interaction between *structural* knowledge, or knowledge about physical state, and *functional* knowledge, knowledge about change and events. The L-G graph provides a powerful description of the structural image features presented in an event, and the SPN model offers a description of the functional behavior. The NN (or other adaptive) model provides the capability of learning behavioral patterns for classification of posture and activity, and forecasting possible events in a free environment.

*Keywords: Representation and Interpretation of Activities, Events, L-G graphs, SPN graphs*

## 1. Introduction

In the recent years the scientific accomplishments in image understanding and visual languages fields have been shifted into the video understanding domain. Several research

efforts have been presented in the literature to achieve objects motion in a sequence of images, and some cases basic understanding of human behavior, such as description of a human that carries a box, a human that walks, and simple activity description [1-20]. Specifically, relevant research work has been done on trajectory guided tracking and recognition of actions [14], stochastic temporal models of human activities [13], Bayesian approach to human activity recognition [12], complex visual activity recognition using a temporally ordered database [11], layered probabilistic recognition of human action [10], learning and recognizing human dynamics in video sequences [9], probabilistic recognition of human actions [8], recognizing human action in time-sequential images using hidden Markov models [7], visual understanding of dynamic hand gestures [15] etc. These methodologies, however, provide neither a complete framework for the interpretation of an event, nor means to evaluate the cause of the event.

This paper targets the area of human activity interpretation and recognition with the purpose of developing methods that will isolate and describe activities of interest, while ignoring the vast amount of irrelevant information.

One reason to seek this capability in a digital government is to amplify the capabilities of humans in monitoring and surveillance applications. This assistance can take different forms, such as cueing attention to unusual or proscribed events to allow a human observer to efficiently monitor more information without error through boredom or overwhelming raw data sizes. Robust detection and tracking of humans is a challenging problem due to the changeable morphology of the human body. Complex human movements (e.g. walking, running) can give rise to widely different human appearances. The problem becomes even more difficult when considering multiple moving objects which might occlude or destruct the motion of each other, and complex non-static natural backgrounds. Surveillance can, for example, be for criminal and law enforcement, for watching for needed information or services in e.g. a post office or other public facility, or monitoring for safety reasons in a dangerous public environment such as a subway terminal, or monitoring for safety and safeguards of critical infrastructure such as nuclear facilities or shipping ports.

Another application of this technology in a digital government is the enhanced support of virtual collaborations. The transmission of human posture and activities in a greatly compressed data structure is crucial to the cultivation of collaborative environments. Also, this has important implications to anonymous interaction, where modeled collaborators can be "blinded" by the abstraction of certain specifics, while still project the full range of human expression.

This paper addresses the need to track humans and objects in a visual scene, and assess the activities in the scene. Starting from low-level vision operations of segmentation and contour generation, characterization of the scene objects as graph descriptions will allow recognition and characterization of the objects. From object characterization, a correspondence from structural object features to functional attributes will result in a description of the activity. The model proposed here is based on the implementation of three main components: (1) segmentation and classification of moving objects, (2) tracking and integration of extracted data on humans and human manipulation of objects, and (3) recognition of human activity.

## 2. Region based Fuzzy-like Segmentation

The whole segmentation scheme is shown as Figure 1. It has three stages — smoothing, edge detection and segmentation. The initial smoothing operation is intended to remove noise. The smoother and edge detector algorithms are presented in [21] and not elaborated upon here. Within the segmentation approach, edge information is utilized from a smoothed input image.

The segmentation algorithm uses edge information and the smoothed image to find
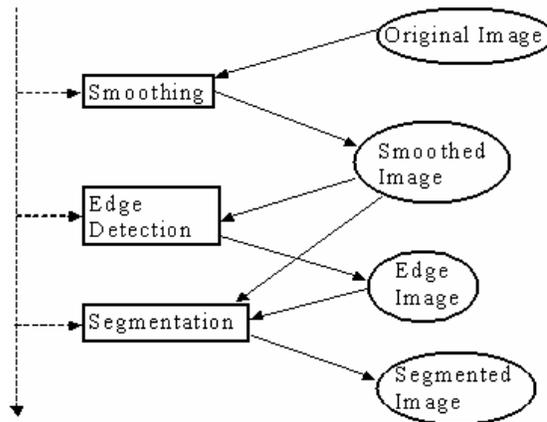


*Figure 1. Image processing steps.*

segments present within the image. The processes involved in this segmentation procedure are listed as follows:

1. Find big and crisp segments.
2. Expand segments based on homogeneity criteria.
3. Expand segments based on dichromatic reflection model.
4. Expand segments based on degree of farness measure.
5. Apply an iterative filter.
6. Find medium size segments.
7. Expand segments using homogeneity criteria and degree of farness.
8. Fill in blank regions.
9. Apply an iterative filter.

It processes big regions first. It expands them based on three criteria: homogeneity, dichromatic reflection model and degree of farness measure. Then it applies an iterative filter and processes the medium size regions next. It follows the same procedure expanding them based on homogeneity and degree of farness criteria. For medium size regions the same procedure is used with the exception that the dichromatic reflection model is not applied.

## 3.  The Synergistic Model

*It is known that an event is the result of a sequence of actions, and this event is characterized both by structural and functional information. The recognition of a human activity is strongly related with the ability of describing and interrelating events.* The capability of describing, analyzing, recognizing events leads to understanding, interpreting, and predicting human activities. A synergistic methodology based on Local/Global (L-G) graphs, Stochastic Petri-nets (SPN) and Neural nets (NN) provides a framework to store and exploit this information, as described in Figure 2.



NN model

isomorphism

SPNG model

mapping

L-G model
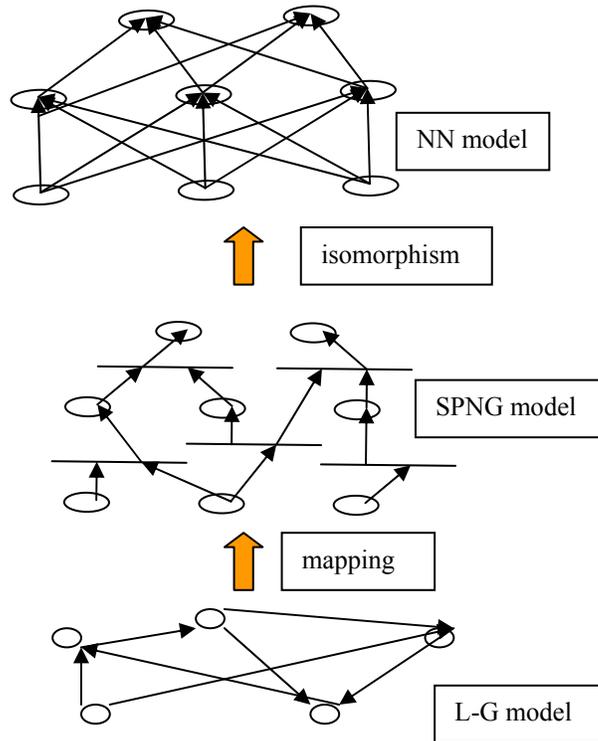
*Fig. 2: The synergistic model*

The L-G graph provides a powerful description of the image structural features presented in an event, the SPN model offers a way to describe functional behavior of the changes or operations, and the NN model provides the capability of extracting and learning behavioral patterns. The framework described organizes these models in a synergistic hierarchical manner, where the advantages of one model at a particular level are

employed by the model of the next level. More specifically, the L-G graph maps its structural descriptions into the SPN model by converting it into a Stochastic Petri-net graph (SPNG) that holds both structural and functional representations. Due to high complexity (a great volume of states for each object) at the SPNG level, a search process becomes difficult and time consuming. Thus, by mapping the SPNG features into a NN model the complexity problem becomes tractable based on the ability of the neural nets to generalize and abstract, providing relevant, though difficult to interpret, models of behavioral categories and predictions. *Note that some important features of this synergistic model are the ability of the neural part to learn certain sequences of actions and predict the output before their completion, also the ability to analyze a sequence of actions and determine the possible "causes".*

## 3. The L-G Graph

The L-G graph is capable of describing with accuracy and robustness the features contained in an image. The local (LG) region graph, the skeleton graph (SG) and the global (GG) image graph components are described below.

### 3.1. The Region or Local Graph

A Local graph G holds information of a contour –line of an image region after segmentation.

$$G = N_1 a^c_{12} N_2 a^c_{23} N_3 \ldots N_k a^c_{k1} N_1 \otimes N_i a^p_{ij} N_j \otimes \ldots \otimes N_n a^{rd}_{nm} N_m \ldots$$

where, $\otimes$ represents the graph relationship operator, and each Ni maintains the structural features of the corresponding line segment, thus, Ni = { sp, orientation (o), length (le), curvature (cu)}, and $a_{ij}$ holds the relationships among these line segments, thus, $a_{ij}$ = { connectivity (c ) , parallelism (p) , symmetry (s), relative magnitude (rm), relative distance (rd), etc}.

### 3.2. The Skeleton Graph [22]

The skeleton graph is based on the efficient generation of the regions skeletons after the regions segmentation process. The line segments of the skeleton of a region are interrelated with each other through a graph with attributes in a similar way with the one generated by the contour of a segmented region. Thus, the skeleton graph of a region's skeleton is

$$G = K_1 a^c_{12} K_2 a^c_{23} K_3 \ldots K_k a^c_{kq} K_q \otimes K_i a^p_{ij} K_j \otimes \ldots \otimes K_n a^{rd}_{nm} K_m \ldots$$

where, $\otimes$ represents the graph relationship operator, and each Ki maintains the structural features of the corresponding line segment, thus, Ki = { sp, orientation (o), length (le), curvature (cu)}, and $a_{ij}$ holds the relationships among these line segments, thus, $a_{ij}$ = { connectivity (c ) , parallelism (p) , symmetry (s), relative magnitude (rm), relative distance (rd), etc}.

The missing elements for a global visual perception of an image are: the color (or texture) of each region, its relative geographic location (distance and angle) among the other regions, its relative size in regards with the other regions, etc. One way to obtain these additional features is the development of the global image graph GG.

*3.3. The Image Global Graph*

The global image graph attempts to emulate a human-like understanding by developing global topological relationships among regions and objects. More specifically, for each image region $M_i$, a skeletonization task is performed and the final center of gravity $GCg(i,x,y)$ is defined [23].

When all the final centers of gravity have been defined for every image region, the global image graph is developed :

$$GG(Ak) = (P_1R_{12}P_2) \ \Phi_{23} \ (P_1R_{13}P_3) \ \dots \ (P_1R_{1n-1}P_{n-1}) \ \Phi_{n-1n} \ (P_1R_{1n}P_n)$$

where $P_i$ is a node that represents an image region graph, its color, and its $GCg(i,x,y)$, Rij represents the relative distance between two consecutive Gcg, and the orientation of each dg, $\Phi$ij represents the relative angle between consecutive distances dg(i) and dg(j).

An important feature of the L-G graph is its ability to describe 3-D scenes. The only difference between 2-D from 3-D is that in 3-D the local graph will represent 3-D surfaces and the global graph will appropriately interrelate them .

## 4. The SPNG Model

The graph models described above have the capability of holding structural information about targets. The functional behavior of a target is described by the states in which a particular target could be changed to after an appropriate trigger state is satisfied. A successful and powerful model capable of describing (or modeling) the functional behavior of a system is the Stochastic Petri-net (SPN). Thus, in order to maintain the structural features of the graph model and the functional features of the SPN model, a mapping is presented here, where the SPN model is transformed into a SPN graph model as follows:

$$m : G \rightarrow SPNG$$

where, $\{N_i\} \rightarrow \{P_i\}$, graph-nodes correspond into SPN places , and $\{a_{ij}\} \rightarrow \{t_{ij}\}$, relationships corresponds into SPN transitions.

*Example: See Figure 4. The SPNG illustrates a target that has four different states (Places $P_i$, i=1,2,3,4). Each place $P_i$ has its own structural features transferred from the corresponding graph node $N_i$. The transitions $t_{14}$ and $t_{43}$ represent relationships among the same parts of a target and a stochastic distribution of time required to fire that transition. The $t_{21}$ transition requires no time to fire.*
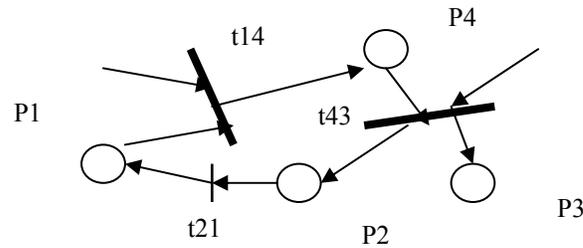


Figure 4. An example SPNG model.

**5. The Isomorphism of SPNG onto NN**

The SPNG model has to deal with a great volume of information due to many functional states associated with each object. Thus, the interpolation and correlation of such amount of information becomes difficult and time consuming. An alternative solution is the use of a neural network model. The neural model alone, however, is not able to efficiently deal with structural representation of knowledge, particularly in a way that can be developed and/or understood by humans. A solution proposed here is the transferring of the SPNG structural and functional features into a NN via an isomorphism: $I : SPNG \rightarrow$ NN, where the places $P_i \rightarrow n_i$, where $n_i$ is a neuron-nodes of a NN, and transitions $t_{ij} \rightarrow w_i$, where $w_i$ represents a threshold logic net). Although this isomorphism contains no mapping to a known learning structure, it demonstrates the capability of neural-like models to contain the appropriate information.

**6. Video Understanding System**

*6.1.Multiple Cameras*
 To implement a system that understands human activity, the viewing system should be able in general to image the tracked human(s) in a broad area over a period of time that is long compared to the events of interest. Using a fixed single camera has limitations since it restricts tracking to a very narrow area due to the restricted viewing angle of the system. A moving camera with some degree of rotational freedom increases capability in a large environment, but complicates the implementation by adding the motion estimation of both the viewing system and the subject of interest. In the system described here, multiple fixed cameras mounted in the area of interest to track and monitor the motion of individuals in sequences of monocular images are used. Since occlusion is view angle specific, multiple cameras will reduce the chance the occlusion is present in all views. Multiple cameras will also alleviate the difficulty when certain views are confused. To improve the quality of the segmentation and tracking, the system will also use information from multiple calibrated cameras. Although the fusion of information from multiple cameras will improve the segmentation and tracking, there will still be expected levels of error and low confidence observations. To further improve the results, the system to be implemented will not be strictly feed-forward, from low-level operators to high level recognition processes, but will rely feedback from the recognition process and the casting of an activity in larger context.

*6.2. Recognition and Learning*
We have previously described how specific procedural activities can be incorporated into our system for activity recognition in scene situation analysis, as well as the assessment of the results of activity in constrained situations. The system has also the capability of this framework to learn normal and abnormal patterns from the data. Given a time-sequence of humans, their tracks, and a description of their position and activities, we use an neural adaptive model to characterize and learn normal behaviors, and thus also recognize anomalous or unusual behaviors, or identify behaviors like a set specified.

Starting from a tracking record of humans and their activities normal profiles can be built up at various levels of description complexity, from simple path tracking based on Markov Models, or SPNG, to complex characterizations of sequential activities using time-series and spatial descriptions of arbitrary complexity, such as in neural models. These models can then be used to isolate unusual, suspicious, or alarming behavior from the identified scenes. The isomorphism between these descriptions learned dynamically in a neural formalism, and the structural and functional descriptions isolated through the scene recognition provides the crucial link for combining learning capabilities of the adaptive systems with the human and object description structures that developed for underlying scene feature recognition.

## 7. Methods and Tasks

### 7.1. Detection of Moving Objects
The system is initialized by acquiring measurements of the scene over a number of video frames. The goal is to build a powerful representation of the background which will allow us to extract the moving objects in the scene more robustly. Towards this objective, we propose using an eigenspace representation of the background. What is interesting about this representation is that when an image containing new objects is projected onto the background eigenspace, then its reconstructed counterpart does not contain the new objects anymore. This representation needs to be updated over time to account for changes in lighting conditions and new objects in the scene. In a long term, this eigenspace will describe the range background appearances that have been observed. The key to using the eigenspace background representation is on our ability to update it efficiently over time. We deploy recent results in numerical mathematics (recursive Singular Decomposition techniques) to implement incremental approaches that allow us to update the eigenspace of the background in real time. To further improve the segmentation results, we will use (i) fusion from segmentations obtained using other viewpoints (multiple cameras), (ii) information from the tracking component (predicted locations), and (iii) feedback from the high-level human activity recognition component.

### 7.2. Target Detection, Extraction and Representation
Connected components analysis, and morphological operations and regions model-based synthesis are used to improve segmentation quality. Each segmented object will then be divided into a number of regions which will be used for building its L-G graph. Specifically, each human will be divided into a number of regions using color, texture, and motion information. Each region will be represented by a mixture of Gaussian learned by using the EM algorithm. In this task the automated detection, extraction and representation in graph forms of a variety of targets, such as human, objects, or other actors, will be performed. In particular, each segmented image frame will be described by the Local-Global (L-G) graph model. The L-G graph model provides a flexible representation of the targets and their surrounding in an image frame independent from rotation and shifting. The target detection problem presents two possible scenarios: a) the targets are known to the system, and b) the targets are unknown. For the first scenario, the target is known to the system, it means that its L-G model is available. Thus, the system searches for the target L-G graph form at the image L-G graph representation by

using a fuzzy like matching algorithm. When a sub-L-G graph (from the image L-G graph) matches the targets specifications, then that sub-graph is extracted and temporary saved for the next processing task described below. In case that the target is unknown, then the L-G graph of the image frame is exhaustively searched and all possible recognizable or desirable sub-L-G graph forms are extracted and interrelated with their surrounding are saved for further use in later frames.

### 7.3. Tracking of Interesting Moving Objects

A common approach to tracking non-rigid objects is based on using high-level parametric models representing the various object parts (e.g., legs, arms, trunk, head etc in the case of human tracking) and their connections to each other. However, these methods are difficult to apply in real-world scenes due to the difficulty of acquiring and tracking the requisite model parts (e.g., specific joints such as knees, elbows or ankles in human tracking). Another problem with this approach is that a separate model is required for each type of objects to be tracked (e.g., humans, animals, etc.). In this project, we propose to build dynamic models of appearance of the objects being tracked. These models will enable robust tracking but at the same time provide useful information for classifying moving objects as rigid or non-rigid. The key to building dynamic models of appearance lies on using an incremental eigenspace approach such as the one described earlier for background modeling. Using multiple cameras will facilitate this since different views of the object will be obtained quickly. To build the eigenspace of an object, the segmented moving object will be resampled in a canonical frame throughout tracking, despite changes in scale and position. Its estimated bounding box will be used to resize and resample it into a canonical view. To classify the object as rigid vs. non-rigid, the distances of the views used to update the eigenspace will be analyzed (non-rigid objects are expected to create widely different views, thus, the distances will be large compared to the ones obtained when building the eigenspace of rigid objects). To build the larger object graph, the system searches for the sub-L-G graph form of the same target. When the sub-L-G graph is discovered, the two graphs are connected by maintaining the relationships between the sub-graphs of the same in consecutive frames. The same process is repeated and the output is the tracking path of a target in a sequence of video frames. When a target is detected in a specific frame, the L-G graph provides the capability to interrelate each recognizable target with any other target in the same frame.

### 7.4. Representation of Actions and Events

*Definition:* An action $A_m$ represents the mapping from a state $(S(i,t) \rightarrow S(j,t') : A_m : \Sigma x S \rightarrow S$, where $\Sigma$ represents the set of actions and S the set of states of a target. Thus, an action $A_m$ could be described as the SPNG equivalent to $G_f$ graph that interrelates the same target into different (or consecutive) frames.

*Definition:* An event $E_i (T_i, T_j)$ between two targets $T_i$, and $T_j$ is the result of a set of actions $A_m$ executed by a certain order on these targets.

Example: Figure 5. Event (*catching a ball; target-1 the hand, target-2 the ball* ), Actions (*the five fingers change status from the "open" state into the "closed" state due to coming ball in a synchronized manner described by the SPN*).
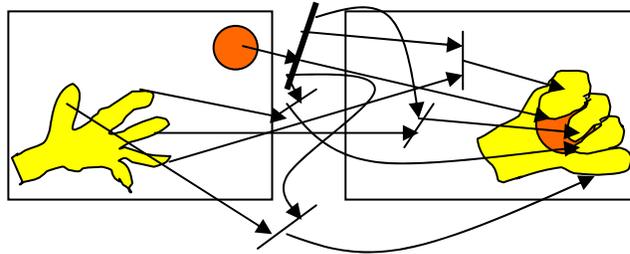
Figure 5. An example of a catching event using an SPNG for two consecutive frames:  A human hand. and a ball : a) Frame –1: the hand in the open state and a ball in the area coming to its direction;  b) Frame-2 : the hand into a closed state catching the ball. The ball plays the role of the triggering action to make the transition of the targets parts (fingers) from one state into another.

*7.5. Actions Recognition, Events Interpretation and Activity Recognition-Understanding*
The recognition of an action, based on its definition, is based on the efficient representation of the SPNGs associated with the states and transitions involved with the particular target. The recognition process is actually a matching of the states, transitions and their order of appearance against the order of target-states and transitions available in a database. The interpretation of an event is actually the recognition of the sequence of actions involved in an event. This means that certain sequences of actions will represent events in a database, thus, when a sequence of actions has been extracted from a video, the sequence goes against the sequences available in the events database. The human activity is defined as a sequence of actions and a set of events. Thus, the recognition of the human activity will be the recognition of the actions and events involved within. Here the system is appropriately trained with a variety of activities before its use in real examples.

*7.6. Scenarios for Demonstrating the Model*
The synergistic model described here is used both in controlled and uncontrolled environments. Surveillance in controlled environments, which have specific rules that can be monitored and enforced, has two distinct aspects. The first is surveillance in a highly controlled environment, in which specific access and procedural rules must be followed, for security and safety reasons. These include identification, authorization, and access rules; two-person and other supervision rules; and rules of physical procedure when using sensitive and/or hazardous materials. The second aspect of controlled environment surveillance is the characterization of general, or even individual, behavior to monitor for anomalous scenarios. In a sensitive facility, this is a way to address the insider threat. Many cases of insider compromise have been characterized by markedly unusual behavior that went unrecognized, from such gross indicators as unusual work schedule changes and after-hours access, to more subtle indicators such as unusual work habits,

paths, or access. Although it is not being suggested that every example of anomalous behavior would indicate a significant threat, we know that it is possible, as demonstrated by human surveillance in e.g. department stores, to detect certain kinds of unusual behavior that should be alerted. Adaptively learning behavior and calling attention to unusual circumstances for review by the facility monitoring personnel is a complex and very important aspect of this surveillance application. A key application for surveillance in uncontrolled environments, where there are no specific rules for activities to follow, is embodied in monitoring airports for alarming behavior.

*7.7. An Illustrative Simple Example*
In this section we provide an illustrative simple example to present the potential of the methodology described here. In particular, three consecutive image frames were used, see Figure 6. In these synthetic frames, the segmented human is synthesized by using the model based region synthesis approach. The sequence of changes (high-level) of the states of his body parts (head, (blue), arm (red), hand (green), and legs (yellow)). In particularly, from each image frame the main parts of the segmented human body are extracted, represented and interrelated using the L-G graph. At this point, the SPNG connects these parts in consecutive frames to determine the changes of the states that take place. Between the first and the second frame, the token activates or causes the activation of the head to "look at" a certain direction, at the same time the arm is moving up and the legs change position. The SPNG graphically presents and connects these changes. The same happens between the second and the third frames, where the SPNG connects the body parts using the affect of the previous token. The important issue here is that the SPNG provides the ability of synchronizing the actions performed by each body part. In this particular example the NN model has no visual contribution since the involvement of learning cannot be shown from one example.
This is a highly simplified demonstration that does not reveal the complexity of matching these states and inferences to the extracted data. The management of this complexity in the knowledge bases and matching tools is a challenge that is in progress in the project.


## 8. Conclusion

This paper presented the overall synergistic model needed for the representation and interpretation of human behavior (activity and events). In particular, it has shown the models (fuzzy-like segmentation of the images, L-G graph for representing objects and regions belonging to the same object, SPNG for representing the functional behavior of objects, the NN was showing here since has no visual representation) used in a synergistic way for achieving the representation, recognition and interpretation of human behavior (activity, events). The simulated simple example was provided to demonstrate the potential of the model. The multiple cameras description of the system was given here mainly to show potential applications of the model. The synthesis of the regions that belong to the same object and the synthesis of the multiple views of moving targets (or objects) has been implemented in [24].
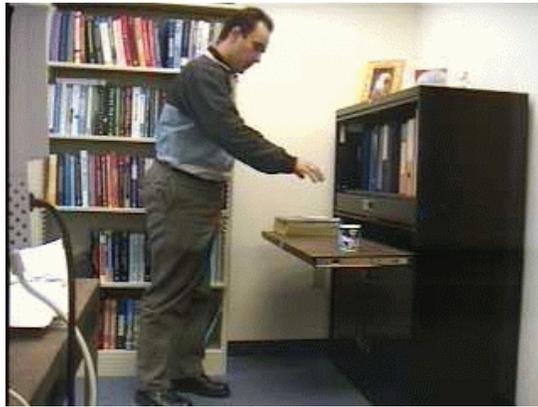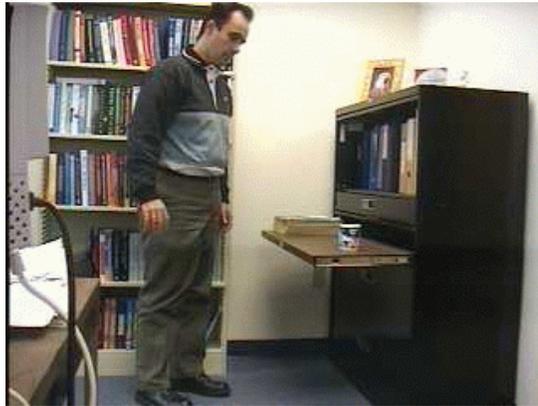
Fig.6a: The original frames

token
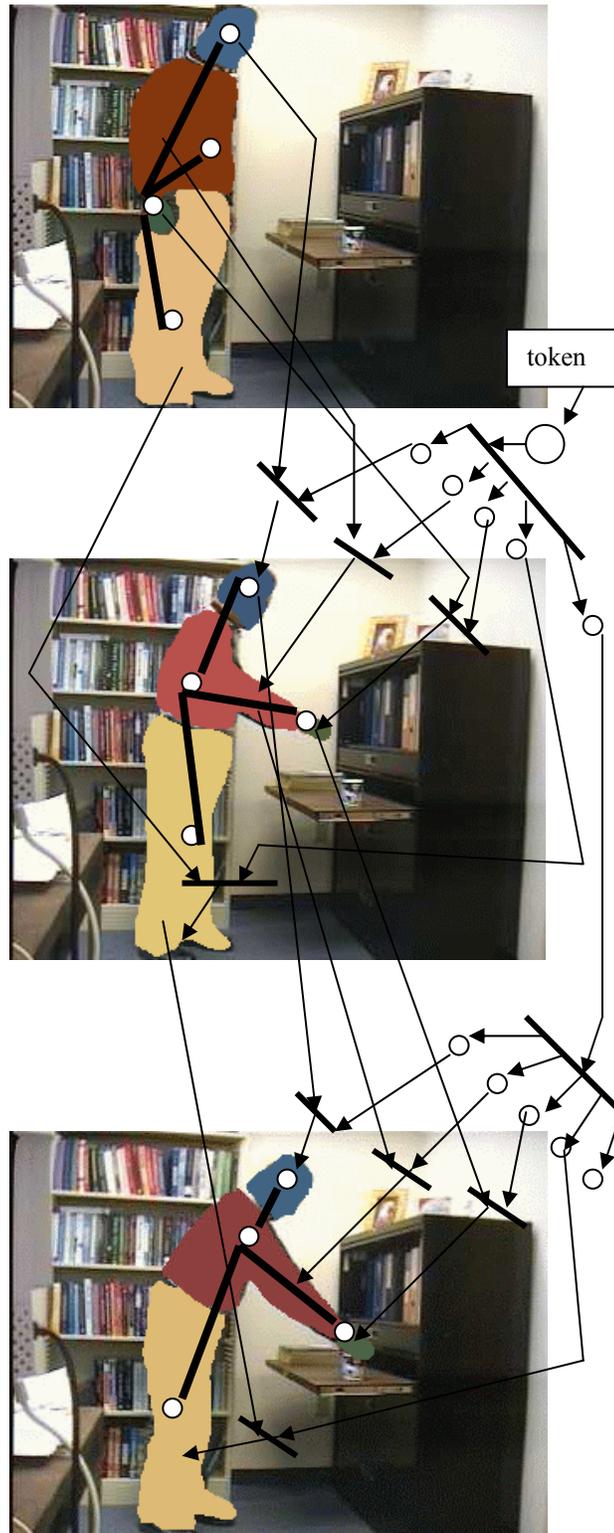
Fig. 6b: The Segmented frames with the L-G and SPN  visual representations

token
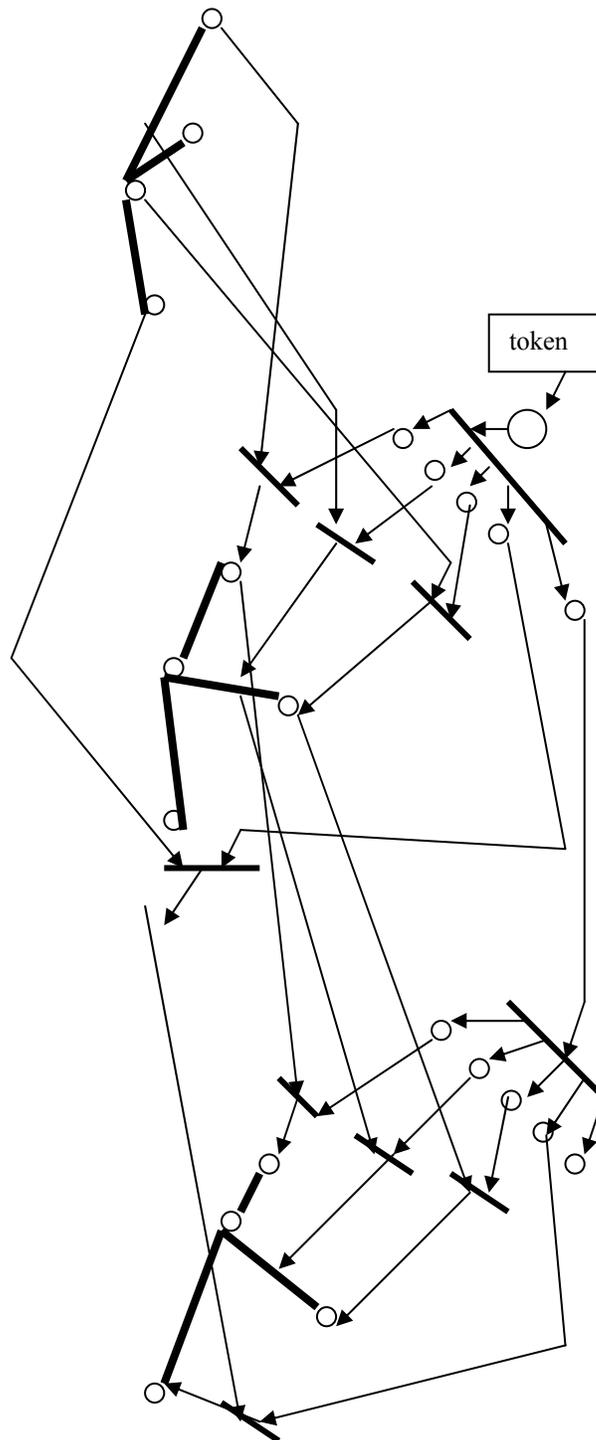
Fig. 6c: The L-G graphs  and the SPNG graphs are represented by  with thick and thin lines respectively.
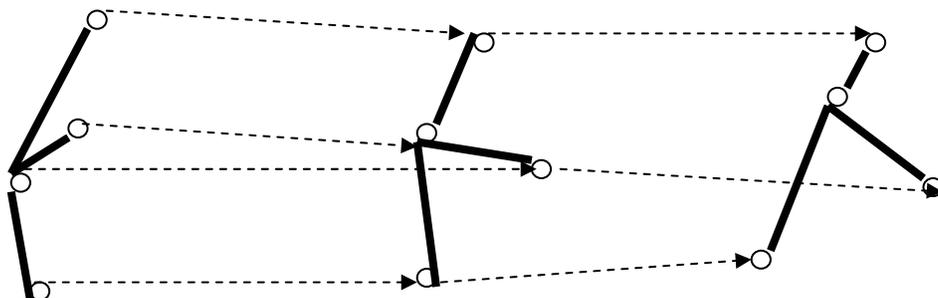
Fig. 6d. The connection of three consecutive movements (actions) of a human body based on the L-G graph representation. Note here that the center of gravity are located on the skeleton of each human standing. These center of gravity may represent regions of connection points of singularities.

## References

[01] N. Bourbakis, A Neural-based KB using SPNGs in sequence of images, AIRFORCE & SUNY-B-TR-1991,45 pages, Nov. 1991.

[02] J.Gattiker& N.Bourbakis  Representation of Structural and Functional Knowledge using SPN Graphs, Proc. IEEE Conf. on SEKE 1995, MD.

[03]N. Bourbakis, HVP: Measuring Human Perception using the L-G graph, SUNY-B-TR- 1997.

[04]  N. Bourbakis and R.Andel, ATR from sequence of images Proc. ICTAI, CA,1997

[05]  N. Bourbakis, Multimedia Information Systems using SPNGs, ACM Conf. LasVegas, 1997, invited talk.

[06]  N. Bourbakis, Facial expressions using L-G&SPNGs , SUNY-B-TR-1997

[07] J.Yamato, J.Ohya and Kenichiro, "Recognizing human action in time-sequential images using Hidden Markov Model", IEEE Conf CV. 1992, pp.379-385

[08] C.Bregler, "Probabilistic Recognition of Human Actions", UCB-TR-May-1996,p.28

[09] C.Bregler, "Learning and Recognizing Human Dynamics in video sequences" IEEE Conf. on CVPR, Puerto Rico, June 1997

[10]  J.Feldman and C.Bregler,MICRO-TR-96-103, 1996-97.

[11] S.Bhonsle, A.Gupta, S.Santini, M.Worring and R.Jain, "Complex visual Activity Recognition using a temporally ordered DB", UCSD-TR-1998

[12]  A.Madabhushi and JKAggarwal, "Bayesian approach to Human Activity Recognition", UT-Austin, TR-97-ARP-275

[13] M.Walter, S.Gong and A.Petrou," Stochastic Temporal Models of Human Activity", UW-TR-UK-1998

[14]  R.Rosales and S.Sclaroff, "Trajectory Guided Tracking and Recognition of Actions", BU-CS-TR-99-002, p.29

[15] M.Yeasin and S.Chaudhuri, "Visual understanding of dynamic hand gestures", PR Journal 33,11,2000, 1805-1817.

[16] N.Bourbakis, IJAIT ATR

[17] G. Bebis, S. Uthiram, and M. Georgiopoulos, " Face Detection and Verification Using Genetic Search", International Journal on Artificial Intelligence Tools, vol 9, no 2, pp. 225-246, 2000.

[18]G. Bebis and K. Fujimura, "An Eigen-space Approach to Eye-Gaze Estimation", ISCA 13th International Conference on Parallel and Distributed Computing Systems, pp. 604-609, Las Vegas, 2000.

[19] N.Bourbakis, J.Gattiker and G.Bebis,A model for interpreting human activities and events from video,  IEEE Conf. on TAI-2000, Vancuver, Canada, pp.132-139.

[20] N.Bourbakis, EIRMA: a formal  language for representation and interpretation of activities and events based on LG graphs, IEEE SMC, sub.

[21] P.Yaun, X.Yuan, A.Mogzadeh, D.Goldman and N.Bourbakis  A fuzzy-like approach to edge detection in colored images, IAPR Pattern Analysis and Applications, vol.4,4,272-282,2001

[22] N.Bourbakis and D.Goldman, Recognition and acquisition of digital line segments with unevenness with applications to recognition of handwritten characters and fingerprints, IFAC IJEAAI,12,273-279,1999

[23] N.Bourbakis, Emulating human visual perception for measuring differences in images using an SPN graph approach, IEEE T-SMC, 32,2,191-201, 2002

[24] N.Bourbakis and P.Yuan, Representing and recognizing 3-D objects in images using LG, IEEE T-SMC, sub.