Monitoring *Trichogramma* Activities From Videos

An adaptation of cognitive vision system to biology field

Mohamed Becha Kaaniche, Francois Bremond, Monique Thonnat
INRIA, Sophia-Antipolis Research Unit, Orion Project
2004 route des Lucioles B.P. 93,
06902 Sophia-Antipolis Cedex, France
{Mohammed-Becha.Kaaniche, Francois.Bremond, Monique.Thonnat}@sophia.inria.fr

Abstract

This paper describes a new adaptation of a cognitive vision system in biology field. Our goal is to monitor Trichogramma activities from image sequence. More precisely we aim at recognizing several scenarios of the parasitoid behavior when it parasites the eggs of the Mediterranean flour moth Ephestia kuehniella Zeller (crops-harmful butterfly). The main advantage of our current work is the use of accurate features for characterizing mobile objects shapes. We first establish different requirements to describe Trichogramma activities involved in the laboratory experiment context. We second explain the steps taken to adapt a general video understanding framework containing a scenario recognition algorithm to reach our objective.

1. Introduction

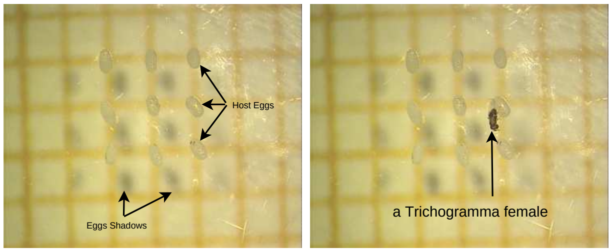
Activity monitoring is an important field of research but it usually deals with human activities. We are interested in applying it to the biology field. *Trichogramma* species are recognized as important biological control agents to substitute pesticides in field crops, forests and fruits. As a parasitoid of caterpillar pests, it protects several vegetables such as corn, rice and sugarcane. Current studies are focused on analyzing the variations of handling-time and on understanding their foraging mechanisms for screening better agents for biological control and improving their efficiency to control their hosts when they are released in the field. To conduct this works, it is essential to understand the parasitoid behavior. Currently, the video sequences of laboratory experiments (see Fig. 1 for sample) are analyzed manually by experts. Due to the hugeness of the data to collect from plentiful video sequences, we are interested on automating this task by using scenario recognition techniques.

Unfortunately, video understanding is focused generally on recognizing human activities. Furthermore, behaviors are usually recognized through the study of trajectories and positions of studied objects and using a priori knowledge about the scene. This is quite sufficient when we deal with scenes having large field of view and simple human activities. But, we often do not have enough information to accurately determine more complex behaviors and especially non-human activities. Moreover, we need to process visual features characterizing the shape of the mobile object so that we can identify its behavior.

Our aim is to adapt an automatic video interpretation system to recognize *Trichogramma* behaviors which is composed of a vision module and a scenario recognition module. The system takes two types of inputs: (1) a video stream acquired by camera(s) and (2) a priori knowledge concerning scenario models predefined by experts and the 3D geometric and semantic information of the observed environment. The output of the system is the set of recognized scenarios at each instant.

In general words, the system performs through three steps. First, a low-level image processing algorithm subtracts the current frame with the background frame and detects moving regions. Then a tracking algorithm tracks the detected regions and computes their trajectory. Finally, the scenario recognition module identifies the tracked moving regions as mobile objects and interprets the scenarios that are relative to their behaviors.

To reach our goal, we use VSIP (Video Surveillance Intelligent Platform), described in [1], including a scenario recognition module based on [15]. We also integrate a module to handle *Trichogramma* specific behaviors. In this paper we focus on the wrapper module and on the set of parameters required for tuning the vision module. We first present the state of art in the related fields. Second, we introduce the proposed method. Third, we show the early results and discuss them. Finally, we conclude by summarizing the contributions and exposing our future work.



(a) Background with host eggs (b) *Trichogramma* in drilling

Figure 1: The figure shows the background of a sample image sequence and a sample frame from the video. Image (a) illustrates the nine hosts eggs and their shadows. Image (b) illustrates a *Trichogramma* female in drilling phase.

2. State of Art

2.1. Shape Features

There are two main categories of methods studying the visual features of the shape of a mobile object: methods based on 2D appearances and methods using 3D models. The majority of the methods based on 2D appearance use the same schema. First they detect the principal extremity parts of the body. Then, based on these detections, they search for the secondary parts of the body which are the articulations. The method proposed by [9] proceeds in three steps: determining the center of gravity of the human silhouette, computing the orientation of the upper half of the body and finally estimating the significant points such as feet, hands, elbows and knees by using a heuristic contour analysis of the human silhouette. Reference [18] is a real time system which uses a multi-class statistical model of color and shape to obtain a 2D representation of head and hands in a wide range of viewing conditions.

Another approach, described in [6], segments the silhouette from the background and computes the vertical and horizontal projections of the silhouette to determine the global posture of a person (standing, sitting, crawling-bending and laying) and his/her orientation relative to the camera (front view, left side view and right side view). To recognize the posture the system compares the current projections to the model of projections realized for a set of predefined posture and point of view. Then it determines the body parts by analyzing the contour of the silhouette. The principal drawbacks of these methods are the dependency on the point of view and their focus mostly on human-shapes.

Concerning methods using 3D models, a model is defined as a set of geometrical objects such as parallelepipeds, spheres or truncated cones and parameters which define the relations between these objects such as spatial constraints. This model is defined in a high dimensional phase space (dimension depends on the degrees of freedom of the model). Several pre-established and known 3D Models are used as a priori knowledge in such methods. In [11], an alternative representation of the phase space which supports a more efficient use of the different constraints is proposed. This method needs a frontoparallel view of the camera.

To handle ambiguities in such methods and to estimate precisely the depth, several cameras may be used. For instance, [10] proposes a system using three cameras all around the person. The method compares the projections of 3D model of a person on an image with the detected silhouettes. This process is iterated by computing a force that will move the 3D model towards the current silhouette. The final parameters of the 3D model constitute an estimation of the real posture. The main drawbacks of these methods are the utilization of too many parameters difficult to tune and the requirement of several cameras to obtain a better precision.

In this work, we combine the 2D methods described in [9] and [6] to extract features characterizing a non-human shape. Methos using 3D models are ill-adapted since only one camera has been provided in the experiment and no *Trichogramma* 3D model is available for the current time.

2.2. Scenario Recognition

Automatic Video Interpretation has been a problem of focus in cognitive vision in the last decade. The last years witnessed a more practical and user-centered development of this field. However, it remains focusing on people and vehicle tracking. Through several approaches which aim at recognizing temporal scenarios we can retain two main categories based on (1) a probabilistic/neural network combining potentially recognized scenarios and (2) a symbolic network.

As a computer vision researcher, the probabilistic/neural network [12] is an adequate approach. The nodes of such network correspond usually to scenarios that are recognized at a given instant with a computed probability. For example, [8] proposed an approach to recognize a scenario based on a neural network (time delay Radial Basis Function). Reference [7] proposed a scenario recognition method that uses concurrence Bayesian threads to estimate the likelihood of potential scenarios. These probabilistic methods are useful, in the case of noisy images, to give an interpretation of the scene while taking into account the stochastic variations of the analysis. For our work, we choose classical filtering techniques to get rid off these variations and to obtain coherent data that can be associated with symbolic values.

In artificial intelligence, researchers use symbolic networks; their nodes correspond usually to the boolean recognition of scenarios. For example, [14] used a declarative representation of scenarios defined as a set of spatiotemporal and logical constraints. To reduce the processing time for the recognition step, they proposed to check the consistency of the constraint network using the AC4 algorithm.

More recently, [4] defined a method to recognize a scenario based on a fuzzy temporal logic. Concurrently, [16] present an approach to optimize the temporal constraint resolution by ordering in time the sub-scenarios of the scenario to be recognized. The common characteristic of these approaches is to store all totally recognized scenarios (recognized in the past).

Another approach consists in using a symbolic network and to store partially recognized scenarios (to be recognized in the future). For example, [5] has expressed a temporal scenario as a chronicle which is represented as a set of temporal constraints on time-stamped events. The recognition algorithm keeps and updates partial recognition of scenarios using the propagation of temporal constraints based on the RETE algorithm. The applications are dedicated to the control of turbines and telephonic networks. Reference [3] made an adaptation of temporal constraints propagation for video surveillance. In the same period, [13] have used Allen's interval algebra to represent scenarios and have presented a specific algorithm to reduce its complexity.

Therefore, we extend the method shown in [15] by adapting the ontology used by the scenario recognition algorithm as described in [2] in order to recognize *Trichogramma* behaviors. Our choice is justified by the ability of this method to limit the amount of scenarios stored by symbolic network based methods.

3. Proposed Method

In VSIP, an image processing module detects the moving regions and computes several measures (e.g. their size, location). Then a tracking module tracks the detected regions within the whole video. Because of detection errors, a moving region can either correspond to a noise, to a part of a mobile object (e.g. part of the wings of a *Trichogramma*) or to a mobile object (e.g. a *Trichogramma*). Then the scenario recognition module generates hypothesis to consider the tracked moving regions as a mobile object composed of one or more regions. Finally the scenario recognition module computes the properties of mobile objects and analyzes the scenarios relative to the behavior of the mobile objects based on their properties. Scenarios are recognized in a recursive manner. At level 0, an elementary scenario (i.e. primitive state or event) is recognized directly from the associate mobile object properties. At level n, a scenario is recognized through a combination of sub-scenarios recognized at level n-1. To describe a scenario we use a modeling formalism as proposed by [2].

Before the current work, the system could only handle human activities and detect only human and vehicles. Up to eight mobile object properties were computed: height, width, speed, motion direction, current location, trajectory (set of previous location), the distance to a reference object

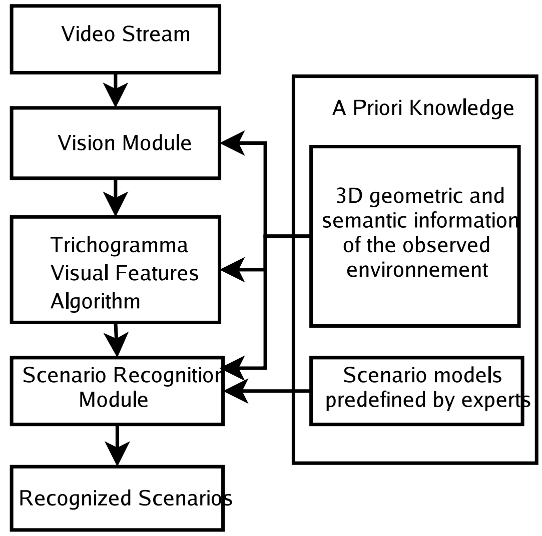


Figure 2: Flowchart of the proposed Automatic Video Interpretation System.

and the number of moving regions that compose the mobile object. To adapt this video understanding platform to *Trichogramma* behavior recognition, we have performed three tasks. First, we have tuned the vision module parameters and a priori knowledge to adjust the context initialization and the segmentation process. Second, we have designed a visual feature algorithm dedicated to the recognition of parasitoid (such as *Trichogramma*) and to the computation of its specifics properties. Finally, we have extended the scenario recognition module (as shown on Fig. 2) to take into account the visual feature module outputs.

3.1. *Trichogramma* Behavior Description

Studies on parasitoids behaviors are conducted since the sixties. Reference [17] shows one of the researches on *Trichogramma* behavior which analyses the variations of handling-time in the parasitoid activities. The experiment consists in offering groups of host eggs to isolated *Trichogramma* females that one observed under a microscope until all host eggs are parasitized. The output of the microscope is digitalized and stored as a video stream. Our goal is to build the history of six *Trichogramma* activities. First, we detect every entry ("Enter" event) and exit ("Exit" event) in an experimental zone surrounding the host eggs; if the *Trichogramma* female exits from this zone and stays out more than sixty seconds, the tracking is stopped. Then, we recognize the "Walk" event which corresponds to the walking of the *Trichogramma* females in the experimental zone between host eggs. Finally, we focus on the duration and time-bound of the three phases of egg laying behavior. The three phases are: (1) antennal drumming, (2) ovipositor drilling and (3) oviposition.

The first three actions can be easily detected in VSIP because they can be computed using only mobile object position. However, the three phases of egg laying behavior are specific to the *Trichogramma* activities. Thus, local descriptors are needed to identify these three phases. First, we focus on the antenna positions and movements to identify the antennal drumming phase. Second, we detect the quasi-immobility of the *Trichogramma* female during the drilling and the oviposition phases. Finally, we distinguish between these two later phases by detecting the slight vibration of the *Trichogramma* abdomen to insert, completely, its egg-laying organ (i.e. ovipositor) into the host egg.

Nonetheless, due to the delay needed by a *Trichogramma* to be interested in host eggs, the experiment begins with a phase where a glass dish is reversed to prevent *Trichogramma* from escaping. If the *Trichogramma* is not interested in the host eggs, the experiment is stopped. The glass dish reversing phase introduces a background modification in the image sequence, thus the video understanding process starts only after this phase. To extend the automate analysis through the whole video, we need to modify the experiment steps.

3.2. Detection and Tracking of *Trichogramma*

Basically, in case of one static observing camera, the vision module contains four main components: the acquisition component, the context and background component, the segmentation component and the frame to frame (short term) tracking component.

The acquisition component manages the source of the video, the format of the captured image sequence and its frame rate (typically between 4 and 25 frames/second). It provides the current frame of the video to process at the current iteration.

The context and background component detects the host eggs (using contextual information) and the mobile object (using background information) in the current frame. This component was adapted to treat sequences captured from a camera view point enabling the computation of mobile object height. However, in the experiment, the image sequences are captured by a vertical camera with a non-focal lens. So, we have extended this component to handle this situation where only two dimensions can be observed.

The segmentation component detects the moving regions by subtracting the current image from the background image. These moving regions, associated with a set of 2D features like moving pixel density or region position, are called BLOBS. The 2D features correspond to image dimensions measured in pixels. After noise removing, a set of 3D features like 3D position, width and height, are computed for each blob. The 3D features correspond to the physical world dimensions measured, in the current case, in micrometers. The blobs associated with their 3D features are called MOBILE OBJECTS. We have adjusted the segmentation component parameters to adapt the quality of the segmented mobiles in the scene observed through microscope equipment which implies a strong light intensity due to the experimental conditions.

The short-term tracking component links from a frame to the next frame all detected mobile objects. The output of this module is a graph containing the detected mobile objects updated over time and a set of links between blobs detected at time t and blobs at time t-1.

3.3. *Trichogramma* Visual Feature Algorithm

To identify the three phases of the *Trichogramma* egg laying behavior, we need to extract the visual features characterizing the insect shape and interpret them as events. Hence, we have designed an algorithm in order to handle these aspects. The *Trichogramma* Visual Feature Algorithm first classifies the segmented mobile object as a *Trichogramma* female by controlling the already computed features such as width and height. Then, it detects the antennas and the ovipositor of the *Trichogramma*. Finally, it determines their events (e.g. Drumming, Drilling) by computing their positions and monitoring their movements.

The Visual Feature Algorithm classifies the detected mobile object into two classes: *Trichogramma* or unknown. We have defined a specific probabilistic model which learns the 3D size of the insect in a preprocessing phase. *Trichogramma* female is recognized by comparing the current 3D features to the predefined model. Then, the algorithm computes additional visual features to be used by the scenarios recognition module.

In order to detect the parts of the body of the parasitoid, we proceed through five steps. First, we compute the center of gravity of the mobile object recognized as the *Trichogramma* body. Second, we compute its orientation by calculating autocorrelations of pixels through x axis (C_{x^2}) and y axis (C_{y^2}) and their intercorrelation (C_{xy}): the orientation is deduced by applying (1). Third, we normalize the blob corresponding to the parasitoid body with a rotation to position it horizontally in order to obtain a self-referential of the insect. Fourth, we calculate the main axis and orthogonal axis projections of the insect (self-referential) by counting the number of pixels of each axis as shown by Fig. 3.

$$\theta = \arctan\left(\frac{C_{xy}}{C_{x^2} - C_{y^2} + \sqrt{(C_{x^2} - C_{y^2})^2 + C_{xy}^2}}\right) \quad (1)$$

Finally, we determine the length and the position (in the projections) of the antennas and the ovipositor by matching the extremities of the projections with a 3D template corresponding to their average size.

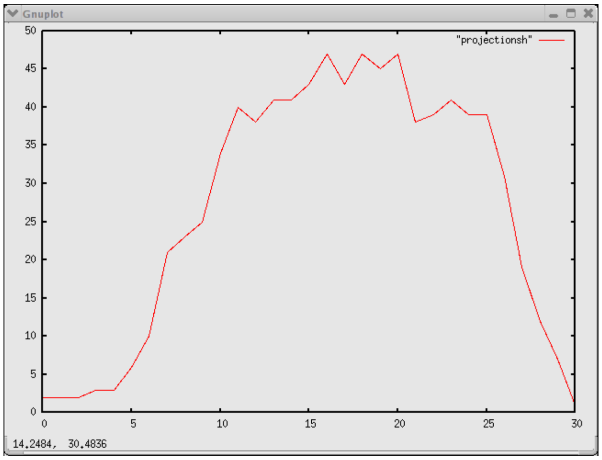


Figure 3: Main axis projection of the *Trichogramma* in its self-referential

After that, we can locate these body parts in the blob and retrieve their positions in the scene. To generate the antenna drumming and ovipositor events, we maintain a history of the positions of the antennas and the ovipositor. By analyzing this history, we can monitor the movements of these parts of the insect and determine if the insect is in a drumming phase, in a drilling phase or in an oviposition phase.

3.4. Recognition of Scenarios Revision

The scenario recognition module, given a priori knowledge of the scene and the scenario models, has to detect which scenario is happening from a stream of observed individuals tracked by a vision module at each instant.

The scenario models are based on the concepts of "state", "event" and "scenario". A state is a spatio-temporal property valid at a given instant or stable on a time interval. An event is a meaningful change of state. A scenario is any combination of states and events and is represented as a constraint network whose nodes correspond to sub-scenarios and whose edges correspond to temporal constraints. Temporal constraints are propagated inside the network to avoid an exponential combination of previously recognized sub-scenarios. The scenarios are modeled in terms of "physical objects" (individual or zones of interest etc.), "components" (which can be primitive-states, composite-states, primitive-events or composite-events) and "constraints" between the physical objects and/or the components (constraints can be temporal, spatial or logical). To describe the scenarios relative to the *Trichogramma* behavior, we use a declarative language specifying generic events which are defined in the video event ontology detailed in [2].

In order to recognize the phases of the egg laying behavior, we have matched the visual features computed by the Visual Feature Algorithm with three new primitive-states.

\# Drumming Scenario
CompositeEvent(Drumming,
  PhysicalObjects((tricho : **Insect**),(z : **Zone**))
  Components((close : **PrimitiveEvent** close_to_zone(tricho,z)))
  Constraints((!**IsStopped**(tricho))
    (**AntennaeMovement**(tricho))
)
  Alarm(**AText**("Trichogramma is Drumming in " + z-¿**Name**)...)
)
\# Drilling Scenario
CompositeEvent(Drilling,
  PhysicalObjects((tricho : **Insect**),(z : **Zone**))
  Components((close : **PrimitiveEvent** close_to_zone(tricho,z)))
  Constraints((**IsStopped**(tricho)))
  Alarm(**AText**("Trichogramma is Drilling in " + z-¿**Name**)...)
)
\# Oviposition Scenario
CompositeEvent(Oviposition,
  PhysicalObjects((tricho : **Insect**),(z : **Zone**))
  Components((drilling : **CompositeEvent** Drilling(tricho,z)))
  Constraints((!**IsOvipositorMoving**(tricho)))
  Alarm(**AText**("Trichogramma is in Oviposition in " + z-¿**Name**)...)
)

Figure 4: An extract from the description of the scenario models representing the three phases of the egg laying behavior.

We have extended the scenario description language by adding these states. The first state indicates whether the *Trichogramma* antennas are drumming a host egg. The second state corresponds to the quasi-immobility of the insect during the drilling and oviposition phases. The last state is relative to the slight movement of the ovipositor discriminating the two last phases of the egg laying behavior. Fig. **??** shows an extract from the description of drumming, drilling and oviposition scenario models which define the three added states. For instance, in Drumming scenario, two physical objects are involved: an insect "tricho" and a zone of interest z corresponding to an egg. To recognize this scenario, one component (sub-scenario) has to be recognized indicating that the *Trichogramma* is close to the egg to parasitize and two constraints needs to be verified: First that the *Trichogramma* remains in motion and second that their antennas are in movement.

The algorithm, part of the scenario recognition module, performs the recognition of the scenarios in two steps: (1) recognition of elementary scenarios (primitive-states and primitive-events) and (2) recognition of composite scenarios (composite-states and composite-events).

During the first step, all elementary scenarios are recognized. To recognize a given elementary scenario model, a loop is performed on all possible sets of physical objects followed by a verification of the corresponding atemporal constraints until all combinations of physical objects have been tested. Once a set of physical objects satisfies all constraints, the elementary scenario is recognized and a scenario instance p is generated and attached with the corresponding scenario model, the set of physical objects and the recognition time t. The scenario instance is then stored in the list of recognized scenarios. If at the previous instant, a scenario instance p' of same type (same model, same physical objects) was recognized on a time interval $[t_0, t-1]$, the two scenario instances are merged into a scenario instance recognized on the time interval $[t_0, t]$.

For the second step we define a composite scenario as a set of sub-scenarios ordered in time, as described in [15]. Each sub-scenario corresponds to a temporal variable in the corresponding scenario model. To speed up recognition process, a scenario model is decomposed in a preprocessing phase into a set of simple scenario models containing at most two sub-scenarios. The recognition of a composed scenario model S_c is triggered by a scenario template, which has been generated when the last sub-scenario p_e terminating S_c has been recognized. The scenario template contains S_c and the scenario instance p_e with its list of physical objects that partially instantiate S_c. As S_c is composed of two sub-scenarios, only one sub-scenario p_s starting S_c remains to be found. If such scenario instance already exists in the past and p_s and p_e satisfy S_c constraints, the scenario S_c is recognized and stored in the list of recognized scenarios.

4. Evaluation and Discussion

To validate our work, we have conducted tests under rigorous conditions. This section describes and discusses the experimental results. First, we describe the system and the test set characteristics and the criteria for experimental evaluation. Second, we show our early results. Finally, we discuss our proposed method and its results.

4.1. Implementation Details

The results are computed on a PC based on an Intel Pentium IV processor and Linux operating system. Our test set consists of a long video (46 minutes) representing a typical laboratory experiment. The image sequence was captured by a digital camera (color image 720x576 pixel at rate of 25 frames/second) placed at the front-end of a microscope. As already explained, we truncate the first five minutes of the video to avoid the background modification caused by the glass dish reversing phase. The acquired video is annotated manually by a human expert. These results are confronted to those obtained automatically by the extended cognitive vision system. For each behavior we compute two indicators: the precision and the sensibility. The precision is the ratio between the numbers of instances well detected automatically and the number of instances detected manually (Ground Truth). The sensibility is the ratio between false-detected instances and manually-detected instances. The content of the test video is resumed by table 1.

Table 1: Number of Occurence of the Parasitoid Activities in the Test Video

Behaviors	Number of occurence
Enter	18
Exit	19
Walk	49
Drumming	42
Drilling	25
Oviposition	21
Total	194

Table 2: *Trichogramma* Behavior Detection Results

Behaviors	Precision	Sensibility
Enter	100%	0%
Exit	100%	0%
Walk	100%	0%
Egg laying behavior phases	85%	15%
Drumming	70%	5%
Drilling and Oviposition	15%	10%

4.2. Results

The outputs of the four components of the vision module at an instant t are shown in Fig. 5: (a) The context and the backgroud image, (b) the cuurent frame to handle, (c) the segmented current frame (blob identified) and (d) the tracked mobile object. Fig. 6 shows the ouput of the scenario recognition module at two intants: When it recognize a Walk scenario (a) and when it recognize a Drilling scenario (b). As shown in table 2, we have obtained acceptable results as feasibility proof. The miss detection of Drumming and Oviposition activities is first due to the segmentation. Besides, the distinction between these two phases is even difficult for expert human. However, we have a low sensiblity which means that the system detects too many false positive scenarios. This is due to the approximative definition of the scenario models. A finer design of these models is currently under development.

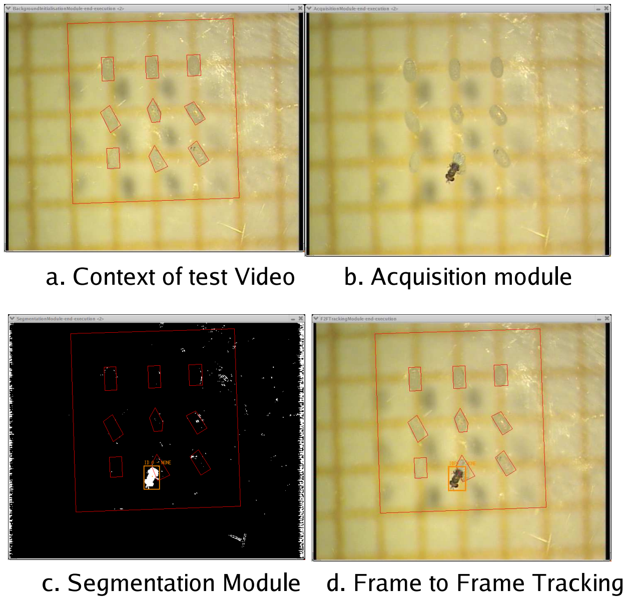


Figure 5: Outputs of the four main components of the vision module.

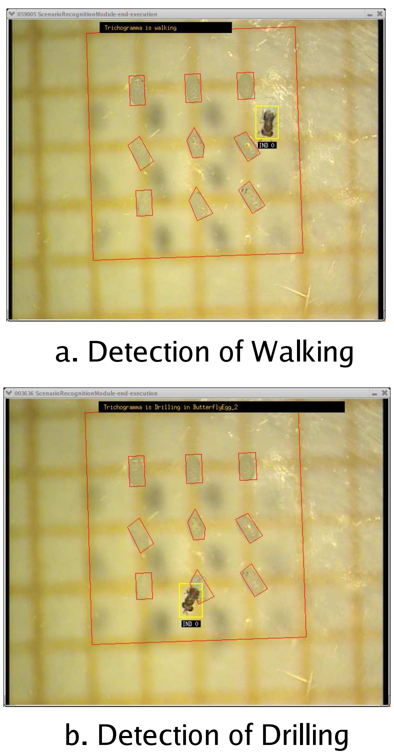


Figure 6: Output of the scenario recognition module.

4.3. Discussion

The first problem in this application is to discriminate the egg laying phases which is not even possible for experts in some cases. Moreover, when expert are able to distinguish this phases, they are not always able to explain how they manage to do it. Thus, there is a main problem of the modelisation of the scenarios of interest. Another problem in this method is the segmentation process. It still has some imperfections: the anteneas and the ovipositor are sometimes not segmented which are the main causes of the detection problem of the three phases of egg laying behavior. In addition, the descriptors of the walk behavior are not accurate: it is defined by the transition between two states which are *Trichogramma* is close to egg i and then *Trichogramma* is close to egg j. Another definition of the walk behavior, corresponding more to the expert definition, is that the *Trichogramma* is in motion and is not overlap any egg zone.

5. Conclusion and Future Work

We have extended a cognitive vision system to monitor *Trichogramma* activities. Our approach computes the visual features characterizing mobile object shape to distinguish between the three phases of the egg laying behavior and allow the recognition of the global activities. After extending the vision module of this system, we have concieved a module which extracts these visual features and bind them to a scenario recognition module. In this paper, we have shown that this behavior recognition approach can be extended over physical objects than human beings. The main challenge is to be able to detect shape features characterizing the body parts of physical objects. As a feasibility proof, we have obtained acceptable results according to the complexity of the egg laying behavior. Previous work has concentrated on studying human behaviors. This paper summarizes an attempt to generalize the utilization of cognitive vision system to monitor non-human activities such as insect parasitoid activities.

This work can be improved by developing the following aspects. First, we plan to improve the segmentation algorithm in order to have a better determination of the three egg laying behavior phases. Currently, we focus on applying program supervision techniques to dynamically tune the parameters of the segmentation algorithm. A second task consists in refining the definition of the walk activity by a better understanding of the knowledge of th expert. Third, in order to develop an operational system, we shall automate the description of the context of the scene. For instance, we have, now, to define the context of each new experiment due to the variations of the background. We are planning to develop a learning module for automating the acquisition of the experimental context which will define the context of the scene without expert help. Finally, in the long term, we also plan to add a new front-end tool which will collect the output of the scenario recognition module for all experiments and mine these outputs to deduce the frequent activities and their probabilistic law: it is claimed that the behavior of the *Trichogramma* while selecting a host egg can be described by the dynamic game theory.

References

[1] A. Avanzi, F. Bremond, C. Tornieri and M. Thonnat, "Design and Assessement of an Intelligent Activity Monitoring Platform," *EURASIP Journal on Applied Signal Processing, Special Issue on "Advances in Intelligent Vision Systems: Methods and Applications"*, Vol. 2005:14, pp. 2359-2374, 11/08/2005.

[2] F. Bremond, N. Maillot, M. Thonnat and V. T. Vu, "Ontologies for Video Events," *INRIA Research Report RR-5189*, May 2004.

[3] N. Chleq and M. Thonnat, "Real-time Image Sequence Interpretation for Video Surveillance Applications," *International Conference on Image Processing (ICIP'96):Proceedings IEEE ICIP'96*, Vol 2, pp. 801-804, Switzerland, 09/1996.

[4] R. Gerber, H. Nagel and H. Schreiber, "Deriving Textual Descriptions of Road Traffic Queues From Video Sequences," *The 15th European Conference on Artificial Intelligence (ECAI'2002)*, Lyon, France, 26/07/2002.

[5] M. Ghallab, "On Chronicles: Representation, On-line Recognition and Learning," *The 5th International Conference on Principles of Knowledege Representation and Reasonning (KR'96)*, USA, 11/1996.

[6] I. Haritaoglu, D. Harwood and L.S. Davis, "Ghost: A Human Body Part Labelling System Using Silhouettes," *14th International Conference on Pattern Recognition*, Brisbane, Australia, 1998.

[7] S. Hongeng, F. Bremond and R. Nevatia, "Representation and Optimal Recognition of Human Activities," *IEEE Proceedings of Computer Vision and Pattern Recognition*, USA, 2000.

[8] A. J. Howell and H. Buxton, "Active Vision Techniques for Visually Mediated Interaction," *Image and Vision Computing*, 2002.

[9] S. Iwasawa, K. Ebihara, J. Ohya and S. Morishima, "Real-time Estimation of Human Body Posture from Monocular Thermal Images," *Conference on Computer Vision and Pattern Recognition*, 1997.

[10] S. Iwasawa, J. Ohya, K. Takahashi, T. Sakaguchi, S. Kawato, K. Ebihara and S. Morishima, "Real-time, 3D Estimation of Human Body Postures from Trinocular Images," *Faculty of Engineering*,Seikei University, 2000.

[11] T. B. Moeslund and E. Granum, "3D Human Pose Estimation using 2D Data and an Alternative Phase Space Representation," *Procedure Humans 2000*, Hilton Head Island, South Carolina, 2000.

[12] N. Oliver and A. Pentland, "Graphical Models for Driver Behaviour Recognition in a Smartcar," *Proceedings of IEEE Intl. Conference on Intelligent vehicles 2000*, Detroit, Michigan, 10/2000.

[13] C. Pinhanez and A. Bobick, "Human Action Detection using PNF Propagation of Temporal Constraints," *M.T.T. Media Laboratory Perceptual Section Report*, No. 423, 04/1997.

[14] N. Rota and M. Thonnat, "Activity Recognition from Video Sequences using Declarative Models," *14th European Conference on Artificial Intelligence (ECAI'2000)*, Berlin, W. Horn (ed.) IOS Press, Amesterdam, 20-25/08/2000.

[15] V.T. Vu, F. Bremond and M. Thonnat, "Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition," *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*,Acapulco, Mexico, 9-15/08/2003.

[16] V.T. Vu, F. Bremond and M. Thonnat, "Temporal Constraints for Video Interpretation," *The 15th European Conference on Artificial Intelligence (ECAI'2002)*, Lyon, France, 21-26/07/2002.

[17] E. Wajnberg, "Analysis of Variations of Handling-time in *Trichogramma* Maidis," *ENTOMOPHAGA*, Vol.34 (3), pp. 379-407, 1989.

[18] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-time tracking of the Human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19:7, pp. 780-785, 1997