Jornadas de Seguimiento de Proyectos, 2009
Programa Nacional de Tecnologías Informáticas

DEVLENA: Descripción de Escenas de Vídeo en LEnguaje NAtural TIN2007-62568

Juan Moreno García *
Oreto Research Group
Universidad de Castilla-La Mancha

Abstract

The overall objective of this project is to develop a cognitive vision system to describe video sequences in a reduced natural language. A representation formalism to describe the video sequence in a comprehensible way must be created, this formalism will be used to obtain the description of the video sequence in a reduced natural language. All the process must be done in real time. To do this, we use as input video sequences without decompress codified in MPEG2 format, so the time of decompressing the video is saved. The created algorithms work without human intervention during all the process. The obtained system will be a mobile system that can be used in different places. The final system will be applied to two real applications of video monitoring to test the representation and the algorithms created.
Keywords: Video sequence description, Knowledge representation, Multi-agent systems, Fuzzy logic

1 Proposed objectives of the project

This section details the objectives of the project, the human and material resources to obtain them, and the time table proposed to do this. The following sections show these objectives.

1.1 Objectives of the project

The main objective of this project is to develop a cognitive vision system to describe video sequences using a reduced natural language. That is, we want to describe video sequences in a descriptive and qualitative way. A formalism representation must be created to describe the video sequence, and the description of the video sequence in a reduced natural language will be obtained by using this representation.

The obtained system must verify the following qualities:

- The capability to work in real time, to do this, we initially believe in the use of the distributed multi-agent systems that allow distributed execution in different machines.

*Email: juan.moreno@uclm.es

- As input of our system we use MPEG2 video without decompressing since this format contains enough information to reach our objectives, and it allow us to save the time of decompression.
- The system will work in an autonomous way, that is without human intervention. The system directly learns from the video sequence in an autonomous way.
- The system must be a mobile system, to do this, it will be used mobile devices as wi-fi cameras, solar panels, etc.

Two real applications will be created to test the algorithms and the created video sequence representation. Both applications belong to the field of the video monitoring, these applications are efficient in cost and necessary at present. So, we think that this scope is appropriate and useful to test our system.

The first application will use the created techniques in order to detect automatically non-habitual visual events and the automatic indexation of the recorded sequences. The movement description of the different objects in the video sequence is desired to be able to catalogue them.

The second application will use the capacity of the system to work in real time in order to narrate the events of the video to a human operator who does not arrange at those moments of graphical interface. More concretely, the system will indicate in a sonorous way to the human operator the situation of the non-habitual objects, so the operator is totally informed about the real situation of the objects by means of some type of sonorous interface (radio, mobile telephone, etc) and he can do the correct actions.

The web page of this project is *http://raro.oreto.inf-cr.uclm.es/apps/devlena/index.php*.

1.2 Human resources and material resources

Firsts of all, we want to comment that the money granted in the concession has been inferior to the initial request, eliminating the subsidy for the hiring of an Engineer in Computer Science. This fact has brought delays in the software implementation to test the proposed algorithms, models and representation formalisms.

Our research group has two types of resources to make this project: human resources and material resources.

- Human resources: Initially, the project counts with seven researchers, three of them are Doctors and the rest of them are no Doctors, although during the development of the project a researcher (Luis Rodríguez Benítez) has completed its formation obtaining the degree of Doctor. On the other hand, the incorporation to the project of two new researchers was asked for during the development of this project. These researchers are the Doctor Caridad Perez de los Reyes - she collaborates in the task of equipping the system with mobility, mainly as regards the power subject, and the doctorate student David Vallejo Fernández - that participates in the design and implementation of the platform of the multi-agent system.
- Material resources: All the researchers are members of the Oreto Research Group (*www.-oreto.inf-cr.uclm.es*). This group has two research laboratory, one in Ciudad Real and another one in Toledo. The computer science material and the artificial vision material

(video cameras, antennas, power amplifier, feeding kit, etc.) necessary for the development of the project has been bought with the money subsidized with this project. Also these materials will allow us to continue with the research line of the project in the future.

Subtasks	Description
$T1$	Study of the state of art of the cognitive vision
$T2$	Study of the state of art of the multi-agent systems
$T3$	Design of the cognitive vision multi-agent system
$T3.1$	Development of the cognitive vision multi-agent system
$T3.2$	Design of the agents that compound the subtasks T3.1
$T3.3$	To equip the system with the capacities of real time execution and mobility
$T4$	Design of the cognitive vision multi-agent system
$T4.1$	Development of the general architecture of the multi-agent system
$T4.2$	Development of the agents that compound the subtasks T4.1
$T5$	Design, development and implementation of the two real proposed applications
$T5.1$	Study of the first application
$T5.2$	Study of the second application
$T5.3$	Apply the architecture to the first application
$T5.4$	Apply the architecture to the second application

Table 1: Subtasks of the project

1.3 Time table

Figures 1 and 2 show the time table of the project tasks where each task and its subtasks are described in Table 1.

2 Level of obtained success in the execution of the project

This section shows the level of the success of the project, to detail this, the section is divided in three subsections.

2.1 Obtained objectives and not obtained objectives

In the beginning of the project we have a lot of problems with the segmentation and tracking of the video sequence. These problems are related to the over-segmentation and merge of regions. We spend long time in solving these problems, and they are not resolute absolutely still. Finally, our segmentation and tracking method obtains acceptable results [7, 13, 10, 11] and it makes use of video sequences in MPEG2 format without decompress. Currently we are investigating in this line in the Thesis of Cayetano Solana Ciprés to improve the obtained results in our segmentation and tracking methods [13]. The found solutions do not use the multi-agent technology in the segmentation and tracking phases as a consequence of these problems as we indicated in the project request memory. So, in this project we do not use the

Actividades/Tareas	Centro Ejecutor	Persona responsable y otras involucradas	Primer año (*)	Segundo año (*)	Tercer año (*)
T1	UCLM	Juan Moreno Luis Jiménez Luis Rodríguez	XIXIXIXI		
T2	UCLM	José Jesús Castro, C. González, E. Castillo, E. Domínguez	XIXIXIXI		
T3.1	UCLM	Juan Moreno Luis Jiménez José Jesús Castro	IXIXIXI		
T3.2	UCLM	Juan Moreno Luis Jiménez José Jesús Castro	IXIXIXIXIXI		
T3.3	UCLM	Luis Rodríguez, C. González, E. Castillo, E. Domínguez, Ingeniero contratado	IXIXIXIXIXIXI		
T4.1	UCLM	Todo el equipo	IXIXI		
T4.2	UCLM	Todo el equipo	IXIX		

Figure 1: Time Table 1

multi-agent technology in the segmentation and tracking phases but the desired objectives are reached with another solutions.

The obtained objectives are the following:

- Our method is able to work with video encoded by using MPEG2 without decompress [7, 8, 9, 11].

- We develop a representation formalism that describes the motion of the objects of a video sequence. Our algorithms directly learn from the video sequence. They find the macroblocks which represent motion in the image by studying the motion vectors of MPEG2 video. So, the motion vectors are used to detect the motion in the sequence combined with fuzzy logic to manage the vagueness caused by working with video without decompressing [9]. Then our method uses concepts like linguistic motion vectors, linguistic blobs, linguistic objects [7, 8, 9, 10, 11, 12].

- The created method works in real time [13].

- The first real application proposed to detect automatically the non-habitual visual events and the automatic indexation of the recorded sequences. Our work in this line is determined by several factors: (1) The use of fuzzy models to compare different behaviours in traffic video sequences [10]; (2) The use of a Mealy Machine to recognize actions in video sequences based on the representation formalism exposed previously [12]; (3) Analysis of the normality to detect abnormal behaviors [1, 2, 5, 4].

Actividades/Tareas	Centro Ejecutor	Persona responsable y otras involucradas	Primer año (*)	Segundo año (*)	Tercer año (*)
T5.1	UCLM	Juan Moreno Luis Jiménez L. Rodríguez, Ingeniero	| | | | | | | | | | |	| | | IX|X| | | | |	| | | | | | | | | | |
T5.2	UCLM	José Jesús Castro, C. González, E. Castillo, E. Domínguez	| | | | | | | | | | |	| | | IX|X| | | | |	| | | | | | | | | | |
T5.3	UCLM	Juan Moreno Luis Jiménez Luis Rodríguez, Ingeniero	| | | | | | | | | | |	| | | | IX|X|X|X|X|X|X|X	| | | | | | | | | | |
T5.4	UCLM	José Jesús Castro, C. González, E. Castillo, E. Domínguez	| | | | | | | | | | |	| | | | IX|X|X|X|X|X|X|X	| | | | | | | | | | |
T6	UCLM	Ingeniero contratado	X|X|X|X|X|X|X|X|X|X|X|X	X|X|X|X|X|X|X|X|X|X|X|X	| | | | | | | | | | |

Figure 2: Time Table 2

- The obtained methods are easily to establish as a mobile system, currently we are working in this objective.

Finally, we want to comment respect to the second real application that we have not got enough time in this project to test this problem. This will be a future work.

2.2 Found difficulties during the project

The main problem has been in the development of the segmentation and tracking from MPEG2 video sequences without decompress, we spend a lot of time in this task.

Another problem is related to the duration of this project, when we requested the subsidy of this project we express that we want a project with a duration of two years instead of three years in order to be able to request a coordinate project with the group managed by Dr. Miguel Delgado Calvo-Flores, from the University of Granada, in the projects call of the year 2009, this new project already has been asked for. This has caused that we have only worked during two years.

Finally, we can not to hire an Engineer in Computer Science what it has caused delays in the software implementation of the proposed algorithms, models and representation formalisms.

2.3 Scientific Obtained result

During the execution of this project two Doctoral Thesis have been defended:

- A Thesis relates to the recognition of actions on the signal of compressed video by using fuzzy logic [9]. This Thesis formalizes, develops and implements a cognitive vision system that makes the segmentation, tracking and analysis phases obtaining as output the description of video sequences in a reduced natural language.

- A Thesis that presents a multi-agent architecture to the optimization of rendering [3]. This Thesis presents a multi-agent architecture for the process of rendering.

Three publications in Journals with Impact Factor (Journal Citation Reports) will be published [10, 1, 15].

Nine papers in national and international conferences are published: [2, 4, 6, 7, 8, 11, 12, 13, 15].

3 Indicators of results

Now, we expose the information to evaluate the results of this project.

- Objectives' attainment degree of the project: We want to highlight that we obtain a representation formalism that is qualitative and descriptive. Our method does the segmentation and the tracking, and learns directly from the video sequence. Our method works in an autonomous way, we used video encoded in MPEG2 without decompress, and our system works in real time. We believe that our method can be convert in a mobile system. We only work in the development of the first real application proposed. The not obtained objectives are the following: (1) The segmentation and tracking have been done without using the multi-agent technology, although this technology is used in another research lines to solve the first application proposed. This work has been obtained a base multi-agent platform that can be extended to the segmentation and tracking. (2) The implemented applications do not have commercial appearance since they have been used for our tests.

- Relevance and originality of the obtained results: We have obtained publications in journals with impact factor in the Journal Citation Report, and two Thesis are defended with the maximal scope.

- The technological production is the implemented software. The algorithms of segmentation, tracking and analysis [7] are programmed by using the python and C languages, although this has not commercial appearance. A multi-agents system [15] has been implemented according to the specifications of the FIPA committee (http://www.fipa.org/). In other works, there are specific agents designed according to this specifications in order to guarantee the management, communications and inter-operative between the different agents of the platform. This multi-agents platform uses ZeroC ICE (http://www.zeroc.com) that allows the agents to be implemented in different programming languages, it can be executed in different hardware or software platforms.

- Usefulness of the results and relations with the socio-economic environment: The obtained representation formalism and algorithms are useful to the video surveillance. In the future it can be used to detect abnormal situations in public spaces. In present, the risks have

been reached and the necessity to protect this zones of terror attacks. These types of solutions are efficient in cost.

- Formation of human resources: This project has been contributed in the obtaining of the degree of Doctor of two of our researchers. Six doctorate students are researching in the line presented in this project under the supervision of researchers of our research group (www.oreto.inf-cr.uclm.es).

- Collaborations with other European or international researcher groups: Our research group has excellent relations with national and international groups. Currently we maintain interchanges of the researchers with the *Software Competence Center* of Hagenberg (Austria) whose main researcher is the Dr. Gerhard Weiss, and with the *Digital Imaging Research Center* of the University of Kingston in Londres whose main researcher is the Dr Paolo Remagnino. Also, during this project, we have established a collaboration with Dr. Gerardo Fernandez-Escribano from the I3A of Albacete (http://www.i3a.uclm.es/) that collaborate with us in the codification and decodification of the MPEG2 format.

4 Conclusions

This projects allow us to begin in the research of the description of video sequences. Our initial idea was to work with multi-agent system during all the phases of the cognitive video system, but the problems undergone in the initial phases forced us to use other techniques and do not used agents in segmentation and tracking phases. On the other hand, the analysis phase has been focused of different ways obtaining valid results that has lead to the publication of papers in journals with impact factor.

In brief, a project executed during two years has reached an interesting research line that has provided us the publication of papers in journals with impact factor, Doctoral Thesis and a new line of work for the Doctoral Student of the group. We believe that is a good line of work for the future.

As future work, we must to finish the not obtained objectives, mainly the improve of the segmentation and tracking, and the fusion of the representation formalism and the developed algorithm with the multi-agent architecture developed during this project.

References

[1] J. Albusac, D. Vallejo, L. Jimenez, J.J. Castro-Schez, L. Rodriguez. Intelligent Surveillance based on Normality Analysis to Detect Abnormal Behaviors. *International Journal of Pattern Recognition and Artificial Intelligence*. Accepted to be published.

[2] J. Albusac, J.J. Castro-Schez, D. Vallejo. Learning maximal structure rules with pruning based on distances between fuzzy sets. *Proceedings of Information processing and management of uncertainty in Knowledge-based systems*,441-447, Malaga, Spain, 2008.

[3] Carlos Gonzalez Morcillo. Arquitectura multi-agente para la optimización del proceso de rendering. *Tesis Doctoral*, 2007/11/26, Sobresaliente Cum Laude, Doctorado Europeo.

[4] L. Lopez-Lopez, J. Albusac, J.J. Castro-Schez, L. Jimenez. Semantics-provided environment views for normality analysis based intelligent surveillance. *International Conference on Agents and Artificial Intelligence (ICAART'09)*, Oporto, Portugal, Febrero, 2009.

[5] L.M. Lopez-Lopez, J. Albusac, J.J. Castro-Schez, J. Jimenez, J.A. Mateos. Multi-Camera Information Fusion to achieve Global Normality Analysis in Intelligent Surveillance. *International Journal of Computer Vision and Image Understanding*, Sometido.

[6] J. Moreno-Garcia, J.J. Castro-Schez, L. Jimenez. A new method to compare Dynamical Systems. *Proceedings of the International Fuzzy Systems Association 2007*, LNCS, Springer Berlin, 533-542, Cancun, 18-21 June 2007, DOI 10.1007/978-3-540-72950-1-53.

[7] L. Rodriguez-Benitez, Juan Moreno-Garcia, J.J. Castro-Schez, J. Albusac, L. Jimenez. An approximate reasoning technique for segmentation on compressed MPEG video. *2nd International Conference on Computer Vision Theory and Applications (VISAPP 2007)*, 184-191, Barcelona, Spain, 2007.

[8] L. Rodriguez-Benitez, J. Moreno-Garcia, J. Albusac, J. J. Castro-Schez, L. Jimenez. Seguimiento de objetos representados linguísticamente utilizando técnicas de razonamiento aproximado. *II Simposio sobre Lógica Fuzzy y Soft Computing (EUSFLAT 2007)*, 237-244, Zaragoza, Spain, 2007.

[9] Luis Rodriguez Benitez. Razonamiento aproximado aplicado al reconocimiento de acciones sobre la senal de video comprimido. *Tesis Doctoral*, 2008/03/14, Sobresaliente Cum Laude.

[10] L. Rodriguez-Benitez, J. Moreno-Garcia, J.J. Castro-Schez, J. Albusac, L. Jimenez. Automatic Objects Behaviour Recognition from Compressed Video Domain. *Image and Vision Computing*. DOI 10.1016/j.imavis.2008.07.002.

[11] L. Rodriguez-Benitez, J. Moreno-Garcia, J.J.Castro-Schez, and L. Jimenez. Fuzzy logic to track objects from MPEG video sequences. *Proceedings of Information processing and management of uncertainty in Knowledge-based systems*, 675-681, Malaga, Spain, 2008.

[12] L. Rodriguez-Benitez, J. Moreno-Garcia, J.J. Castro-Schez, C. Solana, L. Jimenez. Action recognition in video sequences using a Mealy machine. *Proceedings of World Academy of Science Engineering and Technology*, 2008 July 25-27, 42-48, Prague, Czech Republic.

[13] C.J. Solana-Cipres, L. Rodriguez-Benitez, J. Moreno-Garcia, G. Fernandez-Escribano. Algoritmo de segmentación de vídeo en tiempo real sobre vídeo comprimido H.264/AVC. *ESTYLF 2008*, Mieres (Spain), 299-306. 17-19 Septiembre 2008. ISBN: 978-84-691-5807-4.

[14] D. Vallejo, J.J. Castro-Schez, J. Albusac, L. Jimenez, C. Glez-Morcillo. Integrating a Standard Communication Protocol into an E-Commerce Environment based on Intelligent Agents. *9th International Conference On Enterprise Information Systems (ICEIS)*, 55-60. 2008. ISBN: 978-989-8111-39-5.

[15] D. Vallejo, J. Albusac, C. Gonzalez, L. Jimenez, J. Moreno-Garcia. A Cognitive Surveillance System for Detecting Incorrect Traffic Behaviors. *Expert Systems With Applications*. Accepted to be published in 46(3) of this journal.