MEMORIA CIENTÍFICO- TÉCNICA PARA ACCIONES INTEGRADAS

1. RESUMEN DE LA PROPUESTA (Debe rellenarse también en inglés)

INVESTIGADOR PRINCIPAL: Antonio J. Bandera Rubio

INSTITUCIÓN BENEFICIARIA (CENTRO I+D): Universidad de Málaga

TÍTULO DEL PROYECTO: Visual Attention using Combinatorial Pyramids (VACOP)

RESUMEN
(breve y preciso, exponiendo solo los aspectos más relevantes y los objetivos propuestos)
El objetivo de este proyecto es el desarrollo de un sistema artificial de atención, que filtrará la información visual adquirida para así detectar los objetos relevantes de la escena. En este sistema, el mapa de saliencia que determinará qué regiones de la escena son más relevantes será construido usando un proceso de agrupación perceptivo, que será implementado usando una estructura jerárquica. En concreto, para conseguir que se pueda trabajar en los niveles superiores de esta estructura conservando toda la información topológica de la escena, se estudiará el empleo de estructuras basadas en mapas combinatorios. Tomando como entrada distintas características de bajo nivel (contraste color, profundidad o bordes, por ejemplo), el proceso de agrupación perceptivo deberá obtener de la secuencia de vídeo distintos elementos u 'objetos pre-atentivos', que formarán parte de objetos reales presentes en la escena. La posibilidad de tratar con entornos dinámicos se conseguirá incluyendo un mecanismo de seguimiento de objetos en movimiento, que será también implementado sobre la misma pirámide combinatoria usada para la detección de objetos de interés.

PROJECT TITLE: Visual Attention using Combinatorial Pyramids (VACOP)

SUMMARY
(brief and precise, outlining only the most relevant topics and the proposed objectives)
The aim of this project is to develop an artificial visual attention model. This mechanism will be responsible of extracting the relevant objects of the real scene from the broad input of visual information provided by the acquired video sequence. In order to build the saliency map associated to the sensed scene, a novel perception-based grouping process will be employed. This grouping process will be performed by a hierarchical irregular structure which should be able to preserve the topological information at upper levels. Thus, these levels could be used to perform the perceptual grouping process. Specifically, this project proposes to study the application of combinatorial pyramids to accomplish this task. Taking into account low-level image features (e.g. colour contrast, edge or depth information), the perception-based grouping process will extract from the input video sequence, the different parts or 'pre-attentive objects'. These parts will correspond to units of visual information that can be bound into coherent and stable objects. In addition, the ability to handle dynamic scenarios will be included in the proposed model by introducing a tracking mechanism for moving objects. This mechanism will be also performed using the same hierarchical structure, reducing the computational time.

2. DESCRIPCIÓN DEL PROYECTO DE I+D CONJUNTO

(Máximo 2 páginas)

Deben tratarse aquí: la finalidad del proyecto; los antecedentes y estado actual de los conocimientos científico-técnicos, incluyendo la bibliografía más relevante; el interés de la colaboración propuesta, los grupos nacionales o internacionales que trabajan en la temática del proyecto.

Overview of the proposal

In biological vision systems, the attention mechanism is responsible for preselecting possible relevant information from the sensed field of view so that the complete scene can be analyzed using a sequence of rapid eye saccades. In the recent years, efforts have been made to imitate such attention behaviour in artificial vision systems, because it allows to optimize the computational resources as they can be focused on the processing of a set of selected regions only. The aim of this proposal is to define an object-based framework of visual attention, which will be able to handle dynamic environments. This system will integrate bottom-up (data-driven) and top-down (model-driven) processing. The bottom-up component will determine salient 'pre-attentive objects' by integrating different features into the same hierarchical structure. Specifically, we propose to achieve these perception-based grouping process using a combinatorial pyramid. Thus, the image topology will be preserved at upper levels, allowing to correctly encode relationships among image regions. It must be noted that these 'pre-attentive objects' or 'proto-objects' will be image entities which will not necessary correspond with a recognizable object, although they will possess some of the characteristic of objects. It could be considered that they will be the result of the segmentation of each frame of the input video sequence into candidate objects (i.e. grouping together those input pixels which will be likely to correspond to parts of the same object in the real world, separately from those which are likely to belong to other objects). This process will group the image pixels into entities that can be considered as segmented perceptual units. We will also study the possibility to perform the perceptual-grouping process on the 2D + time or 3D + time space. The top-down component could make use of object templates to filter out data and shift the attention to objects which are relevant to accomplish the current task (e.g. human faces in a human-robot interaction framework). Generic knowledge could be used to select potential areas of attention in this component. If the knowledge is acquired before, it could led to a hierarchy describing the structure of an articulated object with abstract properties of the entities (e.g. connectivity, articulation…). Such information can be efficiently used in the top-down search to focus quickly on the more discriminant parts of the objects. Finally, in a dynamic scenario, the location and shapes of the objects may change due to motion and minor illumination differences between consecutive acquired images. To deal with these scenes, a tracking approach for 'inhibition of return' will be employed. This tracking process will be conducted using the same combinatorial hierarchy.

Theoretical models of visual attention

Probably one of the most influential theoretical models of visual attention is the spotlight metaphor, by which many concrete computational models have been inspired (Koch and Ullman, 1985)(Itti, 2002). These approaches are related with the feature integration theory, a biologically plausible theory proposed to explain human visual search strategies. According to this model, these methods are organized into two main stages. First, in a preattentive task-independent stage, a number of parallel channels compute image features. The extracted features are integrated into a single saliency map which codes the saliency of each image region. The most salient regions are selected from this map. Second, in an attentive task-dependent stage, the spotlight is moved to each salient region to analyze it in a sequential process. Analyzed regions are included in an inhibition map to avoid movement of the spotlight to an already visited region (Itti, 2002). Thus, while the second stage must be redefined for different systems, the preattentive stage is general for any application. Although these models have good performance in static environments, they cannot in principle handle dynamic environments due to their impossibility to take into account the motion and the occlusions of the objects in the scene. In order to be able to track moving objects, an attention control mechanism must integrate depth and motion information (Maki et al, 2000).

The previously described methods deploy attention at the level of space locations (space-based models of visual attention). They have some intrinsic disadvantages: in a normal scene, objects may overlap or share some common properties.Then, attention may need to work in several discontinuous spatial regions at the same time. If different visual features, which constitute the same object, come from the same region of space, an attention shift will be not

required. Other approaches deploy attention at the level of objects. They are based on the assumption that attention must be directed to an object or group of objects, instead to a generic region of the space. Object-based models of visual attention provide a more efficient visual search than space-based attention. Besides, it is less likely to select an empty location. In the last few years, these models of visual attention have received an increasing interest in computational neuroscience and in computer vision (Orabona et al, 2007)(Pylyshyn, 2001). Therefore, these models will reflect the fact that the perception abilities must be optimized to interact with objects and not just with disembodied spatial locations i.e. be data dependent and adaptive. Thus, visual systems will segment complex scenes into objects which can be subsequently used for recognition and action.

(Koch and Ullman, 1985) Koch, C., Ullman, S.: Shits selective visual attention: Towards the underlying neural circuitry. Human Neurobiology 4, 219-227 (1985)

(Itti, 2002) Itti, L.: Real-time high-performance attention focusing in outdoors color video streams. Proc. SPIE-HVEI, 235-243 (2002)

(Maki et al, 2000) Maki, A., Nordlund, P., Eklundh, J.: Attentional scene segmentation: integrating depth and motion. Computer Vision and Image Understanding 78(3), 351-373 (2000)

(Orabona et al, 2007) Orabona, F., Metta, G., Sandini, G.: A proto-object based visual attention model. Proc. WAPCV, LNAI-4840, 198-215 (2007)

(Pylyshyn, 2001) Pylyshyn, Z.: Visual indexes, preconceptual objects, and situated vision. Cognition 80, 127-158 (2001)

Justification of the proposed bilateral action

The aim of this project is to establish a forum where the research groups of the Pattern Recognition and Image Processing (PRIP) of the Vienna University of Technology and the TIC-125 (ISIS) and FQM-296 Groups of the Universities of Málaga and Seville can exchange ideas and work together to develop an object-based visual attention system based on a topology-preserving, combinatorial pyramid. The ISIS group has experience on the development of visual perception system for mobile robots, which was acquired under projects TIN2005-01359 and VISOR, funded by the Spanish Government and by the European Robotics Network (EURON), respectively, and which is currently being addressed under project TIN2008-06196, funded by the Spanish Government. On these implementations, the visual information is grouped using a hierarchical structure which does not preserve the topology of input images. Therefore, the information encoded at upper levels is not complete, and interesting relationships between image parts, e.g. inclusions, are lost. On the other hand, the PRIP group has a broad experience on the use of topology-preserving hierarchical structures. Thus, they proposed the original idea of using dual graphs to encode the information inside each hierarchy level to simultaneously deal with regions and edges. Besides, they have also been one of the first research groups in using combinatorial maps to encode this same information about regions and edges into a single directed graph. They are currently working in related projects funded by the Austrian Government (CHIC, TWIST-CV, Cognitive Vision) or by the European Union (EuCognition II, MUSCLE), and are partners in a EU FP7 project proposal about the study of topology within the context of 4D digital imagery (under review). This project has been applied by Dr. Real, from the FQM-296 group.

International research groups

- Models of visual attention

Center of vision research (Prof. Tsotsos) - http://cvr.yorku.ca/home/index.html

Princeton University (Prof. Treisman) - http://weblamp.princeton.edu/~psych/psychology/home/index.php

- Topology-preserving hierarchical structures (Larger group: IAPR TC15)

Université de Poitiers - SIC (Prof. Lienhardt, Prof. Damiand) - http://www.sic.sp2mi.univ-poitiers.fr/

ENSICAEN (Prof. Brun) - http://www.greyc.unicaen.fr/~luc/index_eng.html

INSA - Lyon (Prof. Jolion) - http://rfv.insa-lyon.fr/~jolion/

LIS Lab – Grenoble (Prof. Montanvert) - http://www.lis.inpg.fr/english/ind_enseign.html

DISI Genova (Prof. de Floriani) – http://www.disi.unige.it/

3. OBJETIVOS DEL PROYECTO

(Máximo 1 página)

3.1. Describir con **claridad, concreción y de manera realista los objetivos** del proyecto de colaboración

- Development of a perception-grouping process based on a combinatorial pyramid
- Detection of relevant 'pre-attentive objects' using this novel grouping scheme
- Tracking and topological analysis of relevant image items using the same hierarchical structure
- Integration of the detection and tracking processes inside an attention mechanism at a semi-attentive stage
- Generic hierarchy-controlled processing. This generic knowledge can be acquired before and led to a hierarchy describing objects of interest with abstract properties (e.g. connectivity, articulation…)

3.2. Indicar el **interés y el valor añadido de la colaboración propuesta**, para los objetivos del proyecto

The central item of the proposal is the use of a topology-preserving hierarchical structure inside an attention mechanism. The PRIP group of the Vienna TU is one of the pioneering research groups on the study and development of such type of structures. Specifically, we understand that the combinatorial pyramid could be used to perform a correct perception-based grouping of the input image, allowing the detection of salient image regions. Selected objects could be represented by hierarchical templates, which could be useful to implement a fast tracking process inside the same pyramid structure. In the context of artificial vision systems, this will allow us to implement the so-called 'inhibition of return'. Although the ISIS group has experience on the development of the described architecture, previous works have been conducted using hierarchies which do not preserve the image topology at upper levels. This causes that several relationships among image regions, such as inclusions, must be studied at the base level of the hierarchy, i.e. using the original input image. This is computationally expensive, a feature which is not tolerable for a visual attention mechanism. On the other hand, the FQM-296 group provides the possibility to represent objects using topological invariants. They are currently collaborating with the PRIP group. This proposal will allow that these three groups can integrate their research activities inside an interesting, real application.

3.3. Indicar los **antecedentes y resultados previos,** del equipo solicitante español y extranjero, en la temática del proyecto

During the last years, the group of Ingeniería de Sistemas Integrados (ISIS) has been granted with two projects, TIN2005-01359 and VISOR, where the main guidelines of the currently proposed system were initially drawn. Specifically, in the last project, the ISIS group proposed, together to the ISR group of Coimbra, an artificial visual system for social robots based on the principles of active vision and task-oriented perception. The recently approbed TIN2008-06196, funded by the Spanish Government, is a natural continuation of this work, but it is now mainly focused on the attentive stage of the visual perception system. The FQM-296 group has a dilated experience in representing objects by topological invariants. On the other hand, the group of Pattern Recognition and Image Processing (PRIP) of the TU of Vienna is currently working on several projects funded by the Austrian Government (TWIST-CV, Cognitive Vision...) whose main goal is to develop a framework that provides solutions to practical problems of cognitive processing using approaches that strongly use structure and hierarchical representations.

3.4. Indicar **novedad y relevancia de los objetivos**, utilidad de resultados y relaciones con el entorno socio-económico:

An attention mechanisms can be the central part of any vision system which tries to process the entire perceived scene in a fast way. The bottom-up component of visual attention could be useful to develop autonomous robots which will be able to unsupervisedly detect visual, natural landmarks for localization and mapping. The top-down component of visual attention could be employed to detect human faces or face features (e.g. eyes) for face or iris recognition when the whole image is not ease to process. Currently, the ISIS group is working on the development of a top-down image processing approach, which goes from human faces to eyes, for iris recognition inside of the CENIT project 'Investigación en Tecnologías para la Gestión de la Migración' ('Integra').

4. METODOLOGÍA Y PLAN DE TRABAJO

(Máximo 2 páginas)

Se deben **detallar y justificar** las actividades de colaboración que se proponen realizar y si es posible, incluyendo cronograma.

El plan de trabajo debe desglosarse en actividades o tareas, fijando los hitos que se prevé alcanzar en cada una de ellas. En cada una de las tareas, deben indicarse el centro ejecutor y las personas involucradas en la misma.

Detallar el plan de visitas e intercambios que se propone realizar, indicando la previsiones, calendario, duración de las mismas y personas implicadas.

Project schedule

Task 1 - Tracking objects using combinatorial pyramids (01/2010 – 10/2011)

Outcome – The development of a tracking framework based on the representation of the object to track (template) and the perceived scene by means of combinatorial pyramids.

People involved: W. Kropatsh, A. Ion, E. Antúnez, N. Artner, D. Shao, R. Marfil

Task 2 – Definition of the general framework of the visual system (01/2010 – 01/2011)

Outcome – In order to develop the visual attention mechanism, it is necessary to define a software architecture where the different modules run (interfaces functionality…). It will be also necessary to describe the aim to reach for each module. This work is extended for one year, as we understand that it could be changed. The final implementation of the software architecture will be achieved at the same time that the rest of tasks.

People involved: A. Bandera, R. Marfil, F. Sandoval, W. Kropatsch, A. Ion, Y. Haxhimusa, A. Palomino

Task 3 – Generation of bottom-up saliency maps (01/2010 – 07/2011)

Outcome – Bottom up saliency maps can be obtained from different low-level image features. The integration of the saliency measure inside of a hierarchy of graphs will provide a first version of an object-based attention mechanism. In the TIC-125 group, we have a first version of this mechanism, which runs using the Bounded Irregular Pyramid. This pyramid does not preserve the topological information. The aim is to change the employed hierarchy by a Combinatorial Pyramid.

People involved: R. Marfil, J.P. Bandera, A. Palomino, R. Vázquez-Martín, P. Núñez, A. Ion, E. Antúnez

Task 4 – Integrating top-down component inside the attention mechanism (01/201-12/2011)

Outcome – Top-down component of the visual attention mechanism will permit to search for interest objects in the perceived scene. This will allow modulating attention by means of the task to reach (e.g., it the robot has a map of visual landmarks of the environment, then it could look for these landmarks in advance). This task must deal with the problem to match the hierarchical representation of the object templates and the representation of the perceived scene. Both representations will be conducted by means of combinatorial pyramids.

People involved: Y. Haxhimusa, W. Kropatsch, A. Bandera, R. Marfil, F. Sandoval

Task 5 – Analysis of topological invariants (01/2010-12/2011)

Outcome – In these last years, the interest of the pattern recognition community in using precise object representation provided by algebraic topology has increased. In this project, we consider computing and using homology in the context of measure the importance of the 'pre-attentive objects' detected at the upper-levels of the hierarchy. Thus, typical saliency measure driven by low-level image features will be enhanced with other measures provided by the topological information of the objects.

People involved: P. Real, H. Molina-Abril, A. Ion, A. Bandera

Task 6 – Evaluation of the proposed attention mechanism (1/2011-12/2011)

Outcome – The proposal of visual attention mechanism will be tested in real situations. The integration of new components will be achieved in a sequential way. That is, it is expected that we can firstly test the bottom-up component without take into account topological information; then, topological information to measure saliency can be added in a second step; and this mechanism could be finally extended to deal with dynamic environments when the tracking process using combinatorial pyramids will be finished. Top-down component of visual attention will be probably added at the last stages.

People involved: All researchers

Meetings

Four meetings will be organised to share information and experiences. In these meetings, the main advances of both groups will be presented and discussed. The tentative dates for these meetings will be:

- June-2010
    - Place: Vienna
- December-2010
    - Place: Sevilla
- June-2011
    - Place: Vienna
- December-2011
    - Place: Málaga

Researcher exchanges

It will be also desirable to encourage researcher exchange between both groups. A tentative timetable of these stays could be:

- Spring 2010: Walter Kropatsch at Seville/Málaga (1 month)
- June-July 2010: Antonio J. Palomino at Vienna (1 month)
- Fall 2010: Nicole Artner at Málaga (1 month)
- June 2011: Rebeca Marfil and Antonio Bandera at Vienna (1 month)
- Spring 2011: Yll Haxhimusa and Adrian Ion at Málaga (1 month)

5. BENEFICIOS DEL PROYECTO

(Máximo 1 página)

Indicar si se trata de una consolidación de una relación previa subrayando lo que aporta de nuevo respecto a la relación existente, o

Indicar si se trata del establecimiento de una nueva relación que resultados se esperan de la misma

Deben destacarse en ambos puntos, entre otros, los siguientes aspectos:

- Participación conjunta en proyectos de investigación internacionales o europeos,
- Contribuciones científico-técnicas esperables del proyecto, beneficios esperables para el avance del conocimiento y de la tecnología: tesis doctorales, publicaciones científicas o de divulgación, patentes, aplicaciones informáticas, informes técnicos,
- Obtención de nueva metodología / tecnología,
- Inicio de nueva línea de trabajo o abordaje experimental,
- Realización de metodología/tecnología que no es accesible en el centro de origen,
- Posible utilidad de resultados,
- Relaciones con el entorno socio-económico.

Hierarchical structures have been widely employed in computer vision or image processing areas. In this research field, both groups have proposed different decimation processes and structures. Specifically, under the supervision of Prof. Kropatsch, the PRIP group has proposed the dual graph pyramid, the first pyramid structure where the topology is preserved at upper levels. In recent years, Prof. Kropatsch has been worked with Luc Bruc in encoding the levels of the hierarchy by means of combinatorial maps. These maps allow to encode the same information of dual graph pyramids into a directed simple graph, and cand be constructed in higher dimensions (e.g. 3D, 4D). On the other hand, the main goal of the research activities of the TIC-125 (ISIS) group in this area has been the development of a fast approach for image segmentation using sequential processing. The PhD Thesis of Rebeca Marfil describes an approach which combines regular and irregular pyramids to achieve this goal: the Bounded Irregular Pyramid. This pyramid has been widely applied inside the ISIS group to perform a task-oriented segmentation of video sequences. Finally, it was integrated inside a first implementation of an object-based attention mechanism which has been employed to look for visual natural landmarks (in a mobile robot navigation framework) and to detect human faces and hands (in a human-robot interaction framework). The research activities of the FQM-296 group are mainly focused on the computing, representing, and using homology in the context of digital images. Recently an EU FP7 project proposal where this group and the PRIP are included has been submitted. This project considers the problem of using Homology in the context of 4D digital imagery.

This integrated action will allow to establish a common forum of discussion where both groups can share ideas and applications. It will also allow to fund brief stays of researchers of both groups to increase the efficienty of the knowledge transfer. The final aim will be to change the object-based proposal of visual attention mechanism, integrating as central item a topology-preserving hierarchical structure. We are sure that the PhD Thesises of Antonio J. Palomino and Esther Antúnez could take special advantages of this collaboration, initiated by a stay of Esther Antúnez at PRIP (Feb.-Dec. 2009). On the other hand, the results obtained from the application of combinatorial pyramids to develop active vision systems could be published in conferences like the WAPCV or GbRPR, and in international publications like Pattern Recognition, Pattern Recognition Letters, Computer Vision and Image Understanding and International Journal of Computer Vision.

6 CAPACIDAD FORMATIVA DEL PROYECTO

(Máximo 1 página)

Debe justificarse que el equipo solicitante está en condiciones de **recibir** personal asociado a este proyecto en estancias cortas de duración inferior a un mes o estancias por un tiempo superior. Debe argumentarse la capacidad formativa del equipo en relación con el equipo extranjero. (Ej. Técnicas nuevas o complementarias a las del departamento de origen).

Debe justificarse que el equipo solicitante está en condiciones de **enviar** personal asociado a este proyecto en estancias cortas de duración inferior a un mes o estancias por un tiempo superior. Debe argumentarse la capacidad formativa del equipo extranjero en relación con el grupo español. (Ej. Técnicas nuevas o complementarias a las del departamento de origen).

Along its history, the Spanish group has supervised the final year projects for over a hundred undergraduate students. These are required by our Universities at Malaga or Seville to obtain the degree of Master and some of them have received prize awards from institutions such as The Spanish Institute of Telecommunication Engineers, Nortel Co. (Canada), the Spanish Institute of Technical Engineers of Telecommunication and the Andalusian Association of Telecommunication Engineers.

Should this project be funded, the Spanish research groups have the required equipment and enough physical space to for the potential scholarship holders. Both groups are proud of its past success in researcher training. Since 1994, 29 PhDs have been awarded to students supervised by members of the TIC-125 and FQM-296 groups. Besides, both groups have been funded in the last years from different projects, allowing them to have an interesting background in the fields of robotics and vision systems, and algebraic topology. Stays of members of the PRIP group at Málaga will allow them to integrate topology-preserving hierarchies inside vision systems which run over autonomous robots.

On the other hand, the PRIP group has a vast experience in the research field of hierarchical representations. Brief stays will allow to the members of the Spanish group to improve their knowledge about this kind of structures and its extension to deal with higher dimensional domains (2D+time or 3D+time). This collaboration will be specially useful for those members which are currently working on their PhD Thesis (J.P. Bandera, Helena Molina-Abril, Antonio J. Palomino and R. Vázquez-Martín).