Application

2015-03-31

2015-04965 **Pauwels, Karl** **NT-2**

Information about applicant

Name: Karl Pauwels

Doctorial degree: 2008-03-31

Birthdate: 19770125

Academic title: Doktor

Gender: Male

Employer: Kungliga Tekniska högskolan

Administrating organisation: Kungliga Tekniska högskolan

Project site: CVAP, Datorseende och robotik

Information about application

Call name: Forskningsbidrag Stora utlysningen 2015 (Naturvetenskap och teknikvetenskap)

Type of grant: Projektbidrag

Focus: Unga forskare

Subject area:

Project title (english): DISTRACT: Distributed Active Perception for Dexterous Object Manipulation

Project start: 2016-01-01

Project end: 2019-12-31

Review panel applied for: NT-2

Classification code: 10207. Datorseende och robotik (autonoma system)

Keywords: computer vision, robotics, machine learning

Funds applied for

Year:	2016	2017	2018	2019
Amount:	1,144,000	1,185,000	1,273,000	1,370,000

Descriptive data

Project info

Project title (Swedish)*

DISTRACT: Distribuerad Aktiv Visuell Förståelse för Flink Objekt Manipulering

Project title (English)*

DISTRACT: Distributed Active Perception for Dexterous Object Manipulation

Abstract (English)*

The ability to manipulate realistic objects with high dexterity is critical for next generation robotic systems. It enables a transition from the constrained industrial settings we see today towards unstructured domestic and natural environments. To achieve this, a robot needs to autonomously extract physical properties of objects and monitor how they are affected by manipulation. This requires a tight coupling between enhanced visual perception and action since interaction is typically required to obtain this information. To avoid occlusions and increase precision, it is also beneficial to involve multiple cameras in the process. Current solutions need to severely restrict the object complexity to achieve the real-time performance required for active exploration.

Fueled by the mobile and gaming industries, advanced cameras and efficient massively parallel processing platforms have become widely available recently. When combined, they can form smart sensors that can not only process the camera information with advanced computer vision algorithms, but can also simulate the appearance and physical behavior of objects in a very realistic manner.

The main objective of this project is to develop a scalable distributed active perception framework that enables multiple such smart sensors to collaboratively extract rich dynamic information from realistic objects in unstructured environments. Our goal is to develop a theoretical communication and learning framework that can be used for distributed on-line robot calibration as well as active state estimation of multiple rigid, articulated and deformable objects.

To achieve this we will combine and enhance methods from the fields of biological computer vision, computer graphics, and physics simulation. We will implement these on commodity massively parallel hardware architectures to enable a continuous real-time interaction between visual simulation and visual perception. We will combine flexible representations with machine learning clustering algorithms to establish the kinematic and dynamic properties of realistic objects in a distributed manner. We will use distributed Monte-Carlo techniques to involve multiple sensors in the evaluation of the information gain of exploratory actions.

The project will span four years and is organized according to three milestones. The system setup and on-line distributed robot calibration will complete by month 18. Once the basic multi-sensor communication framework has been established and accurate measurements can be obtained, the initial learning task will focus on modeling the kinematics of passively manipulated articulated objects, and objects interacting with each other. We plan to realize this by month 30. In the final stage of the project, the learning component will be enhanced to model object deformations as well and to generate the active exploration primitives that optimize and enhance perception.

Each milestone will be evaluated using a humanoid robotic platform complemented with commodity hardware executing realistic domestic manipulation tasks that increase in complexity from milestone to milestone. The methods developed in this project will be made available to the robotics research community as open source extensions to our software framework SimTrack, which will provide the tools to genuinely undertake the combined study of perception and action in realistic dynamic situations.

The results of this project will contribute to a new generation of manipulation robots that can be applied in less constrained industrial settings, such as electronics assembly and logistics, but also in the service industry, medical diagnostics, high-tech organic agriculture, etc.

Popular scientific description (Swedish)*

Vi ser för närvarande en snabb utveckling inom robotik som förväntas leda till stora förbättringar för industrins produktivitet, jordbruk, livskvalitet, katastrofhantering, etc. Dagens autonoma robotsystem agerar antingen efter en förprogrammerad plan som inte kräver någon form av visuell information. Om visuell data använd så kräver det speciella markörer placerade på de föremål vilket roboten skall interagerar med. Att gå från dessa kraftigt begränsade miljöer till ostrukturerade naturliga miljöer kommer kräva av robotar, en färdighet att förstå fysiska egenskaper för okända föremål. Utan det är det kommer det inte att vara möjligt för robotar att på ett precist sett manipulera föremål i sin omgivning.

Tack vare nya framsteg från mobil och tvspelsindustrin har vi nu tillgång till kostnadsefektiva hårdvaruplatformar med möjligheten att analysera kamerainformation och simulera visuella/fysiska egenskaper på ett mycket realistiskt vis. I detta projekt föreslår vi att kominera dessa analys och simuleringsförmågor för att producera smarta sensorer som kan lära sig och förstå fysiska egenskaper av föremål som en robot interagerar med. Vi kommer att utrusta en robot med flerea sådana sensorer och undersöka hur dessa kan samarbeta för att ge roboten en bättre bild av miljön runt roboten.

Resultaten av detta projekt förväntas accelerera utvecklingen av nästa generations manipulatorrobotar men kommer också att vara till nytta för andra användningsområden där avancerade algoritmer för datorseende används, så som exempelvis styrning av drönare och videoövervakning.

Project period

Number of project years*

4

Calculated project time*

2016-01-01 - 2019-12-31

Deductible time

Deductible time

Cause	Months
Career age: 84	

Career age is a description of the time from your first doctoral degree until the last day of the call. Your career age change if you have deductible time. Your career age is shown in months. For some calls there are restrictions in the career age.

Classifications

Select a minimum of one and a maximum of three SCB-codes in order of priority.

Select the SCB-code in three levels and then click the lower plus-button to save your selection.

SCB-codes*	1. Naturvetenskap > 102. Data- och informationsvetenskap (Datateknik) > 10207. Datorseende och robotik (autonoma system)

Enter a minimum of three, and up to five, short keywords that describe your project.

Keyword 1*

computer vision

Keyword 2*

robotics

Keyword 3*

machine learning

Keyword 4

Keyword 5

Ethical considerations

Specify any ethical issues that the project (or equivalent) raises, and describe how they will be addressed in your research. Also indicate the specific considerations that might be relevant to your application.

Reporting of ethical considerations*

The project does not raise any ethical issues.

The project includes handling of personal data

No

The project includes animal experiments

No

Account of experiments on humans

No

Research plan

Research Plan
Distributed Active Perception for Dexterous Object Manipulation – DISTRACT

1 Purpose and Aims

We are witnessing a rapid progress in robotics that is expected to enable a transition from the constrained industrial settings in which robotic systems currently operate towards unstructured domestic settings and natural environments. This evolution is driven by industrial and societal needs[1] and enabled by the simultaneous commoditization of three technologies: 1) standardized robotic platforms; 2) high-resolution color and range sensors; 3) power-efficient massively parallel hardware architectures.

The economic potential can hardly be overstated and major improvements are expected in manufacturing productivity with reduced risks to human operators, in the service and medical domains, in quality-of-life through domestic innovations, and in disaster response through rescue robotics. A common requirement in all these areas is the ability to manipulate realistic objects with high dexterity. To achieve this, a robot needs to autonomously extract physical properties of these objects and monitor how they are affected by manipulation.

Although vision is the most natural modality to approach this, dynamic properties are often hard to observe from a single viewpoint since interaction is typically required to obtain such information. For example, a camera mounted close to the manipulator does not provide a scene overview, whereas an overview sensor, for example mounted on the robot head, does not provide the details required at the manipulation and is frequently occluded. It seems beneficial to exploit multiple sensors together, and coordinate their position to extract the information required for manipulation. See for example the top row of Figure 1B-D where three robot-mounted cameras together provide all the information required to execute a stacking task.

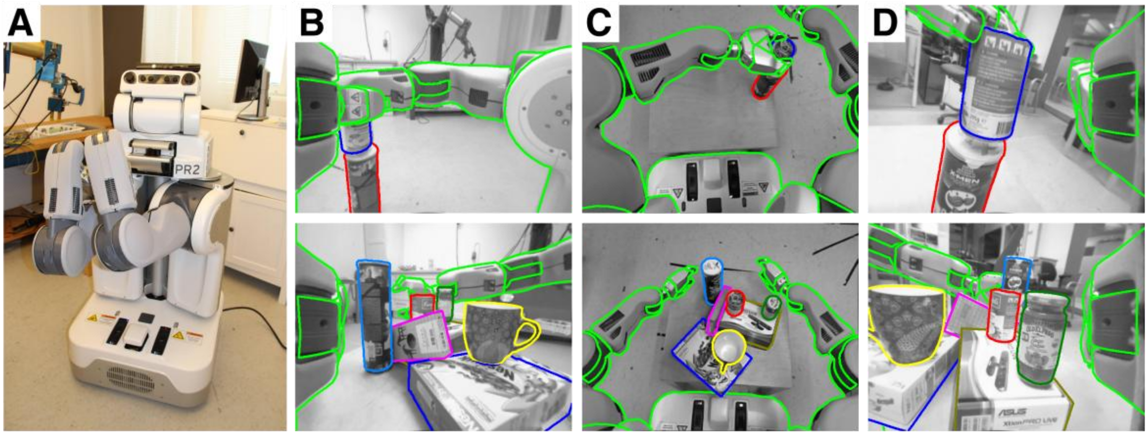

Figure 1: *(A) The PR2 robot at KTH and images recorded with the (C) head and (B,D) arm cameras in a stacking scenario (top B-D) and cluttered scene (bottom B-D). The robot state, as derived from the motor signals and forward kinematics, and the estimated position and orientation of the objects are shown in color.*

Apart from being less affected by occlusions, multiple cameras can together also more accurately localize objects using triangulation. A number of problems arise due to the massive amount of data that is generated by each camera and needs to be exchanged to calibrate and coordinate the cameras with respect to each other. Current approaches are unable to handle this in a reasonable amount of time, which is critical for dynamic manipulation and when working with humans. When humans have to wait for the machine, this is not only inefficient but it can also harm the relationship between human and machine.

---

[1] http://www.eu-robotics.net/

We propose to leverage the commoditization of mobile sensor and processing platforms and combine them into *smart sensors* that can do more than passively relay camera information. Instead they are capable of combining complex vision algorithms with graphics and physics simulations.

The main objective of this project is to develop a **scalable distributed active perception framework** that enables **multiple smart sensors** to collaboratively extract rich **dynamic information** from **realistic objects** in **unstructured environments**.

To achieve this we will design and implement a novel communication framework to allow multiple vision and simulation engines to collaboratively explore a scene in an efficient manner. We will demonstrate the feasibility and scalability of the approach by evaluating this framework in realistic domestic manipulation scenarios such as picking and placing objects, cooking, etc., performed by a humanoid robot equipped with multiple active cameras, see Figure 1A. We will use methods from biologically-inspired computer vision, computer graphics, screw theory [2] (for representing kinematics and dynamics), physics simulation, and machine learning to establish the static and dynamic properties of realistic objects in a distributed manner. A detailed multi-sensor scene representation will be combined with distributed sampling methods to evaluate the information gain of active exploration primitives on the understanding of the scene, and select the next best action.

1.1 Basic Idea and Scientific Challenges

Vision can provide extremely rich information about complex objects. Consider the local motion pattern that emerges when a loaf of bread is poked with a stick, as visualized in Figure 2B. Unlike touch sensors, vision provides information in a large region around the contact point, and thus requires minimal interaction with the object. By condensing the massive information generated by the vision sensor at each video frame, valuable information can be extracted from how the contact energy dissipates over the object, for example to identify object regions with different degrees of hardness.

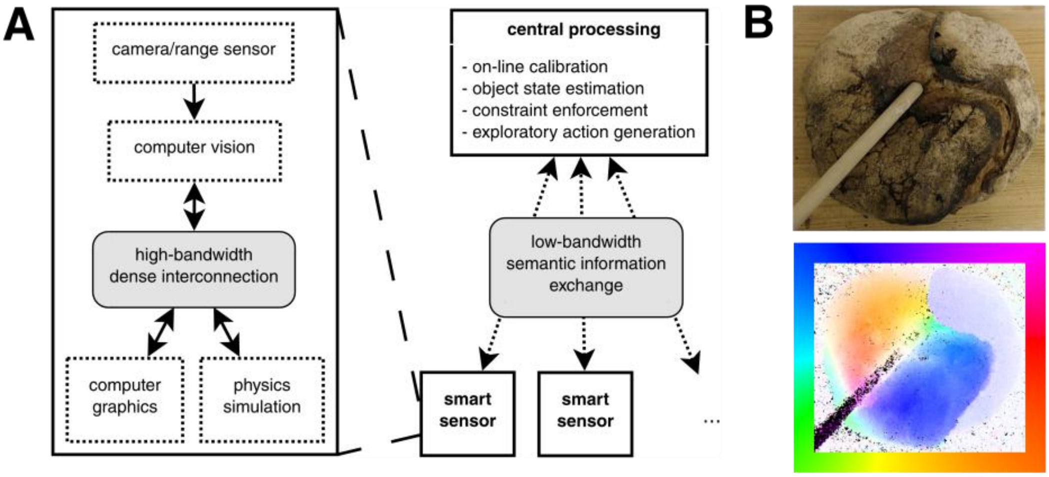

Figure 2: *(A) Framework overview diagram and (B) complex local motion pattern observed when poking a loaf of bread highlighting regions with different physical properties. The color wheel at the border of the bottom image in (B) indicates the direction of local motion at each pixel and the brightness corresponds to the magnitude of local motion at that pixel (best viewed in color).*

The main scientific challenge is concerned with condensing the dense visual information to a degree that can be efficiently communicated between smart sensors while remaining sufficiently rich to allow for object state estimation and active exploration. Since dynamic interaction with the scene is essential, the realization as a real-time system also constitutes a significant technical challenge.

As illustrated in Figure 2A, we suggest to equip each camera with a mobile processing platform that not only extracts low-level visual features, such as the dense local motion shown in Figure 2B

---

[2] R. M. Murray and S. S. Sastry, A mathematical introduction to robotic manipulation. CRC press, 1994.

(bottom), but also manages a local simulation of the appearance, geometry, and physical properties of all relevant entities that participate in the scene, typically the manipulated objects and the robot itself. The locality of the different computational modules enables high-bandwidth interaction between the components which allows for a continuous matching between visual perception and simulation. Information can then be exchanged efficiently between multiple such smart sensors in terms of the drivers of the simulations. A central mechanism can accumulate this condensed information and use it to extract global constraints related to the camera viewpoints and the kinematics and deformation properties of the objects. In turn it can use the efficient communication to enforce these learned constraints across all smart sensors, in this way reducing the uncertainty associated with the low-level sensory information.

We aim at determining a suitable representation for this information exchange with the appropriate level-of-detail required to 1) continuously calibrate the sensors with respect to each other, 2) efficiently communicate the observed object properties between each smart sensor to arrive at a global interpretation of the state of the scene, and 3) to use it to guide the exploration of unknown properties. This will be realized by simultaneously achieving the following objectives:

1) **Overcoming the multi-sensor-semantic bottleneck.** We will develop a closed-loop distributed active perception framework for multiple smart sensors that combines localized high-bandwidth communication and processing with centralized constraint enforcement to efficiently couple low-level vision with high-level semantics related to physical properties.
2) **Collaborative learning of dynamic object properties using active exploration primitives.** The unknown static and dynamic properties of complex objects will be extracted by refining the coupling between vision and physics in response to selective interaction with the object. Potential actions will be generated locally and evaluated globally.
3) **Realization as a real-time open-source software framework.** The system will be developed using high-level programming languages for graphics processing units (GPUs), open source robotics software, and physics simulation frameworks. The smart sensors will consist of commodity mobile CPU/GPU processing boards.

2 Survey of the Field

2.1 Sensory-semantic Bottleneck

Much progress has been made recently on the efficient extraction of semantic information from dense visual input using deep convolutional neural network architectures. [3] These networks consist of a number of interconnected trained layers that extract increasingly more abstract features the further they are located down the network. Human visual processing is commonly separated into two pathways, the ventral and dorsal streams.[4] Current computer vision work on deep learning is focused almost entirely on object or scene classification, which is related to the ventral stream. This stream is concerned with analyzing rather than interacting with the scene, and is thus not sufficient for this project. The second pathway, the dorsal stream, has received less attention. It is concerned with attending to and localizing objects, and extracting the precise depth and motion information required for interaction. Recentely, a real-time motion processing framework based on a similar convolutional architecture[5] was introduced that allows dense feedforward and feedback interaction with higher-level processes, as observed in neurophysiological studies.[6]

To extract semantic information with a degree of precision suitable for object manipulation, it is

---

[3] Y. Bengio, Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009.

[4] M. A. Goodale and A. D. Milner, Separate visual pathways for perception and action, *Trends in Neurosciences*, vol. 15, no. 1, pp. 20-25, 1992.

[5] K. Pauwels, N. Kruger, M. Lappe, F. Worgotter, and M. M. Van Hulle, A cortical architecture on parallel hardware for motion processing in real time, *Journal of Vision*, vol. 10, no. 10:18, 2010.

[6] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A.J. Rodriguez-Sanchez, L. Wiskott, Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.8, pp.1847-1871, 2013.

essential to calibrate the different robot sensors to each other and to the actuators. For a complex robot such as the PR2 shown in Figure 1, this requires a slow and tedious process[7] where large amounts of data are exchanged. Due to model inaccuracies and time-varying properties of the robot, it is in fact infeasible to obtain high precision in this process, especially in a scenario with multiple moving sensors. Therefore, calibration should ideally be undertaken continuously.[8]

Once calibrated, the robot's sensors can be used to extract semantic information from the scene. In this regard, a great deal of work is concerned with model-based object pose estimation using geometry-based, static visual cues such as depth, color, or edges[9],[10]. These approaches have to resort to strong assumptions, such as dynamical models or a single-actor hypothesis,[11] and/or stochastic optimization methods[12] in order to scale to scenarios with many objects and/or multiple cameras observing the scene. Motion information can help in this regard, but is rarely used in real-time systems due to the computational complexity of motion estimation.

The accuracy and robustness of pose estimation can greatly benefit from the use of multiple cameras, as demonstrated recently on hand-pose estimation,[13] but without real-time constraints. In a real-time setting, multi-camera position estimation has led to impressive results on quadrotor swarm control.[14] In that work, a professional fiducial-based (using artificial reflective marker) tracking system was used to estimate the position of the drones.

It has been shown how a redundant representation of pose based on elements from screw theory, and situated in between signal- and symbol-level, can be used to efficiently communicate detailed information between multiple cameras.[15] In our preliminary studies, see Section 5, we have shown how such an approach can be used to successfully estimate object poses using multiple sensors while satisfying real-time constraints. We will apply similar principles to extend this representation to the more detailed physically-based simulations employed in this project. Especially motion information has the potential here to greatly enrich the communication between smart sensors with information related to the kinematic and dynamic properties of objects.

Progress beyond the state of the art*: Our goal is to develop a scalable framework for the extraction and exchange of semantic information related to dynamic object state and robot calibration between multiple combined vision/simulation-engines.*

2.2 Vision-physics Coupling and Active Object State Discovery

Physics simulations have been ignored traditionally in computer vision since the engines require detailed information about object properties that is very difficult to extract due to the ill-posed nature of early vision, which is confronted with a number of inverse problems. Recently much interest has arisen in this field[16],[17],[18],[19]. Most of the physics engines used are complex and operate

---

[7] V. Pradeep, K. Konolige, and E. Berger, Calibrating a multi-arm multi-sensor robot: A bundle adjustment approach, *Experimental Robotics*, vol. 79, pp. 211–225, 2014.

[8] N. T. Dantam, H. B. Amor, H. I. Christensen, and M. Stilman, Online multi-camera registration for bimanual workspace trajectories, *Humanoids*, 2014.

[9] A. Collet, M. Martinez, and S. S. Srinivasa, The MOPED framework: Object recognition and pose estimation for manipulation, *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1284–1306, 2011.

[10] A. Aldoma, F. Tombari, J. Prankl, A. Richtsfeld, L. Di Stefano, and M. Vincze, Multimodal cue integration through hypotheses verification for RGB-D object recognition and 6DOF pose estimation, *International Conference on Robotics and Automation*, pp. 2104–2111, 2013.

[11] N. Kyriazis and A. Argyros. Physically plausible 3D scene tracking: The single actor hypothesis, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9–16, 2013.

[12] I. Oikonomidis, M. Lourakis, and A. Argyros, Evolutionary quasi-random search for hand articulations tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3422–3429, 2014.

[13] L. Ballan, A. Taneja, J. Gall, L. Van Gool, and M. Pollefeys, Motion capture of hands in action using discriminative salient points, *European Conference on Computer Vision*, pp. 640-653, 2012.

[14] D. Mellinger, N. Michael, and V. Kumar, Trajectory generation and control for precise aggressive maneuvers with quadrotors, *The International Journal of Robotics Research*, vo.31, no.5, pp.664-674, 2012.

[15] T. Drummond and R. Cipolla, Real-time visual tracking of complex structures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 932–946, 2002.

[16] N. Kyriazis, A. Argyros, see n.11

[17] M. Salzmann, and R. Urtasun, Physically-based motion models for 3D tracking: A convex formulation, *International Conference on Computer Vision*, pp.2064-2071, 2011.

on the model's vertices and faces[20],[21] which greatly limits the geometric complexity of objects that can be considered. Many commercial techniques have become available recently to easily extract detailed models of the static geometry and appearance of complex objects.[22]

Fueled by the rapid progress in real-time computer graphics, another approach to physics simulation that uses particles as elementary units has been introduced recently.[23] This approach can handle rigid bodies, deformations, fluids, and gases in a uniform way. It does not aim at complete physical realism, but rather at generating visually plausible results, which is exactly what is required to obtain a coupling between physics and vision. Our proposal to combine particle-based physics with our dense computer vision engine is truly innovative and has the potential to substantially increase the complexity of the object types and interactions considered while at the same time simplifying the vision-physics coupling.

It has long been known that perception and action are intertwined and therefore should not be studied in isolation. Existing work on active vision focuses on the control of eye movements,[24] visual servo-control,[25] and more recently on object discovery.[26] Apart from binocular vergence control, multiple cameras are only recently being used in more sophisticated active vision systems, such as visual servoing.[27] We will use multiple cameras on the manipulators to optimize the active exploration of complex objects by considering the discovery of kinematical constraints and deformation modes. Kinematical constraints can arise temporarily for example when objects come in contact with each other or the robot, which can be signaled by the physics engine and modeled as a virtual joint, or when for example moving objects on a table, which can be modeled as a planarity constraint. Much work exists on kinematics discovery in passive[28] and active settings,[29] and on passive deformation modeling.[30] These approaches estimate articulated or deformation modes by factorizing or clustering feature trajectories established over large time spans. Recently, physics simulations have also been used in active deformation scenarios,[31] but the analysis had to be performed off-line due to the computational complexity. We intend to go beyond this by exploiting dense motion coupled with a particle-based physics engine. Since dense motion is a spatially rich cue, this approach allows exchanging time for space, and can thus reduce the processing latency which is critical for dynamic interaction.

The final active component considered in this project is concerned with next best view planning,[32] *i.e.* how to move the camera to obtain maximally useful information. Most recent work

---

[18] D.J. Duff, T. Morwald, T., R. Stolkin, and J. Wyatt, Physical simulation for monocular 3D model based tracking, *IEEE International Conference on Robotics and Automation*, pp.5218-5225, 2011.

[19] B. Zheng, Y. Zhao, Joey C. Yu, K. Ikeuchi, and S.-C. Zhu, Beyond point clouds: Scene understanding by reasoning geometry and physics, *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[20] P. Jia, S. Chitta, D. Manocha, FCL: A general purpose library for collision and proximity queries, *IEEE International Conference on Robotics and Automation*, pp.3859-3866, 2012.

[21] E. Coumans, Bullet game physics simulation, 2011.

[22] http://www.123dapp.com/

[23] M. Macklin, M. Müller, N. Chentanez, and T.-Y. Kim. Unified particle physics for real-time applications. *ACM Transactions on Graphics*, vol. 33, no. 4:153, 2014.

[24] T. Uhlin, P. Nordlund, A. Maki, J.-O. Eklundh, Towards an active visual observer, *International Conference on Computer Vision*, pp. 679-686, 1995.

[25] D. Kragic and H. I. Christensen, Survey on visual servoing for manipulation, *Computational Vision and Active Perception Laboratory, Technical Report*, 2002.

[26] Niklas Bergström, Carl Henrik Ek, Mårten Björkman, Danica Kragic, Scene understanding through autonomous interactive perception, *International Conference on Computer Vision Systems*, pp. 153-162, 2011.

[27] O. Kermorgant and F. Chaumette, Multi-sensor data fusion in sensor-based control: Application to multi-camera visual servoing, *IEEE International Conference on Robotics and Automation*, pp. 4518–4523, 2011.

[28] S. Pillai, M. R. Walter and, S. Teller, Learning articulated motions from visual demonstration, *Robotics: Science and Systems*, Berkeley, 2014.

[29] D. Katz and O. Brock, Manipulating articulated objects with interactive perception, *IEEE International Conference on Robotics and Automation*, pp. 272-277, 2008.

[30] L. Torresani, A. Hertzmann, and C. Bregler, Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no.5, 2008.

[31] B. Frank, C. Stachniss, R. Schmedding, M. Teschner, W. Burgard, Learning object deformation models for robot motion planning, *Robotics and Autonomous Systems*,vol.62,n.8, pp. 1153-1174, 2014.

[32] C. Connolly, The determination of next best views, *IEEE International Conference on Robotics and Automation,*

focuses on 3D scene or object reconstruction[33] and does not consider dynamic settings, where objects are interacting with each other and the robot or deforming. Physics engines can be useful in this regard and graphics simulations readily provide occlusion information. We intend to combine them to determine a viewpoint, or in a dynamic setting a camera motion, that will maximize the information gain.

Progress beyond the state of the art*: We will develop a tight real-time coupling between a massively parallel particle-based physics engine and a massively parallel dense low-level-vision engine and use dynamic mid-level features to enable multiple such vision/physics components to collaboratively explore a dynamic scene.*

2.3 Realization as an Interactive Perception Framework

The computer vision problem is circumvented or simplified in most robotics work by attaching fiducial markers to the objects of interest. These can then be tracked efficiently using open source frameworks such as ARToolkit[34] or professional multi-camera systems that operate in dedicated rooms. Due to the dynamic nature of robotics research, real-time operation is crucial, and more general vision systems are not yet able to deliver this performance.

Some frameworks have emerged recently, with notable examples BLORT[35] and ViSP.[36] Our own contribution, SimTrack[37], a simulation-based framework for object pose estimation, tightly integrates GPU-accelerated dense depth and motion processing with graphics simulation to allow real-time multi-rigid object pose estimation using multiple cameras. SimTrack enables sharing the first stages of the visual processing hierarchy between ventral and dorsal processing which is not only in accordance with biology,[38] but also allows leveraging technological progress fueled by the recent breakthroughs in deep learning. Of particular relevance are the architectural design choices made for upcoming GPU architectures such as NVIDIA Pascal which is claimed to provide a ten-fold speedup in deep-learning applications.[39]

Recently the high performance particle-based physics simulation engine Flex has become available as part of NVIDIA's PhysX framework[40]. The latter has been open-sourced through Epic’s Unreal Engine 4,[41] which should facilitate development in this project. In terms of control and distributed communication, the robotics operating system[42] (ROS) is ideally suited. It incorporates Moveit![43] for real-time motion planning.

Progress beyond the state of the art*: Our scalable architecture exploits dense motion cues and will enable the combined use of many commodity smart sensors, that can for example be used to tile one or multiple robots with cameras.*

3 Project Description

Although we are primarily interested in the theoretical nature of the outlined objectives, due to the inherently dynamic nature of active exploration, a real-time implementation is crucial as well.

---

vol.2, pp.432-435, 1985.

[33] M. Krainin, B. Curless, D. Fox, Autonomous generation of complete 3D object models using next best view manipulation planning, *IEEE International Conference on Robotics and Automation*, pp.5031-5037, 2011.

[34] H. Kato and M. Billinghurst, Marker tracking and HMD calibration for a video-based augmented reality conferencing system, *International Workshop on Augmented Reality*, 1999.

[35] T. Morwald, J. Prankl, A. Richtsfeld, M. Zillich, and M. Vincze, BLORT - The blocks world robotic vision toolbox, *IEEE International Conference on Robotics and Automation Workshops*, 2010.

[36] E. Marchand, F. Spindler, and F. Chaumette, ViSP for visual servoing: a generic software platform with a wide class of robot control skills, *IEEE Robotics and Automation Magazine*, vol. 12, no. 4, pp. 40–52, 2005.

[37] http://www.karlpauwels.com/simtrack/

[38] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A.J. Rodriguez-Sanchez, L. Wiskott, see n.6

[39] J.-H. Huang, Keynote, *NVIDIA GPU Technology Conference*, 2015.

[40] https://developer.nvidia.com/physx-flex

[41] https://www.unrealengine.com/what-is-unreal-engine-4

[42] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R.Wheeler, A. Y. Ng, ROS: an open-source Robot Operating System, *ICRA workshop on open source software*, 2009.

[43] I. A. Sucan and S. Chitta, MoveIt!, http://moveit.ros.org

For example, while poking or stretching an object continuous monitoring is required to prevent damage to the object. The workplan and milestones are developed around realistic scenarios, such as stacking objects of various nature and cooking tasks. These will be evaluated using the humanoid robotic platform shown in Figure 1A complemented with a set of commodity mobile processors. More details of the experimental setup can be found in Section 8. Ground-truth benchmarking datasets will be recording using KTH's Optitrack[44] motion capture room, a professional multi-camera system for tracking reflective markers. All software components will be released as open source ROS modules that extend our software framework SimTrack.

The main goal is to show that the same general theoretical framework can be used for distributed on-line calibration and active state estimation of multiple rigid, articulated and deformable objects.

Each smart sensor will tightly couple a biologically-inspired low-level vision component with a graphics engine and a particle-based physics simulator. All three massively parallel components are executed on a single mobile GPU allowing for high-bandwidth communication. The semantic representation used to coordinate different sensors and to describe novel object characteristics will be developed using kinematics and dynamics elements from screw theory,[45] and using shape-primitives to model object deformation.[46] Multiple such primitives may be required to describe complex objects with varying degrees of deformability.

A strength of the project is the use of commodity hardware and open source software which will allow wide dissemination of the results and facilitate comparisons and contributions from the larger community. We will structure the research according to the main objectives. Each milestone will deliver on all the objectives, but the task complexity will gradually increase from milestone to milestone. This approach facilitates contingency planning since problems are signaled earlier. At each milestone the extensions will be released through our open source software framework SimTrack to ensure rapid dissemination.

3.1 Milestone 1: System Setup and On-line Distributed Calibration (Month 18)

We will first combine our previous works on real-time articulated and multi-camera rigid pose estimation, see Section 5, and apply it to the robot itself in order to achieve continuous self-calibration. This will provide the high degree of accuracy required in subsequent parts of the project.

We will also establish the connection between dense vision, graphics, and the particle-based physics simulation for the rigid body case considering a single camera. The simulated appearance, geometry, and motion generated by the graphics engine are sufficiently accurate to directly interact with the low-level vision component.

We will then extend our preliminary results on multi-camera fusion in the rigid tracking case with physics. Concretely this will allow incorporating a prediction component and enforcing non-penetration constraints. We will use elements from screw theory here to construct a suitably compact semantic representation that allows communicating object dynamics between the smart sensors.

***Measure of success**: On-line distributed calibration on the PR2 robot and distributed multi-rigid object tracking incorporating rigid body physics as applied to pick-and-place and rigid object stacking tasks.*

3.2 Milestone 2: Passive Kinematics Modeling (Month 30)

With the continuous accuracy refinement in-place, we can start using the measurements to gather information about the physical properties of the objects under consideration. Specifically, what constitutes an object in a multi-object scene and/or which kinematic constraints apply to an articulated object. Other examples that can be expressed with the same framework are certain

---

[44] https://www.optitrack.com/

[45] R. M. Murray and S. S. Sastry, see n. 2

[46] M. Müller, B. Heidelberger, M. Teschner, and M. Gross. Meshless deformations based on shape matching. *ACM Transactions on Graphics*, vol. 24, pp. 471-478, 2005.

functional relations between objects, such as an object supporting another, or an object restricted to move on a plane. Once modeled, such constraints can be exploited to reduce uncertainty and ambiguity induced by the sensory information. To achieve this we will extend the now densely coupled vision and physics components with a learning component, *i.e.* a clustering component that operates on the observed dense motion patterns. Initially, we will manually indicate the exploratory actions. We will use off-the-shelf techniques to determine the appearance and geometry of the object(s) prior to manipulation.

Measure of success*: Kinematics discovery of articulated objects demonstrated on simple tasks such as opening and closing a shelf, box or door, and cardboard folding.*

3.3 Milestone 3: Active Deformation Discovery (Month 48)

We will extend the learning component from modeling articulated object parts and links, to identifying deformation clusters with associated deformation complexity. At this point we will also identify exploration strategies using active exploration primitives such as poking and stretching, and viewpoint determination. These will be guided by a representation on the object's surface of the uncertainty of our belief of the object state. Using Monte-Carlo techniques, each sensor will then locally generate and simulate proposal primitives based on its capabilities, in terms of allowable actions, and its need for information to reduce object state uncertainty. The graphics simulations will provide accurate information on occlusions at this stage. The most suitable action will then be selected centrally by considering all the generated alternatives and respecting global constraints, such as collision avoidance.

Measure of success*: Passive and active deformation discovery of deformable objects with application to an autonomous demonstration task such as stacking deformable objects, cutting food, decorating pizza or cake, evaluating food freshness, etc.*

4 Significance

The waves of innovation seen recently in manufacturing towards less constrained industrial applications, such as electronics assembly and logistics, are expected to arrive soon in the service industry and medical diagnostics, high-tech organic agriculture, etc. All these applications share a less precise problem definition and require dexterous manipulation which can be enabled through enhanced perception. As a consequence DISTRACT's outcomes have the potential to impact a wide range of robotic applications by enabling a new generation of manipulation robots.

The timeliness of this work in the research community is evidenced by recent initiatives such as ICRA 2015's Amazon picking challenge and the DARPA robotics challenge, in which vision takes a central role. The community lacks the common software framework DISTRACT aims to provide, as many state-of-the art robotics papers still circumvent vision by using markers or simple colored objects. This project aims to provide the tools to genuinely undertake the combined study of perception and action in realistic dynamic situations. The use of motion as a rich visual cue opens up dynamic control of more complex interactions such as grasping, the final stages of which are still mostly executed blindly in current work. Our initial open source framework SimTrack is already in-use at KTH, the University of Edinburgh, and the University of Bremen, and broader dissemination and announcements are planned, specifically at ICRA 2015 and IROS 2015.

DISTRACT's outcomes will also directly impact industrial robotics through the PI's close collaboration with the team behind the widely-adopted motion planning library Moveit!.

5 Preliminary Results

DISTRACT will leverage results obtained by the PI during previous fellowships and preliminary work performed at the hosting institution. With regard to active vision, we can build on past experience on fixation for motion processing,[47] and binocular vergence control for depth

---

[47] K. Pauwels, M. Lappe, and Van Hulle, M.M., Fixation as a mechanism for stabilization of short image sequences, *International Journal of Computer Vision*, vol. 72, no. 1, pp. 67–78, 2007.

estimation.[48] We developed a biologically-motivated architecture for motion processing,[49] ideally suited for GPU implementation,[50] that achieves an excellent trade-off in terms of speed, accuracy, and robustness. We then added model-based elements to this architecture by tightly coupling the perception with simulation as realized through graphics rendering. This resulted in state-of-the-art performance on real-time rigid[51] and articulated[52] object pose estimation, see Figure 3A. We then applied this to tracking robot manipulators in a way that is robust to calibration inaccuracies.[53] Our most recent work has focused on using multiple moving cameras, rigidly attached to the robot, for object tracking, while accounting for occlusions due to the robot, see Figure 1. We have shown that this increases localization precision through triangulation at the model-level. The above-mentioned work is released to the public as a continuously expanding open-source software framework, SimTrack.[54]

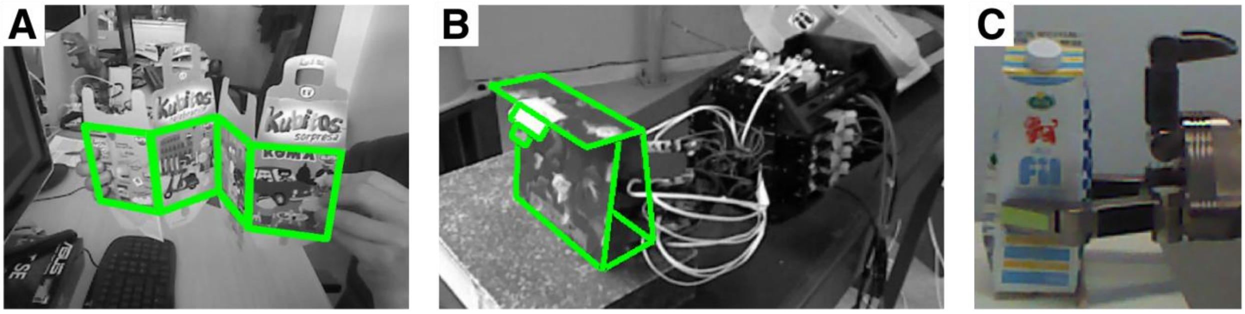

Figure 3: *(A) Tracking a complex articulated object, (B) determining deformable side panel convexity/concavity for robotic cardboard folding, and (C) exploring container contents through squeezing. The tracked object poses are shown in green in A-B.*

We also successfully used our motion processing framework for concavity/convexity estimation in the context of robotic cardboard box folding, see Figure 3B, and for actively exploring the fullness/emptiness of deformable objects through squeezing,[55] see Figure 3C. Finally, we have started a preliminary investigation on using NVIDIA’s particle-based Flex framework for deformable object state estimation. On synthetic data, we have been able to demonstrate the feasibility of determining an object’s degree of deformability by refining the shape matching parameters in the simulation.

Together, these results provide an ideal basis for the work proposed in DISTRACT.

6 Independent Line of Research

The project builds on the work performed during the PI's doctoral studies, in which the core elements of the real-time motion architecture were developed in the context of a number of European projects. The initial architecture was tuned for problems related to autonomous navigation.

---

[48] K. Pauwels, and Van Hulle, M.M., Head-centric disparity and epipolar geometry estimation from a population of binocular energy neurons, *International Journal of Neural Systems*, vol. 22, no. 3, 2012.

[49] K. Pauwels, N. Kruger, M. Lappe, F. Worgotter, and M. M. Van Hulle, see n. 5.

[50] K. Pauwels, M. Tomasi, J. Diaz, E. Ros, and M. M. Van Hulle, A comparison of FPGA and GPU for real-time phase-based optical flow, stereo, and local image features, *IEEE Transactions on Computers*, vol. 61, no. 7, pp. 999–1012, 2012.

[51] K. Pauwels, L. Rubio, J. Diaz, and E. Ros, Real-time model-based rigid object pose estimation and tracking combining dense and sparse visual cues, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2347–2354, 2013.

[52] K. Pauwels, L. Rubio, and E. Ros, Real-time model-based articulated object pose detection and tracking with variable rigidity constraints, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3994-4001, 2014.

[53] K. Pauwels, V. Ivan, E. Ros, and S. Vijayakumar, Real-time object pose recognition and tracking with an imprecisely calibrated moving RGB-D camera, *IEEE/RSJ International Conference on Intelligent Robots and Systems,* 2014.

[54] http://www.karlpauwels.com/simtrack/

[55] P. Guler, Y. Bekiroglu, K. Pauwels, and D. Kragic, What’s in the container? Classifying object contents from vision and touch, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014.

As part of completed postdoctoral fellowships funded by the Spanish (Juan de la Cierva, JCI-2011-11019) and European (Marie Curie, FP7-IEF-301144) governments, at which point the PI's research steered away from the original advisor's, the core architectural element were greatly enhanced and generalized to enable its application in object manipulation problems. The continued use and expansion of this architecture provides a strong indication for its generality and has allowed the PI to define his unique scientific area.

The preliminary work for DISTRACT was undertaken in the context of the RoboHow project (FP7-ICT-288533) at the Centre for Autonomous Systems (CAS), an inter-departmental research center at KTH, Stockholm, Sweden. The PI has added expertise that was not at the core of the group. The research planned in DISTRACT is complementary to RoboHow but the latter's focus is more on multi-modal sensor fusion than on dynamic exploration. Much of the required robotics hardware is already available at CAS and there are people working on related problems. Concretely, the PI co-supervises two doctoral students in relation to the work proposed in DISTRACT. The close interaction with the PhD student in DISTRACT will further develop the PI's supervisory skills.

7 Form of Employment

The project will be executed by the PI in the form of researcher. The funding will be used in part for his employment during the four years of the project, but his position is already financed for a large portion through the RoboHow project. In addition, one doctoral student with a background in physics simulation in relation to real-time graphics will be partly funded by the project.

8 Equipment

We will use the robotics hardware available at CAS, in particular the Willow Garage PR2 robot shown in Figure 1A. This robot is equipped with multiple camera systems on the head and near the end effectors. In addition, CAS has access to a high-precision OptiTrack motion capture system that will be used to obtain ground-truth data for object tracking. We will initially develop and test the software on high performance desktop GPU boards, a number of which have already been donated by the NVIDIA corporation. This will then be transferred to mobile CPU/GPU processing boards based on NVIDIA's Tegra architecture.[56] This transfer is straightforward since the architectures are almost identical. This approach will greatly reduce development and testing effort. These low-cost boards will then enable scaling up the project to incorporate a large number of smart sensors (less than ten).

9 International and National Collaboration

The hosting institution, CAS, has support from the Swedish and European governments. The group has participated and is participating in a number of European projects. The work in this project is complementary to work performed in the center, and specifically in the RoboHow project, although the focus there is more on sensor fusion and less on dynamic exploration. Through this project, collaborations have already been established locally with a number of doctoral and postdoctoral researchers at KTH's CAS, and internationally at the University of Bremen (Germany). Through earlier projects the PI also continues collaborations with the University of Edinburgh (United Kingdom) and Granada (Spain).

The PI is also engaged in a tight collaboration with the Moveit!-team currently located at the Stanford Research Institute, Menlo Park, US, with a focus on dynamic motion planning and its interaction with perception.

---

[56] http://www.nvidia.com/object/tegra-x1-processor.html

Interdisciplinarity

My application is interdisciplinary

☐

An interdisciplinary research project is defined in this call for proposals as a project that can not be completed without knowledge, methods, terminology, data and researchers from more than one of the Swedish Research Councils subject areas; Medicine and health, Natural and engineering sciences, Humanities and social sciences and Educational sciences. If your research project is interdisciplinary according to this definition, you indicate and explain this here.

Click here for more information

Scientific report

Scientific report/Account for scientific activities of previous project

Budget and research resources

Project staff

Describe the staff that will be working in the project and the salary that is applied for in the project budget. Enter the full amount, not in thousands SEK.

Participating researchers that accept an invitation to participate in the application will be displayed automatically under Dedicated time for this project. Note that it will take a few minutes before the information is updated, and that it might be necessary for the project leader to close and reopen the form.

Dedicated time for this project*

	Role in the project	Name	Percent of full time
1	Applicant	Karl Pauwels	30
2	Other personnel without doctoral degree	PhD Student A	80

Salaries including social fees

	Role in the project	Name	Percent of salary	2016	2017	2018	2019	Total
1	Applicant	Karl Pauwels	30	215,000	220,000	226,000	231,000	892,000
2	Other personnel without doctoral degree	PhD Student A	80	440,000	484,000	532,000	586,000	2,042,000
	Total			655,000	704,000	758,000	817,000	2,934,000

Other costs

Describe the other project costs for which you apply from the Swedish Research Council. Enter the full amount, not in thousands SEK.

Premises

	Type of premises	2016	2017	2018	2019	Total
1	Offices	79,000	85,000	91,000	99,000	354,000
	Total	79,000	85,000	91,000	99,000	354,000

Running Costs

	Running Cost	Description	2016	2017	2018	2019	Total
1	Travel costs	conferences	30,000	30,000	30,000	30,000	120,000
2	Material	computers	40,000				40,000
	Total		70,000	30,000	30,000	30,000	160,000

Depreciation costs

Depreciation cost	Description	2016	2017	2018	2019

Total project cost

Below you can see a summary of the costs in your budget, which are the costs that you apply for from the Swedish Research Council. Indirect costs are entered separately into the table.

Under Other costs you can enter which costs, aside from the ones you apply for from the Swedish Research Council, that the project includes. Add the full amounts, not in thousands of SEK.

The subtotal plus indirect costs are the total per year that you apply for.

Total budget

Specified costs	2016	2017	2018	2019	Total, applied	Other costs	Total cost
Salaries including social fees	655,000	704,000	758,000	817,000	2,934,000		2,934,000
Running costs	70,000	30,000	30,000	30,000	160,000		160,000
Depreciation costs					0		0
Premises	79,000	85,000	91,000	99,000	354,000		354,000
Subtotal	804,000	819,000	879,000	946,000	3,448,000	0	3,448,000
Indirect costs	340,000	366,000	394,000	424,000	1,524,000		1,524,000
Total project cost	1,144,000	1,185,000	1,273,000	1,370,000	4,972,000	0	4,972,000

Explanation of the proposed budget

Briefly justify each proposed cost in the stated budget.

Explanation of the proposed budget*

The PI will collaborate with the PhD student on other related projects at CVAP, such as RoboHow (FP7-ICT-288533), for the remainder of the salary costs. The travel budget is required to participate in at least one international conference each year. For equipment, we require one desktop computer and several processing boards with integrated CPU/GPU mobile processor.

Other funding

Describe your other project funding for the project period (applied for or granted) aside from that which you apply for from the Swedish Research Council. Write the whole sum, not thousands of SEK.

Other funding for this project

Funder	Applicant/project leader	Type of grant	Reg no or equiv.	2016	2017	2018	2019

CV and publications

CV

1 Higher education qualification

march 2008	Doctor in Medical Sciences, Katholieke Universiteit Leuven, Belgium
2000 – 2001	Master of Science in Artificial Intelligence, K.U.Leuven, Belgium, graduated summa cum laude
1995 – 2000	Master of Science in Commercial Engineering in Information Systems, K.U.Leuven, Belgium, graduated cum laude

2 Doctoral degree

- 2008, Doctor in Medical Sciences, K.U.Leuven, Belgium, Computational Modeling of Visual Attention: Neuronal Response Modulation in the Thalamocortical Complex and Saliency-Based Detection of Independent Motion, doctoral supervisor: Prof. Dr. Marc M. Van Hulle

3 Postdoctoral positions

2014 – present	Computer Vision and Active Perception Lab, KTH Stockholm, Sweden
2011 – 2014	Computer Architecture and Technology Dept., Univ. of Granada, Spain
2008 – 2011	Lab. voor Neuro- en Psychofysiologie, K.U.Leuven, Belgium

4 Qualification required for appointments as a docent

- The applicant plans to file the application during 2016.

5 Current position

- term of appointment: researcher with research portion: 100%

6 Previous positions and periods of appointment

2014 – present	Postdoctoral Associate	Computer Vision and Active Perception Lab, KTH Stockholm, Sweden
2011 – 2014	Postdoctoral Associate	Computer Architecture and Technology Dept., Univ. of Granada, Spain
2008 – 2011	Postdoctoral Associate	Lab. voor Neuro- en Psychofysiologie, K.U.Leuven, Belgium
2002 – 2008	Research Assistant	Lab. voor Neuro- en Psychofysiologie, K.U.Leuven, Belgium

7 Interruption in research

8 Supervision (co-supervised Ph.D. students, not in capacity as main supervisor)

08/2014 – present	Püren Güler	KTH Stockholm, Sweden
08/2014 – present	Francisco Viña	KTH Stockholm, Sweden
04/2011 – 04/2014	Leonardo Rubio	University of Granada, Spain

9 Other merits of relevance to the application

personal grants and fellowships

07/10/2013	NVIDIA Hardware Donation, 2×Geforce GTX Titan (±2000€)
01/09/2012 – 31/08/2013	Genil Start-up Project for Young Researchers, PYR-2012-9, CEI BioTIC GENIL
15/07/2012 – 15/07/2014	Marie Curie Intra-European Fellowship (IEF), PIEF-GA-2011-301144, European Commission (7th framework)
01/01/2012 – 15/07/2012	Juan de la Cierva, JCI-2011-11019, Spanish Ministry for Science and Innovation,
01/03/2011 – 31/08/2011	Genil Travel Grant for Young Talented Researchers, CEB09-0010, CEI BioTIC GENIL

awards

- Best presentation award for Pauwels, K. and Van Hulle, M.M., Real-time Independent Motion

Detection on the GPU, European Workshop on Advanced Predictive Sensor-motor Control, Judokrante, Lithuania, May 21–24 2009.

- Student paper award for Pauwels, K., Gautama, T., Mandic, D.P. and Van Hulle, M.M. Towards Model-Independent Mode Detection and Characterisation of Very Long Biomedical Time Series, The 4th Int. Conf. on Recent Advances in Soft Computing, Nottingham, UK, Dec. 12–13 2002.

participation in European research projects

- 01/02/2012 – 31/01/2016	ROBOHOW: Web-enabled and experience-based cognitive robots that learn complex everyday manipulation tasks, ICT-2011-288533
- 01/04/2011 – 31/03/2014	TOMSY: Topology based motion synthesis for dexterous manipulation, IST-2009-270436
- 01/03/2008 – 28/02/2011	EYESHOTS: Heterogeneous 3D perception across visual fragments, IST-2007-217077
- 01/02/2006 – 31/06/2009	DRIVSCO: Learning to emulate perception action cycles in a driving school scenario, IST-2002-016276
- 09/01/2005 – 31/08/2008	MCCOOP: Multi-channel cooperativity in visual processing, NEST-2003-012963
- 01/01/2002 – 31/12/2004	ECOVISION: Artificial vision systems based on early cognitive cortical processing, IST-2001-32114

software projects

- SimTrack: A simulation-based framework for scalable real-time object pose detection and tracking, open source (BSD license), available at www.karlpauwels.com/simtrack

teaching experience

- 2002 – 2010, K.U.Leuven, H02B3A Neural Computing: Laboratory Sessions

phd student supervision

- 08/2014 – present	Püren Güler	KTH Stockholm, Sweden
- 08/2014 – present	Francisco Viña	KTH Stockholm, Sweden
- 04/2011 – 04/2014	Leonardo Rubio	University of Granada, Spain

journal reviewer

ACM Trans. on Embedded Computing Systems, EURASIP Journal on Image and Video Processing, Frontiers in Behavioral Neuroscience, Frontiers in Computational Neuroscience, IEEE Trans. on Circuits and Systems for Video Technology, IEEE Trans. on Cybernetics, IEEE Trans. on Neural Networks, IEEE Trans. on Pattern Analysis and Machine Intelligence, IEEE Trans. on Systems, Man, and Cybernetics – Part C, IEEE Trans. on Very Large Scale Integration Systems, IET Computer Vision, Image and Vision Computing, Information Sciences, Journal of Real-Time Image Processing, Neurocomputing, PLoS One, Robotics and Autonomous Systems, Sensors

conference reviewer

ACM SIGGRAPH, IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE International Workshop on Machine Learning for Signal Processing, International Conference on Computer Vision Systems, IEEE/RSJ International Conf. on Intelligent Robots and Systems

journal publications under review

- K. Pauwels, L. Rubio, and E. Ros, Real-time Pose Estimation and Tracking of Hundreds of Objects, IEEE Transactions on Circuits and Systems for Video Technology
- L. Rubio, J. Diaz, E. Ros, and K. Pauwels, Articulated Object Pose Detection from Depth and Appearance by Decoupling Rigid and Non-rigid Pose Components, Computer Vision and Image Understanding
- L. Rubio, R.B. Rusu, J. Diaz, E. Ros, and K. Pauwels, Combined Planar Scene Segmentation and Particle Filtering for Real-time Rigid Object Pose Estimation, Image and Vision Computing

Publication list

Summary

A total of 14 journal publications, 13 of which are JCR publications (9 ranked Q1 and 4 ranked Q2), 6 first author JCR publications (4 ranked Q1 and 2 ranked Q2). In addition, 3 book chapters. Also 20 peer-reviewed conference proceedings, 11 of which first author.

Google Scholar Citations on March 8, 2015

Citation indices	All	Since 2010
Citations	434	353
h-index	12	11
i10-index	17	13

1. Peer-reviewed Original Articles `(IF: Impact Factor, JR: Journal Rank in Category)`

1. Indovina, I., Maffei, V., Pauwels, K., Macaluso, E., Orban, G.A., Lacquaniti, F. (2013), Simulated Self-motion in a Visual Gravity Field: Sensitivity to Vertical and Horizontal Heading in the Human Brain, *Neuroimage*, 71:114–124.
   `Q1 ; IF 2012: 6.252 ; JR: 2/14 (neuroimaging)`

2. Zhu, Q., Nelissen, K., Van den Stock, J., De Winter, F.-L., Pauwels, K., de Gelder, B., and Vandenbulcke, M. (2013), Dissimilar Processing of Emotional Facial Expressions in Human and Monkey Temporal Cortex, *Neuroimage*, 66:402–411.
   `Q1 ; IF 2012: 6.252 ; JR: 2/14 (neuroimaging)`

3. Abramov, A., Pauwels, K., Papon, J., Wörgötter, F., and Dellen, B. (2012), Real-time Segmentation of Stereo Videos on a Portable System with a Mobile GPU, *IEEE Transactions on Circuits and Systems for Video Technology*, 22(9):1292–1305.
   `Q2 ; IF 2012: 1.819 ; JR: 62/242 (engineering, electrical & electronic)`

4. Pauwels, K. and Van Hulle, M.M. (2012), Head-centric Disparity and Epipolar Geometry Estimation from A Population of Binocular Energy Neurons, *International Journal of Neural Systems*, 22(3).
   `Q1 ; IF 2012: 5.054 ; JR: 2/114 (comput. science, artificial intelligence)`

(*) 5. Pauwels, K., Tomasi, M., Díaz, J., Ros, E. and Van Hulle, M.M. (2012), A Comparison of FPGA and GPU for Real-Time Phase-based Optical Flow, Stereo, and Local Image Features, *IEEE Transactions on Computers*, 61(7):999–1012.
   `Q2 ; IF 2012: 1.379 ; JR: 14/50 (computer science, hardware & architect.)`

6. Markelić, I., Kjær-Nielsen, A., Pauwels, K., Baunegaard With Jensen, L., Chumerin, N., Vidugiriene, A., Tamosiunaite, M., Van Hulle, M.M., Krüger, N., Rotter, A. and Wörgötter, F. (2011), The Driving School System: Learning Basic Driving Skills from a Teacher in a Real Car, *IEEE Trans. on Intelligent Transportation Syst.*, 12(4):1135–1146.
   `Q1 ; IF 2011: 3.452 ; JR: 1/28 (transportation science & technology)`

7. Klette, R., Krüger, N., Vaudrey, T., Pauwels, K., Van Hulle, M.M., Morales, S., Kandil, F., Haeusler, R., Pugeault, N. and Rabe, C. (2011), Performance of Correspondence

Algorithms in Vision-Based Driver Assistance using an Online Image Sequence Database, *IEEE Transactions on Vehicular Technology*, 60(5):2012–2026.
`Q1 ; IF 2011: 1.921 ; JR: 6/28 (transportation science & technology)`

(*) 8. Pauwels, K., Krüger, N., Lappe, M., Wörgötter, F. and Van Hulle, M.M. (2010), A Cortical Architecture on Parallel Hardware for Motion Processing in Real Time, *Journal of Vision*, 10(10):18; doi:10.1167/10.10.18.
`Q1 ; IF 2010: 2.805 ; JR: 12/55 (ophthalmology)`

9. Baunegaard With Jensen, L., Kjær-Nielsen, A., Pauwels, K., Barsøe Jessen, J., Van Hulle, M.M. and Krüger, N. (2010), A Two-level Real-time Vision Machine Combining Coarse- and Fine-grained Parallelism, *Journal of Real-Time Image Processing*, 5(4):291–304.
`Q2 ; IF 2010: 0.962 ; JR: 8/19 (imaging science & photographic techn.)`

10. Sabatini, S.P., Gastaldi, G., Solari, F., Pauwels, K., Van Hulle, M.M., Diaz, J., Ros, E., Pugeault, N. and Krüger, N. (2010), A Compact Harmonic Code for Early Vision based on Anisotropic Frequency Channels. *Computer Vision and Image Understanding*, 114(6):681–699.
`Q1 ; IF 2010: 2.404 ; JR: 26/247 (engineering, electrical & electronic)`

11. Pauwels, K. and Van Hulle, M.M. (2009), Optic Flow from Unstable Sequences through Local Velocity Constancy Maximization, *Image and Vision Computing*, 27(5):579–587.
`Q2 ; IF 2009: 1.474 ; JR: 26/92 (computer science, theory & methods)`

12. Pauwels, K., Lappe, M. and Van Hulle, M.M. (2007), Fixation as a Mechanism for Stabilization of Short Image Sequences. *International Journal of Computer Vision*, 72(1):67–78.
`Q1 ; IF 2007: 3.381 ; JR: 5/93 (computer science, artificial intelligence)`

13. Pauwels, K. and Van Hulle, M.M. (2006), Optimal Instantaneous Rigid Motion Estimation Insensitive to Local Minima. *Computer Vision and Image Understanding*, 104(1):77–86.
`Q1 ; IF 2006: 1.548 ; JR: 39/206 (engineering, electrical & electronic)`

14. Wörgötter, F., Krüger, N., Pugeault, N., Calow, D., Lappe, M., Pauwels, K., Van Hulle, M.M., Tan, S. and Johnston, A. (2004), Early Cognitive Vision: Using Gestalt Laws for Task-dependent, Active Image Processing. *Natural Computing*, 3(3):293–321.

2. Peer-reviewed Conference Papers (major conferences marked in bold)

(*) 1. Pauwels, K., Ivan, V., Ros, E., and Vijayakumar, S. (2014), Real-time Object Pose Recognition and Tracking with an Imprecisely Calibrated Moving RGB-D Camera. ***IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)***, Chicago, Illinois, September 14–18 2014.

2. Pieropan, A., Salvi, G., Pauwels, K., Kjellström, H. (2014), Audio-visual Classification and Detection of Human Manipulation Actions. ***IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)***, Chicago, Illinois, September 14–18 2014.

3. Güler, P., Bekiroglu, Y., Pauwels, K., and Kragic, D. (2014), What's in the Container? Classifying Object Contents From Vision and Touch. ***IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)***, Chicago, Illinois, September 14–18 2014.

4. Pauwels, K., Rubio, L., Ivan, V., Vijayakumar, S. and Ros, E. (2014), Real-time RGB-D-based Object and Manipulator Pose Estimation. ***Robotics: Science and System: Workshop on RGB-D: Advanced Reasoning with Depth Cameras***, Berkeley, California, July 12 2014.

(*) 5. Pauwels, K., Rubio, L. and Ros, E. (2014), Real-time Model-based Articulated Object Pose Detection and Tracking with Variable Rigidity Constraints. ***IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***, Columbus, Ohio, June 24–27 2014.

6. Pieropan, A., Salvi, G., Pauwels, K. and Kjellström, H. (2014), A Dataset of Human Manipulation Actions. *IEEE International Conference on Robotics and Automation (ICRA): International Workshop on Autonomous Grasping and Manipulation – An Open Challenge*, Hong Kong, China, May 31 2014.

7. Pauwels, K. (2014), Real-time 3D Pose Estimation of Hundreds of Objects. *NVIDIA GPU Technology Conference (GTC)*, San Jose, California, March 24–27 2014.

(*) 8. Pauwels, K., Rubio, L., Díaz, J. and Ros, E. (2013), Real-time Model-based Rigid Object Pose Estimation and Tracking Combining Dense and Sparse Visual Cues. ***IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***, Portland, June 23–28 2013, pp. 2347–2354.

9. Abramov, A., Dellen, B., Pauwels, K., Papon, J. and Wörgötter, F. (2012), Depth-supported Real-time Video Segmentation with the Kinect. ***IEEE Workshop on Applications of Computer Vision (WACV)***, Colorado, January 9–11 2012.

10. Abramov, A., Aksoy, E.E., Dörr, J., Pauwels, K., Wörgötter, F. and Dellen, B. (2010), 3D Semantic Representation of Actions from Efficient Stereo-image-sequence Segmentation on GPUs. ***Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)***, Paris, May 17–20 2010.

11. Pugeault, N., Pauwels, K., Pilz, F., Van Hulle, M.M. and Krüger, N. (2010), A Three-Level Architecture for Model-Free Detection and Tracking of Independently Moving Objects. *International Conference on Computer Vision Theory and Applications (VISAPP)*, Angers, May 17–21 2010.

12. Baseski, E., Baunegaard With Jensen, L., Pugeault, N., Pilz, F., Pauwels, K., Van Hulle, M.M., Wörgötter, F. and Krüger, N. (2009), Road Interpretation for Driver Assistance Based on an Early Cognitive Vision System. *International Conference on Computer Vision Theory and Applications (VISAPP)*, Lisboa, February 5–8 2009.

13. Pauwels, K. and Van Hulle, M.M. (2008), Realtime Phase-based Optical Flow on the GPU. ***Workshop on Computer Vision on GPU (in conjunction with CVPR 2008)***, Anchorage, June 27 2008.

14. Sabatini, S.P., Gastaldi, G., Solari, F., Diaz, J., Ros, E., Pauwels, K., Van Hulle, M.M., Pugeault, N. and Krüger, N. (2007), Compact and Accurate Early Vision Processing in the Harmonic Space. *International Conference on Computer Vision Theory and Applications (VISAPP)*, Barcelona, March 8–11 2007.
15. Martens, J., Pauwels, K. and Put, F. (2006), A Neural Network Approach to the Validation of Simulation Models. *Winter Simulation Conference*, Monterey CA, December 3–6 2006.
16. Pauwels, K. and Van Hulle, M.M. (2006), Optic Flow from Unstable Sequences containing Unconstrained Scenes through Local Velocity Constancy Maximization. ***British Machine Vision Conference (BMVC)***, Edinburgh, September 4–7 2006, volume 1, pp. 397–406.
17. Pauwels, K. and Van Hulle, M.M. (2005), Robust Instantaneous Rigid Motion Estimation. ***IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***, San Diego, June 20–26 2005, volume 2, pp. 980–985.
18. Pauwels, K. and Van Hulle, M.M. (2004), Segmenting Independently Moving Objects from Egomotion Flow Fields. *Early Cognitive Vision Workshop*, Isle of Skye, May 28–June 1 2004.
19. Pauwels, K., Gautama, T., Mandic, D.P. and Van Hulle, M.M. (2002), Towards Model-Independent Mode Detection and Characterisation of Very Long Biomedical Time Series. *4th International Conference on Recent Advances in Soft Computing*, The Nottingham Trent University, December 12–13 2002, pp. 230–236.
20. Pauwels, K., Gautama, T., Mandic, D.P. and Van Hulle, M.M. (2002), Towards Characterisation of Nonstationary Time Series. *IEE Workshop on Non-Linear and Non-Gaussian Signal Processing (N2SP)*, Peebles Hotel Hydro, July 8–9 2002.

3. Monographs

1. Pauwels, K., (2008), Computational Modeling of Visual Attention: Neuronal response modulation in the Thalamocortical complex and saliency-based detection of independent motion. Ph.D. Thesis. Katholieke Universiteit Leuven, Belgium.

4. Research Review Articles

none

5. Books and Book Chapters

1. Lázár, A., Pauwels, K., Van Hulle, M.M. and Roska, T. (2009), Scene Analysis of Unstable Video Flows - Using Multiple Retina Channels and Attentional Methods, in: *Integrated Circuits, Photodiodes and Organic Field Effect Transistors* (Robert McIntire and Pierre Donnell, eds.), NovaScience: Hauppauge (NY), USA.

2. Orban, G.A., Pauwels, K., Van Hulle, M.M. and Vanduffel, W. (2005), Attentional Suppression Early in the Macaque Visual System, in: *Neurobiology of Attention* (Laurent Itti, Geraint Rees and John Tsotsos, eds.), Academic Press/Elsevier, chapter 71, pp. 429–434.
3. Pauwels, K., Gautama, T., Mandic, D.P. and Van Hulle, M.M. (2004), Towards Model-Independent Mode Detection and Characterisation of Very Long Biomedical Time Series. in: *Applications and Science in Soft Computing Series: Advances in Soft Computing* (Lotfi Ahmad and Garibaldi Jonathon M., eds.), Springer-Verlag, pp. 213–218.

6. Patents

none

7. Open Access Computer Programs

SimTrack: A simulation-based framework for scalable real-time object pose detection and tracking, open source (BSD license), available at `www.karlpauwels.com/simtrack`

7. Popular Science Articles/Presentations

none

CV

Name: Karl Pauwels	**Doctorial degree:** 2008-03-31
Birthdate: 19770125	**Academic title:** Doktor
Gender: Male	**Employer:** Kungliga Tekniska högskolan

Research education

Dissertation title (swe)

Datormodellering av Visuell Uppmärksamhet

Dissertation title (en)

Computational Modeling of Visual Attention

Organisation	Unit	Supervisor
Katholieke Universiteit Leuven, Belgium Not Sweden - Higher Education institutes	Faculty of Medicine	Marc Van Hulle
Subject doctors degree	**ISSN/ISBN-number**	**Date doctoral exam**
30105. Neurovetenskaper	9789058676771	2008-03-31

Publications

Name: Karl Pauwels	**Doctorial degree:** 2008-03-31
Birthdate: 19770125	**Academic title:** Doktor
Gender: Male	**Employer:** Kungliga Tekniska högskolan

Pauwels, Karl has not added any publications to the application.

Register

Terms and conditions

The application must be signed by the applicant as well as the authorised representative of the administrating organisation. The representative is normally the department head of the institution where the research is to be conducted, but may in some instances be e.g. the vice-chancellor. This is specified in the call for proposals.

The signature*from the applicant*confirms that:

- the information in the application is correct and according to the instructions form the Swedish Research Council
- any additional professional activities or commercial ties have been reported to the administrating organisation, and that no conflicts have arisen that would conflict with good research practice
- that the necessary permits and approvals are in place at the start of the project e.g. regarding ethical review.

The signature *from the administrating organisation* confirms that:

- the research, employment and equipment indicated will be accommodated in the institution during the time, and to the extent, described in the application
- the institution approves the cost-estimate in the application
- the research is conducted according to Swedish legislation.

The above-mentioned points must have been discussed between the parties before the representative of the administrating organisation approves and signs the application.

Project out lines are not signed by the administrating organisation. The administrating organisation only sign the application if the project outline is accepted for step two.

Applications with an organisation as applicant is automatically signed when the application is registered.