PROJECT FINAL REPORT

Grant Agreement number: 215866

Project acronym: SEARISE

Project title: Smart Eyes: Attending and Recognizing Instances of Salient Events

Funding Scheme: STREP

Period covered: from 01/03/2008 to 28/02/2011

Project co-ordinator name, title and organisation:

Dr. Marina Kolesnik, Fraunhofer

Tel: +49 -2241-14-3421

Fax: +49 -2241-14-1506

E-mail: marina.kolesnik@fit.fraunhofer.de

Project website address: http://www.searise.eu

TABLE OF CONTENTS

1 FINAL PUBLISHABLE SUMMARY REPORT

1.1 Executive summary

The SEARISE project has developed a **trinocular active cognitive visual system,** the **Smart-Eyes**, for detection, tracking and categorization of salient events and behaviours.

Smart Eyes comprises three cameras with a **wide view angle global camera** for general monitoring of events and an **active binocular stereo system** for fixation and zooming in particular salient events. The global camera is fixed and its visual field is adjusted so as to overlook the whole surveying area. Visual input from the global camera is processed in real time for instantaneous identification of most salient event. Once detected, the binocular cameras start the fixation and following its movements as long as another salient event catches cameras' attention. This triggers a saccadic move towards this new event followed by fixation and pursuit. Smart Eyes acquires video record from the global camera complemented by a high resolution video of detected salient events from the binocular cameras.

Smart Eyes operations are controlled by a **hierarchical neural architecture** in real time. Like the human brain controls the perception and action of a human using knowledge learned so far, Smart Eyes architecture triggers camera actions in response to events observed in a scene while using a statistical model learned through observations of previous events in that scene. In a sense Smart Eyes has human-like capability to **learn** from and to **self-adjust** to ever changing visual input; to **fixate** salient events and to **follow** their motion. To bias its attention Smart Eyes prioritizes between different salient events by performing **visual categorization** of salient events based on environmental context and a given policy rules. Smart Eyes magnifies the attended salient event by automatically adjusting the active cameras' zooming to display the salient action in details.
The neural architecture of Smart-Eyes implements a **computational cognitive model** of the **visual processing** replicating major principles and computational strategies found in the **mammalian visual cortex**. The architecture has hierarchical structure and embeds a pipeline of visual processing modules run in several parallel threads to reach real time operations. The architecture maintains the segregation between the form (ventral) and motion (dorsal) processing streams. The low-level and mid-level processing utilises hard-wired mechanisms found in the cortical areas V1-V2-MT. The high-level processing in the form and motion stream utilises flexible learning mechanisms working on subsequent hierarchical levels. Learning is essentially unsupervised, with only a few labelled classes added manually as task bias. Saliency map utilises statistical information derived from various features from different hierarchical levels. Fusion of the saliency and categorization maps on the top of the hierarchy is complemented by an attention strategy to provide video record of salient events most suitable for human visual observations.

Two Smart Eyes prototypes have been developed and their functionality was tested in the **long-range** scenario during football matches in ESPRIT Arena in Duesseldorf; in the short-range scenarios on the Fraunhofer campus and on the city train platform next to the Arena in Duesseldorf. Smart Eyes tests demonstrated real time capability to cope with massive video input; it operates robustly with 8 to 9 frames per second with slight variations, depending on illumination and complexity of the scene. Smart Eyes operations comprise robust detection, fixation and tracking of salient events but also visualization and storing of up to 3 video streams.

Extensive dissemination activities of the project results included publications, conference attendance, press releases and the presentation of Smart Eyes at the bi-annual trade show SECURITY ESSEN in October 2010.

1.2 Summary of the project context and objectives

The SEARISE project aims at the development of a **trinocular active cognitive visual system,** the **Smart-Eyes**, for detection, tracking and categorization of salient events and behaviours. The system ought to have human-like capability to **learn** from and **self-adjust** to ever changing visual input; **fixate** at salient events and **follow** their motion; perform **visual categorization** of salient events based on environmental context and a set of policy rules. An acting part of the visual system comprises two active stereo cameras, the binocular cameras, which like human eyes automatically fixate the salient object, follow (track) its motion, and then switch the attention to another salient location. The system performs **multi-scale recording** by zooming individual parts of attended events, which might either uncover object's identity or display its salient actions in details. The "brain" of the visual system is a **cognitive model** of **visual processing** replicating computational strategies supported by neurophysiological studies of the **mammalian visual cortex**. The Smart-Eyes system naturally combines an **engineering paradigm** for coordinated eye-like movements of the binocular cameras with an innovative **computational theory** of visual cortex. The system can respond intelligently to events happening in its field of view by continuously switching its attention to those objects or object parts exhibiting most salient actions.

SEARISE pursues three major objectives:

1. Achieve comparable to human level of performance in identification, rough categorization, fixation and pursuit of salient events in complex videos for surveillance applications.
2. Develop the Smart Eyes prototype.
3. Prove the Smart Eyes functionality in real-life environment for observation of large public spaces and restricted in-door areas.

To achieve the above major objectives SEARISE conducted technological development along several lines. **First line** included the design of the visual Smart Eyes system including the necessary hardware development with the implementation of camera control mechanisms. **Second line** concerned with the development of single visual processing modules and the integrated neural architecture for real-time visual analysis. **Third line** included the development of the integrated software framework to be run in parallel on several processing units and to perform the real time visual analysis and cameras control.

Three copies of Smart Eyes system have been built during the project with the one used for developmental and testing purposes in lab experiments, second one for the use in the long range scenario and third one the short-range scenario. Smart Eyes hardware for the long- and short-range scenarios is identical with the only difference in cameras' lenses selected to serve typical scenario distance.

To prove the Smart Eyes functionality extensive **field tests** have been conducted for the long- and short-range scenarios. The task of Smart Eyes was similar in both scenarios: based on the learned model of typical events, identify, fixate and follow most salient event and store the corresponding video record. Thus an additional and important objective of the project was to prove whether and to which extent Smart Eyes can adjust its visual analysis to cope with different type of video input. The only different feature for both scenarios was task bias, i.e. definition of priorities with regard to saliency in a given scenario. This definition came from security experts.

To **benchmark Smart Eyes performance** ground truth data based on psychophysical evidence has been collected and are used for validation of Smart Eyes. This validation checked on the **coincidence with a human detection** of behavioural changes and conspicuous events, defined as the final important objective in SEARISE.

Below we describe each of the above listed objectives in more details.

Smart Eyes system design and camera control

Design of an active vision system comprising a robotic platform complemented by software modules for disparity and motion information around the fixation point had to satisfy real time requirements on the visual processing. The control of the binocular cameras is inspired by paradigms of human vision and will be visually-driven on the basis of saliencies detected in the video stream of a cyclopean fixed camera. This implies addressing the following issues:

- Engineering of the trinocular active visual system.
- Cortical architecture for real-time local computation of disparity and motion feature maps.
- Active binocular fixation and smooth pursuit movements.
- Prototyping of the Smart Eyes system.

Visual processing modules

The goal of the ***motion estimation module*** was to provide a model for motion estimation that mimics the hierarchical architecture of the dorsal pathway in the primate visual cortex. Here, information is processed along a cascade of neural mechanisms that pool activations from the surround with spatially increasing receptive fields. Following neurophysiological findings, early levels of processing contain detectors sensitive to oriented edges and oriented motion, which are represented by model area V1. Intermediate stages of processing spatially integrate these cues with increasing sizes of receptive fields, like model areas MT and MST. Throughout the cascade, top-down influences disambiguate local information based on already stable interpretations on higher representation levels in the hierarchy. These mechanisms had to be implemented in the Full Neural Model inclduing form-motion interaction, adaptive spatial integration as well as detection of spatio-temporal occlusions, motion discontinuities, motion gradients and the ability to process and represent transparent motion. A fast but biologically plausible version, the Algorithmic Approach, had to be real time capable using massive parallel processing on Graphical Processing Unit (GPU).

The goal of the ***surprise module*** is the characterization of the saliency associated with some set of feature responses corresponding to a given pixel location within a feature map. This module had to satisfy the two main requirements:

- Flexibility, to be adapted to the needs of the Smart Eye input by taking into account the distinguishing features between saliency computation within the SEARISE project and prior efforts.
- Integration, to allow interaction with existing feature modules in the SEARISE architecture.
- Optimization, since this module will be used intensively at different levels of the hierarchy.

The goal of the ***fusion module*** was to provide principled means of combining saliency representations into a single unified representation which is crucial and necessary for the operation of the system, specifically, to tell the Smart Eyes where to look. This module had to provide to the system a high flexibility to choose between several ways to combine saliency maps.

The goal of the ***segmentation module*** was to provide a large set of edge-based models and region-based models for image segmentation. This module had to satisfy one main requirement: allow creating and combining a set of segmentation modules, each of them inducing a different term in the corresponding level-set evolution equation.

The objective of the ***form learning module*** was to provide means to further analyse salient regions based on the appearance of certain forms such as persons. The presence of persons serves as a reliable basis for further analysis of person behaviour.

The objective of ***motion pattern learning*** was the incorporation of *task-bias modulation* into the SEARISE system. By *task-bias modulation* we mean the ability of the system to adapt its responses *on-line* to the requirements of expert users. For example, the system should learn to track security-relevant events even when these events are no longer considered salient. Likewise, the system

should learn to ignore events which are salient, but do not usually constitute a security risk, e.g. waving flags in the stadium grandstand. The motivation for *on-line* learning is twofold: firstly, security-relevant events are rare. Thus, their characteristic motion patterns need to be learned from very few examples, or possibly even from single observations. While we did compile a training dataset for both long-range and short-range scenarios, the SEARISE system should be able to learn to recognize unforeseen security-relevant events when they happen. Secondly, legal obstacles prohibit the storage of security-relevant material beyond a short time interval. Thus, the its characteristic motion patterns need to be learned before the material is erased. Finally, motion pattern recognition had to be performed in real-time.

Hierarchical neural architecture

Design of the biologically motivated neural architecture to be run by Smart Eyes had to satisfy several requirements on its structure. These comprise but are not limited to the following:

-Hierarchical and modular design;

-Segregation between motion and form processing;

-Visual modules feeding on different input features along the hierarchy;

-Parallelization of independent processing modules where possible;

-Synchronization of critical processing modules supporting the instantaneous camera control.

- Implementation of Bayesian neural networks for form and motion recognition.

Software framework

The Smart Eyes hardware consists of a computation unit and the Smart Eyes head, which hosts the motor and lens control unit and the three cameras. The head is connected to the compute unit via Gigabit Ethernet.

The framework structure and algorithms have been implemented in C++ and optimized for real time performance. The parameterization of modules and connections as well as the processing schedule is dynamically controlled via python scripts.

To guarantee real time performance, three different parallelization techniques have been adapted in the smart-eyes framework. Firstly, low-level computations have been outsourced to the dedicated graphics hardware, which provides a highly specialized and concurrent architecture for image processing. Secondly, several computationally expensive modules have been paralleled internally using the OpenMP extensions for C++. Lastly, the modules have been clustered into groups which run on separated processing cores, using the MPI standard.

Field tests

The objectives of Smart Eyes in the *Arena scenario* were to detect, fixate with appropriate zoom and to track security relevant events. To define what is security relevant in the Arena, security personnel was interviewed and their observation behaviour was recorded via eye-tracking. The information on the nature of dangerous behaviours was incorporated to bias Smart Eyes attention towards these events. One special feature in the arena was the fact that waving flags were a common destructor for the saliency detection since they pose complex motion, yet are not security relevant. Dedicated flag motion pattern recognition module had to be incorporated into the neural architecture to identify and eliminate spots with waving flags from the focus of attention of Smart Eyes.

Technical objectives of the field test in the Arena scenario were to reach real time performance and precise fixation, zooming and tracking of a salient region.

In essence, the objectives for the *short range scenario* on the railway platform are the same as for the long range scenario. A few technical differences have arisen from the perspective and reduced distance. The latter made feasible the automatic detection of individuals and categorization of their

behaviour in the global view. However, the task of fixation became more difficult due to depth variations within the scene now comparable to the distance to Smart Eyes.

Ground truth data, human psychophysics and validation methods

A benchmark for behavioural pattern recognition should be comprised of a collection of videos pertinent to the application domain as defined for the SEARISE system. For the *long-range Arena scenario*, we created a benchmark dataset, the ***Tübingen hooligan simulator***, by staging relevant events. We chose this approach to circumvent legal obstacles and to ensure balanced statistics of security-relevant events. We compiled this dataset in collaboration with officers of the Düsseldorf police, who also participated in psychophysical experiments aimed at elucidating the relative contributions of expert knowledge and saliency to the search strategies employed by experts and naïve observers.

For the *short-range city train station scenario*, we compiled a benchmark dataset in collaboration with security experts from the Rheinbahn, who are in charge of surveillance at the Düsseldorf subway stations. We interviewed the Rheinbahn experts to determine which events are relevant, and how relevant and frequent these events are. Subsequently, we staged security-relevant events in the subway station at the Arena, where the second SEARISE prototype is installed. The events were recorded by the SEARISE camera.

1.3 Main S&T results/foregrounds

1.3.1 Visual system design and camera control

Starting from defining the requirements of the visual system on the basis of the target application domain, and the technical specifications of the system's components, we have designed and assembled the Smart-Eyes prototype, a trinocular robotic head with 4 degrees of freedom (common tilt, independent pan and neck movement). The prototype has been developed from a previous EU project (Eurohead), and it has been adapted to meet the specific requirements of SEARISE project. The head is equipped with a fixed wide angle camera, mounted stationary with respect to the head base, and other two moving cameras assembled to produce an active stereo head. The active cameras are equipped with motorized optical lenses. Each optical lens has three degrees of freedom which control iris, focus and zoom.

All the cameras are connected to the processing unit through Gbit Ethernet, the trinocular head has a PC interface, through a PC104, with Gbit Ethernet and both the controller cards for motors and optics are compatible with the iCub robot platform (www.robotcub.org).

On the basis of functional testing on the first prototype, to improve the stability of the stereo cameras during active movements the design of the supports has been modified by including C-like supports on which to fix the motorized optics. The final structure (see Box 1) includes also a mechanical tuning system (fine tuning 3 DOFs) that allows a better calibration. In addition, the pan rotation of the neck has been enabled. Hence, altogether the final system has ten degrees of freedom: four for the camera movements and six for the optical lenses.

Miniaturized optical encoders have been mounted inside the motorized lenses to allow a closed-loop control of the zoom.

To protect the SEARISE head from dust and bad weather, a protective dome has been designed, which includes two heater/conditioner to prevent condensation. The dome is made of aluminium (backside) and transparent plexiglass.

Videos acquired by the two active cameras are processed by a software module developed to compute dense bi-dimensional disparity and optic flow maps. The cortical architecture is based on a

distributed population coding, whose units are energy neurons [Ohzawa et al. (1990)], [Adelson Bergen (1985)]. The distributed approaches for the computation of the dense feature maps are embedded in a common algorithmic structure: a filtering stage, with separable Gabor convolutions, a decoding stage and a coarse-to-fine refinement. We extract disparity and optic flow features from a sequence of stereo image pairs, using a distributed bio-inspired architecture that resorts to a population of tuned cells [Chessa et al (2009)b], [Gibaldi et al. (2009)]. Using this approach we obtain reliable estimates, with accuracies comparable to the state-of-the-art algorithms. The performances of the developed algorithms have been compared with the results from the literature, by using standard benchmarking sequences [Baker et al. (2007)] (see Fig. 1 and Fig.2)

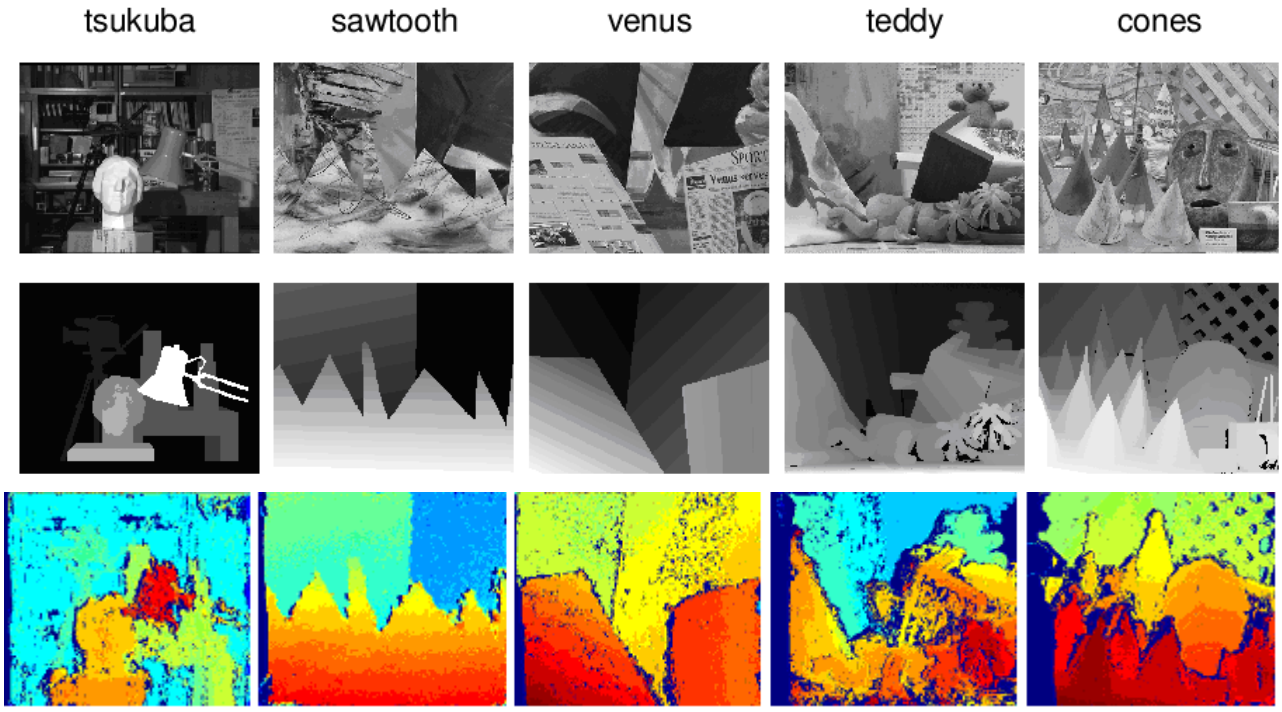


Figure 1. Disparity estimation obtained by the proposed distributed architecture for the benchmarking sequences from [Baker et al. (2007)].

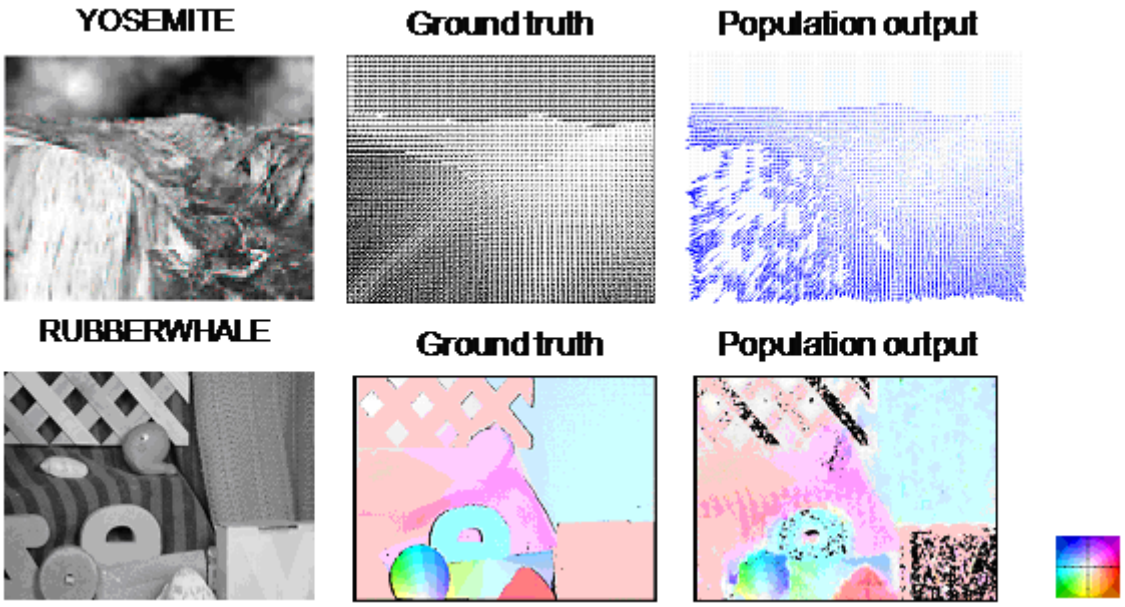


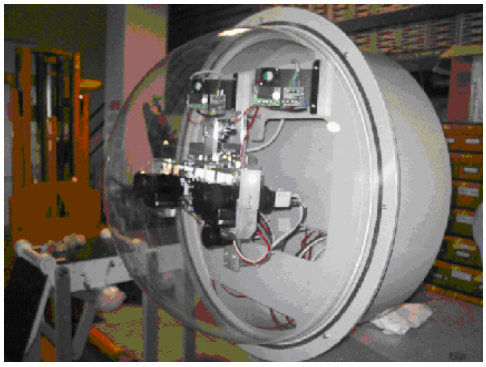
Figure 2. Optic flow estimation obtained by the proposed distributed architecture for the benchmarking sequences from [Baker et al. (2007)].

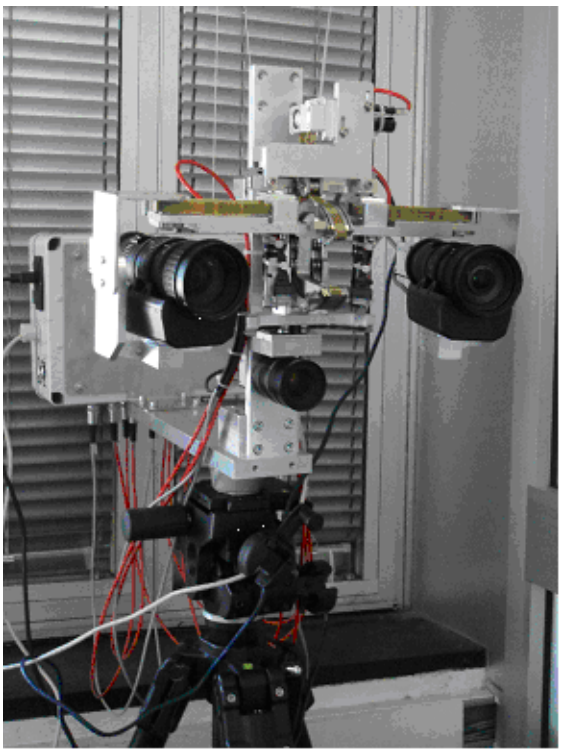
SEARISE HEAD **"Smart Eyes"**

General features

Trinocular vision system with fixed camera + pan-tilt and vergence stereovision positioning device

Height:	40 cm
Width:	43 cm
Depth:	29 cm
Weight:	6.5 Kg + about 2Kg electronic box
Baseline:	30 cm





	Pan eyes	**Pan neck**	**Tilt neck**
Limits:	± 30º (Software limit)	± 10º(Software limit)	± 60º(Software limit)
Acceleration:	5100º/sec²	1600º/sec²	2100º/sec²
Max. speed :	330º/sec	73º/sec	73º/sec
Resolution:	0.03º	0.007º	0.007º
Optical encoder:	512 pulses/revolution	512 pulses/revolution	512 pulses/revolution
Motor voltage:	12 V	12 V	12 V
Gear ratio:	1:80	1:80	1:80
Motor torque:	0.59 Nm	0.59 Nm	0.59 Nm
Backlash:	< 1 arc min	< 1 arc min	< 1 arc min

Cameras and optics

Mono:		
	Resolution	1624 x 1236 pixels
	Sensor area	7.1 x 5.4 mm
	Pixel size	4.4 x 4.4 um
	Focal length	4.8 mm (short range), FOV 73º
	Focal length	12.5 mm (long range), FOV 31º
Stereo:		
	Resolution	1392 x 1040 pixels
	Sensor area	6.4 x 4.8 mm
	Pixel size	4.65 x 4.65 um
	Focal length	7.3 - 117 mm, FOV 47º - 3º

Box 1: Technical description of the final SmartEyes system.

Since in SEARISE we process high-resolution stereo images, a fast implementation of the cortical architecture is necessary. To this end, we have implemented a GPU-based distributed algorithm for the computation of 2D disparity, using the Nvidia CUDA Library. This has reduced computation time with no loss of accuracy. This module is now part of the high performance image processing library.

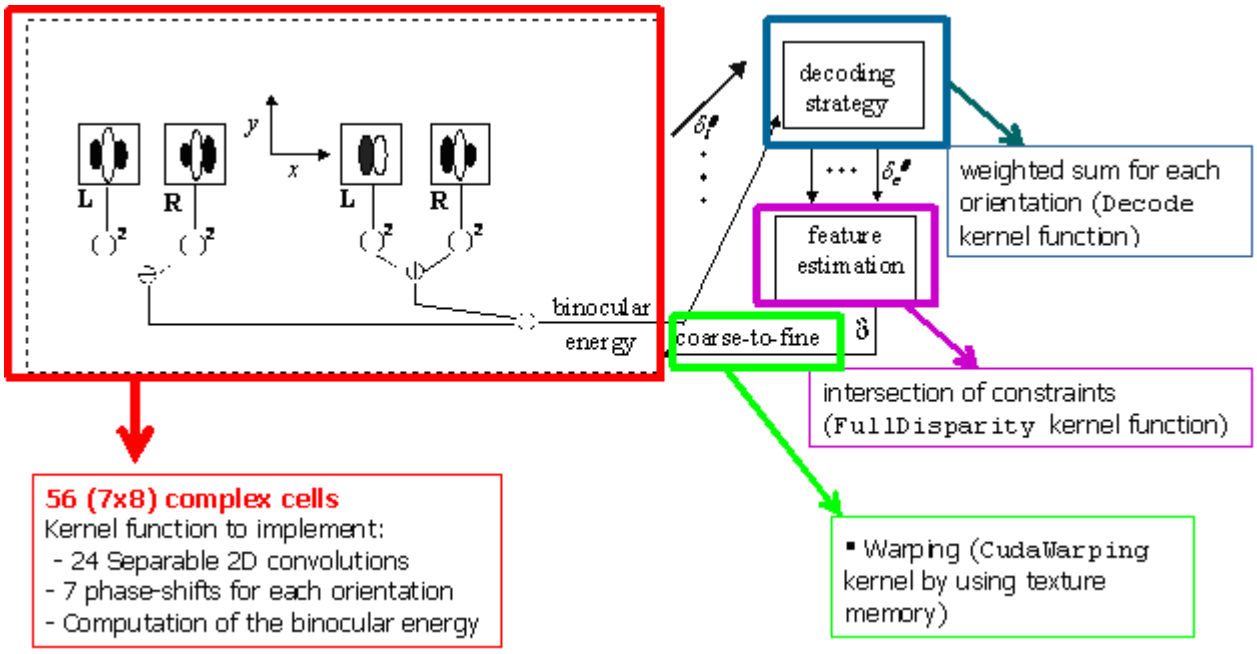


Figure 3. The overall architecture of the GPGPU Disparity Module

Another objective was the development of the modules for the control of the active camera movements. We have focused on the development and testing of a vision-based, biologically plausible, control strategy that fits the Hering's law (e.g., see [Mallot (2000)]), by studying the cooperation of vergence and version movements, as it happens in primate visual systems. Accordingly, the proposed approach [Samarawickrama and Sabatini (2007)] accounts the version and vergence as two independent movements and considers them in parallel and proved to be effective for the control of the robotic system (see Fig. 4).

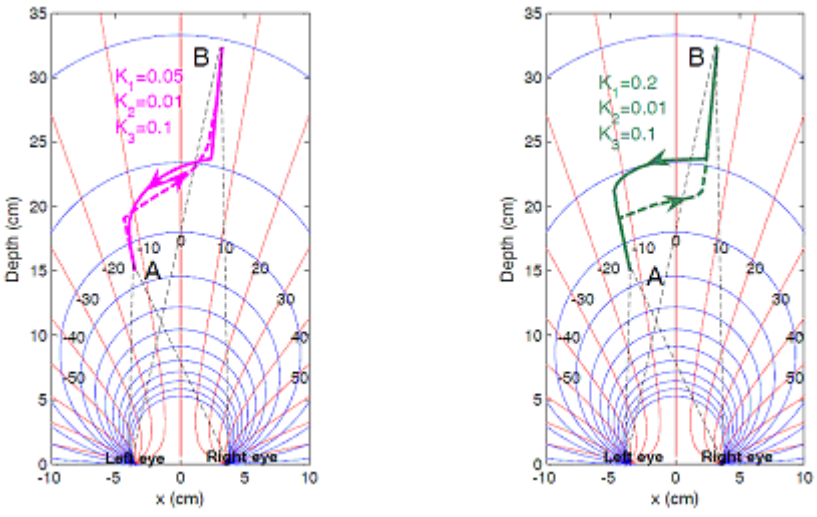


Figure 4. Different trajectories of the fixation point obtained with iCub anthropomorphic head in lab condition. The two cameras change the fixation point from point "A" to point "B" and vice versa.

Additionally, we have described the geometry of the trinocular head in order to accurately move the active cameras with respect to the image coordinate system of the wide angle camera. To properly move the active cameras involves the estimation of the scene depth, hence we have developed an iterative angular-position control based on 2D disparity maps to fixate in depth (see Fig. 5).

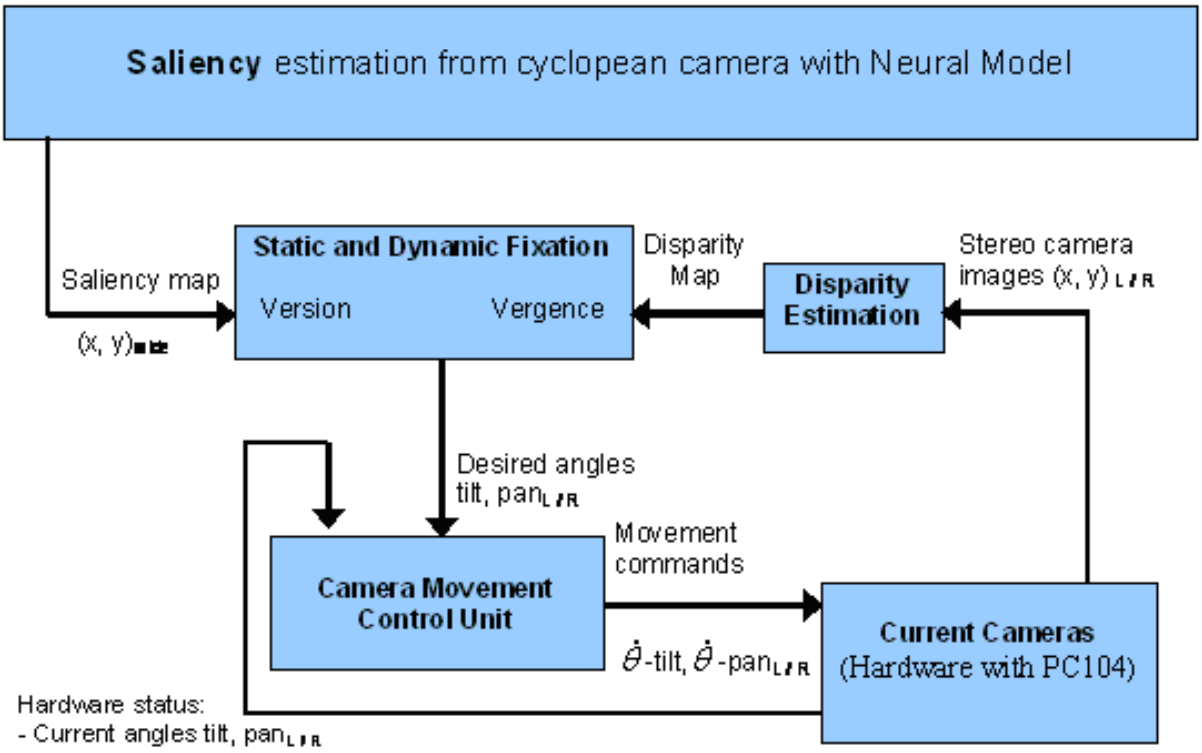


Figure 5. Angular-position control scheme of static and dynamic fixation of the trinocular system.

In the trinocular head remained the problem of the displacement of the image centre in the active cameras when the images were zoomed in or zoomed out. To solve it we have implemented a calibration procedure to determine the optical centre in each active camera. Accordingly, the active images have been centred in their respective optical centres. A performance example of the final system is shown in Fig. 6.

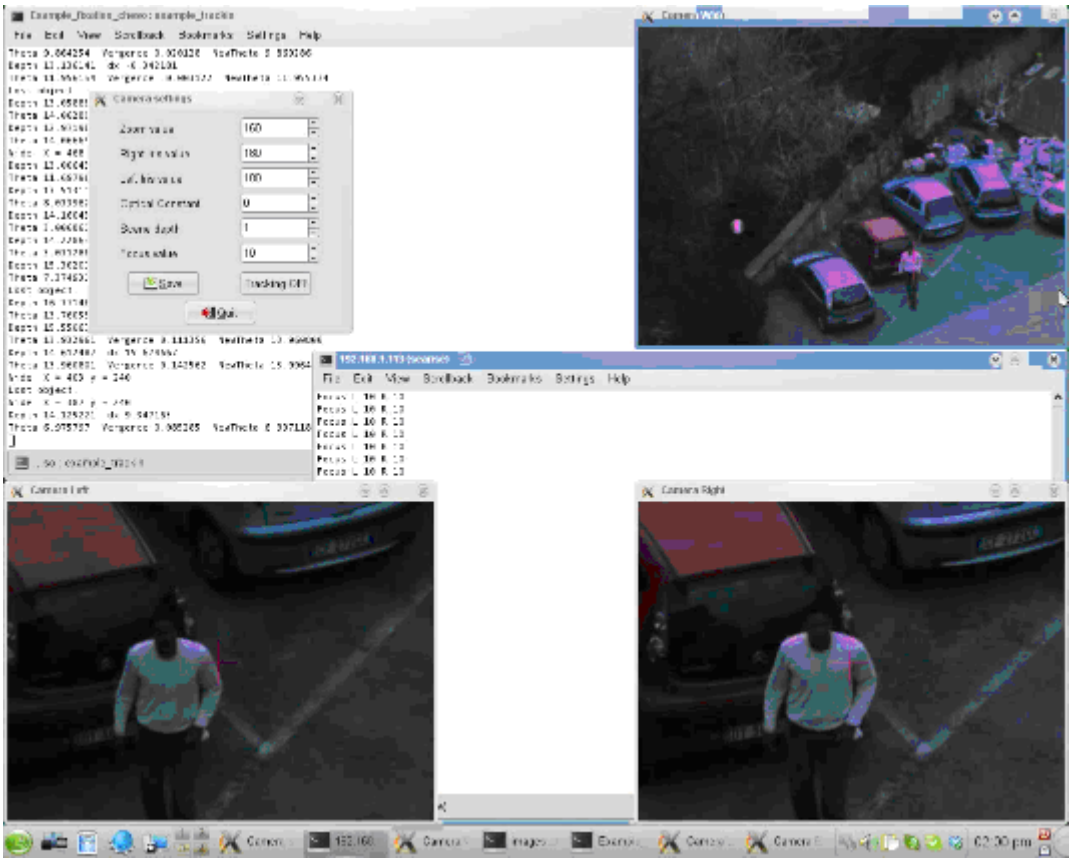


Figure 6. Frame of a sequence recorded by using the trinocular head. This example shows the reference image (Camera Wide) and the active images (Camera Left and Camera Right) at 4x magnification factor. The crosshair is always pointing the same target regardless of zoom variability.

1.3.2 Motion-driven attention and grouping

UULM has developed a **software library** for the extraction and integration of form and motion features. We took inspiration in the known segregation of the cortical visual processing streams into a ventral and dorsal pathway and focused on the processing of shape and motion information, respectively. Mechanisms designated to these pathways follow a core architecture of basic processing stages, that resemble key principles of neural information processing: Linear and non-linear filtering, feedback re-entry (plus nonlinear enhancement) and activity normalisation via shunting inhibition. This network architecture extends previous proposals which have considered feedforward filtering with a non-linear competition to account for non-linear effects in contrast and motion responses of V1 cells [Heeger, Qian et al]. The software library makes extensive use of the parallel processing power if Graphical Processing Units (GPU) where possible.

The **Full Neural Model** implements this architecture and is dedicated to the simulation of neural dynamics in motion and form processing. This model has been developed and further extended with partners INRIA and FhG. This model consists of two stages, namely motion detection and motion integration with dedicated model areas V1 and MT. Both model areas use generic mechanisms of filtering and activity normalization and communicate by recurrent connections. In the Full Neural Model, motion hypotheses are densly represented using cell activities for possible velocities and directions at every image location. Initial motion detection is computed using model area V1 using elaborated Reichardt detectors [Bayerl & Neumann, 2004]. It enhances unambiguous information while keeping ambiguous information. Model area MT integrates motion information of larger spatial regions. Updated motion estimation after integration is fed back to model area V1. The

feedback mechanism is purely modulatory and enhances existing motion hypotheses while it cannot generate new motion activities at places without driving input. With this mechanism, the model is able to disambiguate locally ambigious information.

In parallel, a second model of motion processing was developed. This implementation will be further referred to **Algorithmic Approach**. It offered a promising perspective of efficient processing time and less memory usage by comprising an efficient implementation of the neural architecture in terms of a sparsified representation of velocities [BAYERL & NEUMANN, 2007]. That way it transfers the underlying principles of the Full Neural Model into an algorithm that has a proven complexity of O(N log(N)) with N the number pixel in the processed image. In model area V1 here, inital motion detection has been replaced with a fast mechanism of initial estimation, namely the Census transform. Besides its computational efficienty, another major benefit of the Algorithmic Approach this detection is its ability to detect and represent unlimited motion speed.

Third, a neural model of **local contrast detection and grouping** was developed. This model is able to detect basic form features like discontinuities in the luminance distribution [Hansen & Neumann, 2008]. Here, model area V1 here uses Gabor wavelet responses, whereas model area V2 incorporates long-range grouping of like-oriented contrast responses in order to enhance smooth contours. The grouping stage employs bipole filters which are formed by the combination of two elongated Gaussian sub-fields which are aligned along an oriented axis. The model incorporates recurrent feedback interaction between model areas V1 and V2, using center-surround competition. This suppresses distracting activities in the location-orientation feature space that do not correspond to larger configurations, like noise and outliers.

Quality and speed of the **Algorithmic Approach** has been improved by integrating various extensions of the Full Neural Model and by migrating more parts of the estimation process to the GPU framework. Extensions made to the Algorithmic Version were estimation of motion occlusions and motion gradients [Beck et al, 2008]. Regions of occlusion and disocclusion provide clues about the ordinal structure of the scene and the arrangement of objects and indicate regions of unreliable motion estimates due to a lack of reference points. We compensated this problem by integrating multiple times scale in the estimation. Motion gradients make it possible to detect object boundaries and to distinguish between local acceleration and deceleration as well as direction changes in object motion. Using multiple time scales, correspondences are not only calculated between subsequent frames, but also with frames t+2 and t+3, which allows estimation of subpixel movements. Components ported to GPU processing were initial motion detection with the Census transformation, non-linear transformations of hypotheses, normalisation and integration. This made use of the GPU library developed in the first year. In addition, modulatory input from the estimation of motion energy helps to identify small and slow moving objects that otherwise would be masked by noise.

The **Full Neural Model** has been extended with detailed mechanisms of form-motion interaction in order to disambiguate local motion estimates. This also helps to stabilize motion integration in cases of mutual occlusion and in cases of multiple motions of opaque surfaces [Beck & Neumann, 2010]. In addition, information derived from the static form pathway can be utilized to enhance the motion integration process, particularly at boundaries, and to improve the quality of the computed optical flow. Several variants of form-motion interaction have been investigated, namely the elaborated oriented-diffusion model developed by partner INRIA [Tlapale et al., 2008] and two simpler variants of form modulation derived from local image contrast. All three approaches use form information to control the integration of motion information from different objects. In addition, it sharpens the optic flow by preventing the model to integrate or propagate flow into empty or untextured regions.

The necessity of representing and processing **motion transparency** is two-fold regarding the SEARISE scenario: Mutual occlusion of objects and people in crowded scenes generate multimodal velocity representations in a small neighbourhood which need to be segregated. Also in crowded scenes collectively moving groups of people generate multiple bands of contiguously moving patterns of possibly different widths and directions. The same core model architecture has thus been extended in order to robustly process inputs of **transparent motion**. Examples relevant for the project scenario are crowded scenes in which multiple motions of relatively small objects result in flow patterns of different motion directions. We argue that the ability to robustly handle transparent motion is essential for the development of general purpose vision systems that operate in real-world scenarios where mutual occlusions and semi-transparent configurations occur regularly.

Both estimation modules were extensively tested in the second year as well. In particular, we have evaluated its advantages with respect to the constraints given by the applications within the project.

Also in the second year, partners UUlm and DIBE-Unige specified a **camera tracking framework** that transforms positions in the camera image to camera movement commands. The limitations of the camera head include the ability to only fixate one salient object at a time. This indicated the need of some levels of abstraction when camera hardware and the software framework are consolidated. Based on this specification UUlm developed a tracking module that allows following objects in the visual field of the wide angle camera. The module assumes the camera to be fixed and that objects are moving over a static background. These assumptions hold for the long-range scenario as well as for the short-range scenario. The module separates the foreground from the background with a probabilistic background model. Flow information is additionally used to segment objects in the foreground. This is especially useful in cases where different objects move close to each other with different directions. Flow information is also applied to predict future positions of objects in the scene.

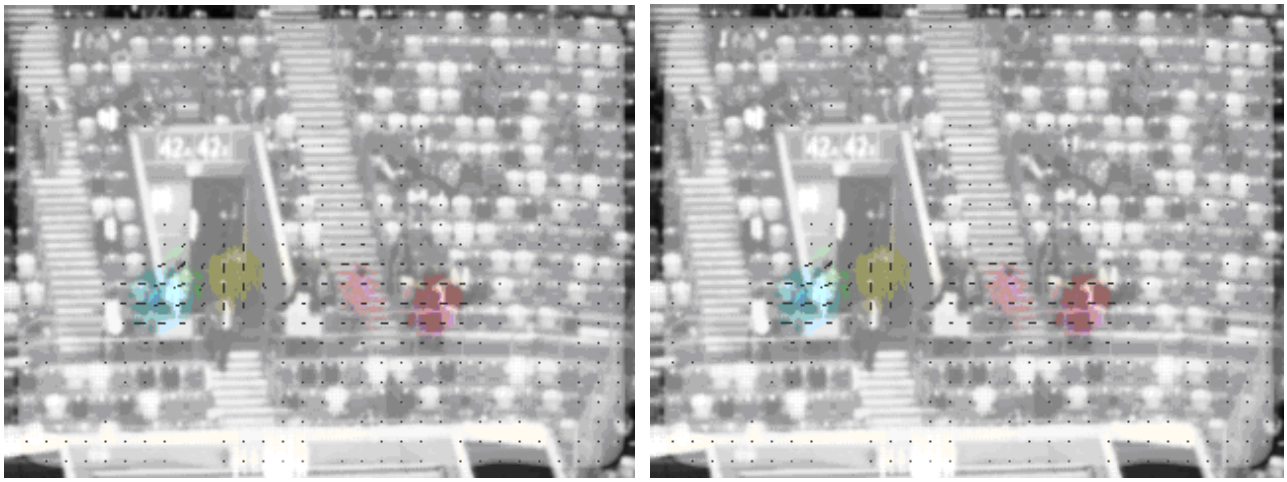


Figure 7. Results from motion model: The optical flow, computed by the Full Neural Model (left side) has higher accuracy for sub pixel movements and produces smoother full field optical flow. For reasonable large displacements on the other hand, it needs a lot of memory and computation time. The output produced by the Algorithmic Approach (right side) can easily handle larger displacements due to its initial

In agreement with reviewers' comments we elaborated the benchmark for the evaluation of biologically motivated models of motion estimation. In contrast to approaches in the field of computer vision, provided stimuli, ground truth and evaluation methodology are tailored to probe biological models of motion estimation, where ideal temporal and spatial behaviour is not adequately defined by classic benchmarks. This was done in collaboration with partner INRIA [Tlapale et al., 2010].

Motion discontinuities that were used to segregate regions of different motion features (direction and speed) are classically estimated using center-surround mechanisms. A full representation of the necessary neural populations for detecting contrasts leads to high computational efforts. We developed a computationally efficient and still biologically plausible mechanism to generate activity for estimation of motion discontinuities. This incorporates an efficient represention of activities, in analogy to the Algorithmic Approach of the motion estimation module. We have also extended the estimation of spatio-temporal **occlusions and disocclusions** by introducing direction-selective occlusions, which allow a more precise diffentiation between various flow configurations.

The **Full Neural Model** was extended with a mechanism of adapting the shape of local MT integration fields with local information about structure of optic flow. This approach improves quality and convergence of estimated motion hypotheses [Ringbauer et al., 2011].

The model for **transparent motion** processing has further been improved. The model developed by [Raudies & Neumann 2010] is able to segregate motion patterns which contain multiple local motion directions and is in agreement with various experimental findings. The model was also applied to real world data and the results were promising. Further investigations are necessary to justify their significance.

In case of the long range scenario, movement usually occurs in one single and slightly slanted depth plane. The residual movement in depth is practically not detectable with the given baseline and resolution of the binocular camera system. This fact changes significantly in the short-range scenario. Here, the camera angle on the scene produces image regions that highly differ in depth, which renders depth cues an important component of scene understanding, classification and segmentation. We have investigated several ways how **disparity and motion** processing can benefit from each other and how the output of those components can contribute to scene segmentation. On the one hand, the close relationship of mechanisms for disparity and motion estimation suggests an interaction between involved cortical areas. In collaboration with partner DIBE-Unige we extended a biologically motivated model of disparity estimation with UUlm methods for discarding potentially mismatching disparity values estimated at regions at motion discontinuities. Furthermore, we proposed possible interactions of disparity and motion estimation for scene segmentation. On the other hand, whereas binocular disparity is an important and obvious cue for a scene's depth, perceptions of relative depth are also experienced from monocular cues. Here, motion information alone is sufficient for this perception. We proposed a model for inferring **ordinal depth** from monocular motion cues using various components of motion-related estimates. This approach uses occlusions and motion energy for (rapid) scene segmentation [Tschechne & Neumann, 2001a|b].

We proposed an extension to the **camera tracking** component to increase robustness towards noise and interruptions in measurements. An intermediate probabilistic filter stage (base on [vo2006]) removes noisy measurements before they are further processed. This also deals with sudden object appearances, splits and merges.

1.3.3 Visual learning of shapes and motion patterns

Form learning and recognition was developed within SEARISE in the form of a fusion of the hierarchical neural architecture concept, which is one of the core concepts of the project, and the well established mathematical tool-set of Bayesian statistics. Based on the established neurophysiological evidence, the learning architecture comprises a hierarchy of converging receptive fields as illustrated in Fig. 8 and it was implemented using Bayesian techniques.

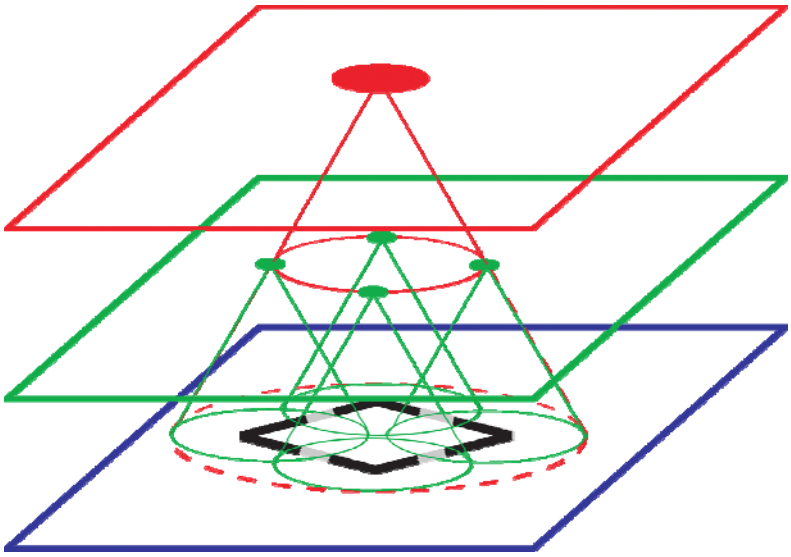


Figure 8. The hierarchy of converging receptive fields which serves as the core concept for form and motion recognition.

The resulting form learning module has been optimized by tuning the learning schedule, taking care of training set bias, and taking care that features extracted on lower layers are effective for the classification task by introducing classifiers on each layer during learning. Furthermore as features we decided to use the established RHOG features instead of the neural model **feature,** as the former seem to preserve more valuable information and have been extensively tuned for the task. The results after these optimizations can be seen in Fig. 9.

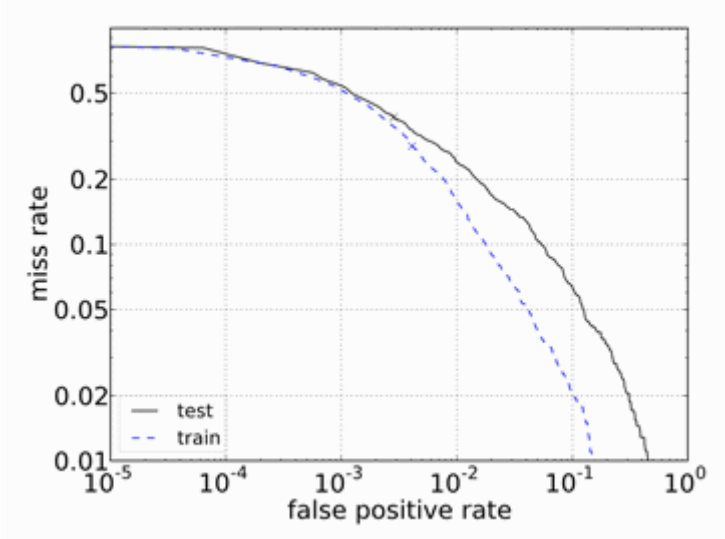


Figure 9. ROC characteristics of the shape recognition module running on the INRIA LEAR dataset.

To be able to use the module as a person detector in live video streams with unknown object size and position the tolerance of the detection performance to shifts and size changes was evaluated on shifted and scaled versions of the INRIA-LEAR person dataset. The results are summarized in Fig. 10. In conclusion the final detector scans the scale-space on a spatial grid of 12 pixels and a series of scales in which successive scales are 20% apart.

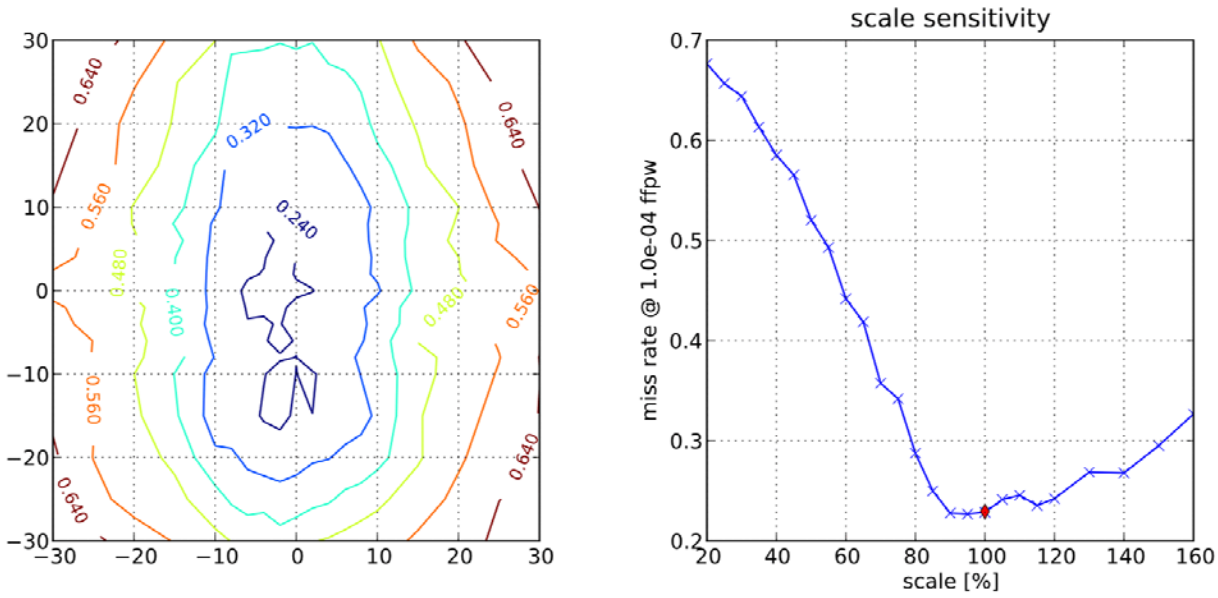


Figure 10. Results of the tolerance tests of the pedestrian recognition module, showing the miss rate as a function of object misplacement (left) and scale change (right).

To assure performance in real time the module was heavily optimized, utilizing the vector processing (SIMD) units available in modern processors, medium- and coarse grained parallelization within and across scales utilizing OPENMP and futures. The module was also ported to CUDA to be able to process some of the scales on the GPU, depending on the overall system load.

The form learning module was also applied to a variety of other tasks, including some industrial tasks, with a few further optimizations such as discriminative fine-tuning using error back-propagation and gradient descent, shift invariant learning, and automatic feature selection. These developments were published to the scientific community in [Oberhoff 2011].

Because it became clear that the chosen approach in principle deals poorly with occlusions we also pursued the integration of segmentation directly into the recognition process: The redundancy introduced by overlapping neighboring receptive fields is lifted by introducing selector variables that assign each input to just one receptive field. It has been found that the resulting ambiguity can be resolved by introducing constraints on the overall number of utilized receptive fields. This direction has shown some promising initial results (Fig. 11) and lead to a publication to the scientific community in [Oberhoff et all. 2011]. This feature has not yet been integrated into the Smart-Eyes system as the increased complexity of the model makes advanced inference algorithms such as Markov-chain Monte-Carlo samplers necessary, which are not real-time capable. But we believe that it is a valuable contribution and basis for future improvements of similar recognition algorithms.

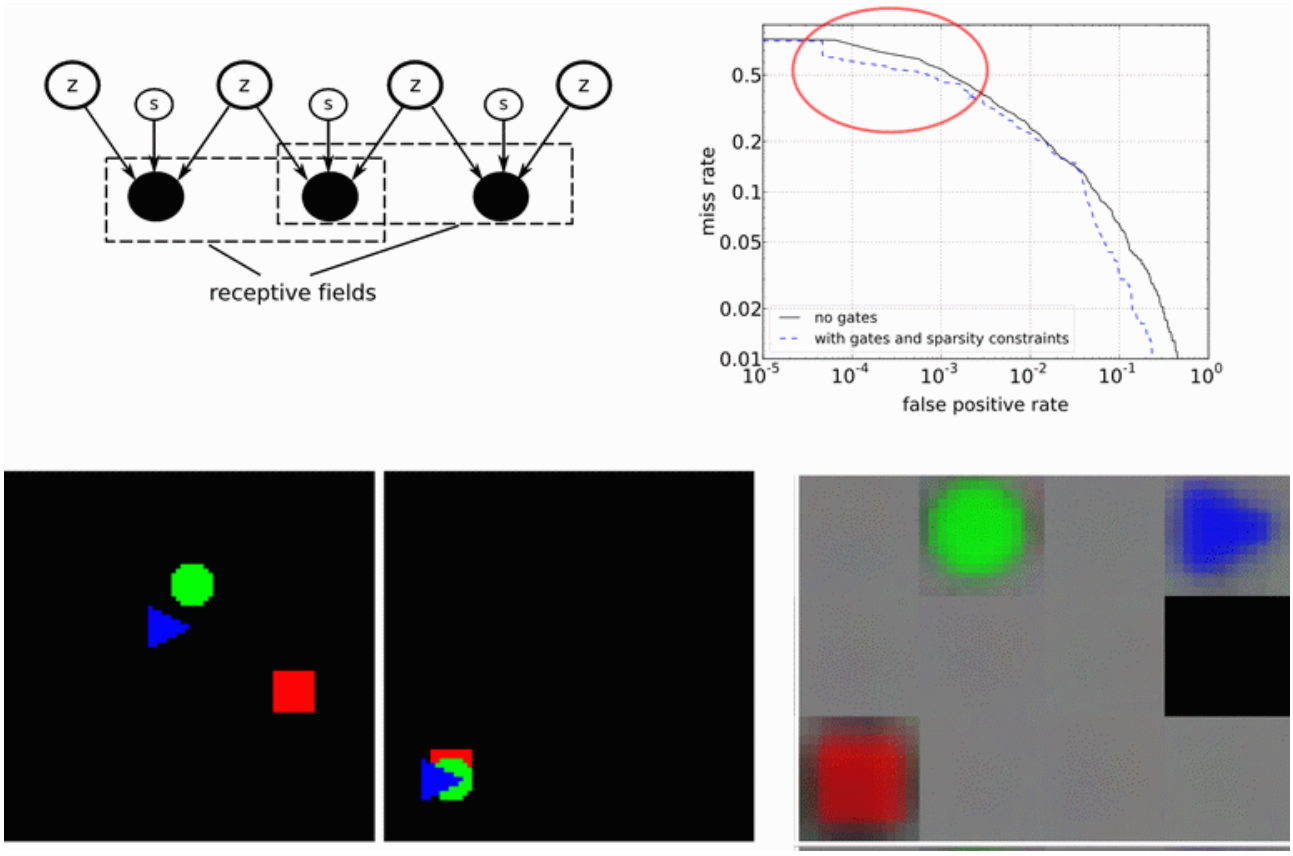


Figure 11. Enhanced shape recognition layer with incorporated segmentation. Upper left shows the modified graphical model in which additional variables are introduced assigning each input pixel to one of several possible receptive fields. Upper right shows the impact on pedestrian recognition on the INRIA-LEAR dataset. Bottom left shows a toy dataset in which geometric shapes move freely including occlusion. The lower right shows that the model succeeds in extracting the underlying shapes, invariant to the movement and occlusion.

The objective of **motion pattern learning** was the incorporation of *task-bias modulation* into the SEARISE system, i.e. the possibility to include task-specific expert knowledge into the visual processing hierarchy which goes beyond low-level saliency. We tested a range of machine learning approaches to this end.

During year 1 we experimented with motion pattern classification via *support vector machines* [Schökopf et al. 1999] and largely hand-crafted optic flow features. We did these experiments on a hand-annotated testbed video acquired during a soccer game in the Arena. We demonstrated that the separation of classes like waving flags, jumping crowds, or people walking along the aisles of the grandstand were possible by looking at optic flow alone.

Moreover, we tested two different unsupervised learning algorithms for the learning of features by Independent Component Analysis (ICA). ICA a common approach for feature learning in computational vision and has been applied, for example, to learning of low-level visual filters (e.g. [Bell & Sejnowski, 1996]) or for the modelling of pictures of faces e.g. [Bartlett et al., 2002]. ICA has been applied successfully to motion patterns, e.g. in the context of robot navigation (e.g. [Ohnishi & Imiya, 2008]). This motivated the application of this technique for the analysis the available motion patterns.

The first applied technique was a standard ICA algorithm (JADE) [Cardoso & Souloumiac, 1993] using an implementation for complex valued data. To apply this algorithm the unit vectors derived from direction detectors were converted into complex numbers forming a complex data matrix. The standard ICA approach assumes that all training patterns are spatially aligned and thus cannot

appropriately model translated patterns. To deal more efficiently with such spatial variations in the analyzed patterns we tested a second new ICA algorithm which is able to recover independent basis patterns even in the case where they occur with spatial displacements in the individual training examples [Omlor & Giese, 2007, 2011]. The algorithm was able to model the optic flow patterns induced by critical events, such as wavig of a flag, accurately by very few extracted components. Also we tested the resulting features using them as input in different classifiers However, due to the optimization involved in demixing each frame, real-time performance could not be achieved. Thus, the ICA approach was given up.

During year 2 we began to test generative probabilistic models on motion pattern classification and novelty detection. We developed a *sequence classifier*, which learns sequences of optic flow features and uses the likelihood of these sequences for label prediction. Label predictions compete with each other via a softmax function, i.e. lateral activity normalization illustrated in Fig. 12.

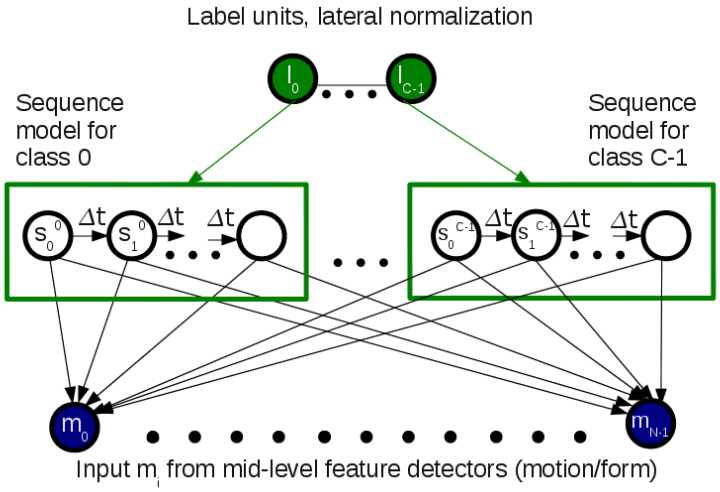


Figure 12. The sequence classifier model. Each class is represented by a sequence of optic flow feature combinations. Blue circles: optic flow feature activations, e.g. the output of the MT layer of the optic flow module from partner **UUlm**. Green boxes: sequence model for each class, *open circles*: feature combination activation for a given time step in each class sequence, *green circles*: label units with lateral normalization. Arrows represent directed generative connections, lines represent undirected connections. Δt indicates a generative connection which operates forward in time.

Concurrently, our efforts for creating a benchmark dataset (see below under "Human psychophysics and benchmark creation" in section 1.3.6) with expert feedback highlighted the importance of an additional requirement for the motion pattern recognition component of SEARISE: learning would have to be *on-line*. This requirement is due to legal obstacles prohibiting the storage of security-relevant material, which would be necessary for batch (off-line) learning. Furthermore, security-relevant events are extremely rare, i.e. we could not expect to have a complete collection of labelled events before deployment. Therefore, new event classes would have to be added while the system was operating. This is easily accomplished in a generative architecture like the one shown above (just add a new label unit and sequence), but much harder in a system using discriminative learning, such as a support vector machine. We implemented *on-line* learning initially via a moment-matching technique [Bishop, 2007] (see deliverable report D4.2. for details) and tested the sequence classifier with both algorithmic and neural optic flow features provided by partner UUlm, finding that they yielded comparable results on the Weizmann action database [Blank et al., 2005].

An important task for the SEARISE system in the context of event classification and detection is the determination of the beginning and the end of a given event, i.e. the temporal segmentation of the event sequence. We experimented with Bayesian binning (BB) to this end. BB's usefulness for this type of task was established previously in the domain of neural modelling [Endres et al. 2008]. A

detailed account of our experiments with action segmentation can be found in [Endres & Giese 2009, Endres et al. 2010A, Endres et al. 2011], exemplary results are shown in Fig. 13.

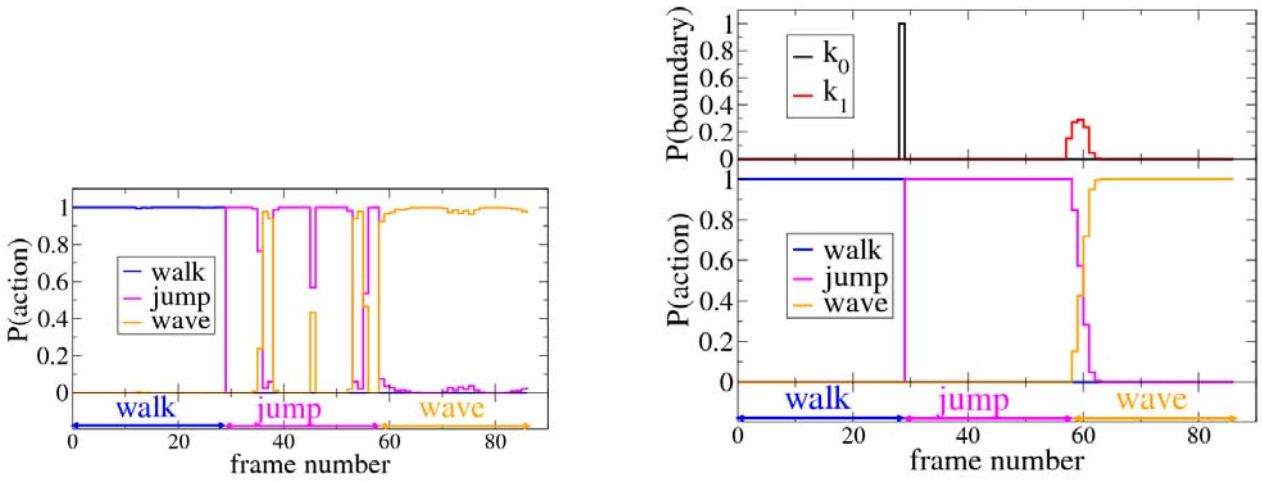


Figure13. Event/action segmentation results on video sequences from [CITE WEIZMANN]. *Bottom, left and right*: the sequence to be segmented contained 3 actions: walk, jump, wave. In the sequence, each action was presented for 28 frames before switching to the next action. *Middle, left*: frame-based event classification results. 'Jump' and 'wave' are hard to discriminate. We computed the expected action identity (e.g. walk,jump or wave) at every point in the sequence. *Middle, right*: event classification with Bayesian binning. The correct event is identified with near certainty. *Top right*: marginal posterior distribution of the event boundaries k_0 and k_1. Inferred boundary posteriors are concentrated at the correct points in time.

During year 3 we decided to switch the learning from moment-matching to *variational Bayesian expectation maximization (VBEM)* [Bishop, 2007], mainly because VBEM casts learning into a convex optimization problem of a lower bound on the marginal probability of the data, and therefore guarantees the convergence of the resulting algorithm to a (local) optimum. Also, we found that using second-order features (here: Gaussian observation models with full covariance matrices) provides better classification performance than first order features (e.g. Gaussians with diagonal covariance matrices).

Open source software: we implemented this version of the sequence classifier as framework module building on the **varmod** library, which is part of the released SEARISE software. This library can not only be used to assemble the sequence classifier, but it contains a collection of C++ templates which allow for the construction of a large variety of singly-connected hierarchical Bayesian models. Furthermore, inference and learning can be efficiently parallelized using the openMP extension to C++, permitting inference in *real-time.*

Testing: we tested the framework module on two scenarios: the benchmark videos from the *Tübingen Hooligan Simulator* [Endres et al. 2010b] and flag detection in the Arena. The latter scenario is important, because the saliency tracker tends to find and get stuck at waving flags in the Arena grandstand. Thus, feedback from the sequence classifier can be used to modulate attention away from these flags. Using the sequence classifier module both for flag detection and hooligan detection is in line with one of the major goals of the SEARISE project, namely to create a system architecture which can adapt to a variety of different tasks via learning, rather than hand-crafting a solution tailored to every sub-problem.

To train the classifier for the flag detection task, we used ≈12000 frames from recordings during soccer matches. These recordings were made with the SEARISE system. The frames were filtered by the attention module, yielding (at most) one salient region per frame. A region was labeled either 'flag', if it contained a flag, or 'relevant' if it did not contain a flag. The regions and labels were

subsequently used as training data for the classifier. For validation, we used ≈3500 frames labeled in the same fashion. Detection results are shown in Table 1.

Saliency only			**Saliency and flag classifier**		
	predicted			**predicted**	
actual	flag	relevant	**actual**	flag	relevant
flag	0.0	0.86	flag	0.83	0.028
relevant	0.0	0.14	relevant	0.017	0.13

Table 1. Performance of flag detection in the Arena scenario.

Clearly, using the flag-trained sequence classifier improves correct detections of relevant salient regions significantly. Moreover, this improvement can be achieved in real-time: we measured the stand-alone framerate of the sequence classifier at 23 fps on 4 cores for this task.

On the *Tübingen Hooligan Simulator* videos, we evaluated the scaling of speed and classification performance with the number of features in the model (Fig. 14).

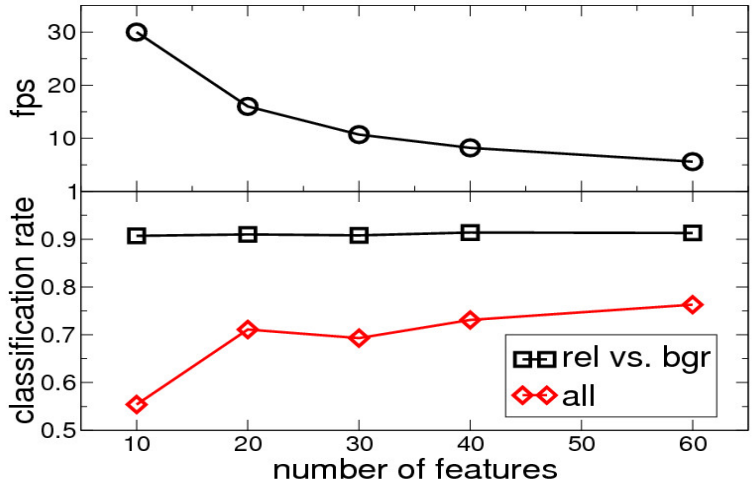


Figure 14. Performance of the sequence classifier framework module as a function of the number of shared features per sequence. *Top*: stand-alone frames per second (i.e. computational speed of this module only) on 4 cores. Frames per second are roughly inversely proportional to the number of features, indicating good parallel efficiency. Real-time performance is achievable with up to 40 features (8 frames per second). *Bottom:* red diamonds: classification rate of all classes in the hooligan simulator; black squares: detection rate of security-relevant behaviours (rel) vs. normal background behaviours (bgr). . Rates are conditional on the saliency module finding the location of the relevant event, if one was present. While the detection rate is relatively constant across different numbers of features, the classification rate increases almost monotonically.

1.3.4 Attention processes guided by saliency and segmentation

Original models of saliency-based visual attention were adapted from [Koch and Ullman 1985]. Early visual features such as color, intensity or orientation are computed, in a massively parallel manner, in a set of pre-attentive feature maps based on retinal input. Activity from all feature maps is combined at each location, giving rise to activity in a topographic saliency map. The winner-take-all (WTA) network detects the most salient location and directs attention towards it, such that only features from this location reach a more central representation for further analysis. Based on this

seminal work, there has been a number of heuristics models directly adapted from it, and the proposal by [Itti et al., 1998] can be considered as a reference one.

More recently, a variety of recent approaches have also demonstrated considerable success in formulating the problem as one based on **principled approaches** via either a probabilistic determination or motivated by information theory. Examples include [Gao and Vasconcelos 2009] (discriminant saliency), [Itti and Baldi 2009] (surprise, iLab Neuromorphic Toolkit) and [Bruce and Tsotsos 2009] (attention based on information maximization). Given the desire to maintain an unsupervised definition of salient behaviors, such approaches were a natural fit with the goals of SEARISE.

Estimation of saliency maps based on extended AIM

During year one, we first had to chose which model was more adapted to our needs. To do so, we made a **systematic exploration** of various notions of saliency that are based on a probabilistic determination using a testbed video (video acquired during a soccer game in the LTU –Arena). From these tests, the **attention based on information maximization approach (AIM)** proposed by [Bruce and Tsotsos 2009] seemed to provide superior judgments to competitors subject to a variety of general evaluation. So we chose AIM as the basis for saliency estimation.

AIM is based on the following idea: A sensible strategy may be to direct attention to visual patterns that are informative on the basis of their context according to a principled definition Existing proposals in this regard appeal to measures of local entropy which equates to high activity regions as a selection strategy. A more intuitive approach is in expressing information on the basis of the predictability of a region on the basis of its context, which is the general idea of AIM.

But SEARISE was a very specific problem domain. **Distinguishing features** between saliency computation within the SEARISE project and prior efforts are:

- The variety of the nature of features
- The temporal learning and the role of context which should be extended
- The necessity to combine of different saliency maps

To take into account these characteristics, we proposed during year one a prototype of a very **generic** probabilistic implementation of saliency computation which affords considerable **flexibility** towards achieving desired behaviour within SEARISE. The extended AIM model provides relevant judgments to a variety of SEARISE specific evaluation.

For example, While the literature currently presents a handful of strategies aimed at exploiting task information in a top-down sense, there is little discussion of the role that contextual information plays in the bottom-up determination of saliency. In [Bruce and Kornprobst 2009] we discussed a variety of issues pertaining to the determination of saliency insofar as **context** impacts on the likelihoods involved in its determination. As such, rather than being a proposal for a specific strategy for saliency computation, we instead focused on more general issues that are important in light of the existing models and will be important for any model of saliency based on a probabilistic formulation.

We have applied these ideas to AIM so that it supports arbitrary **types of spatiotemporal context** so that long term operation of the model allows "characteristic behaviors" to be tied to spatial locations at various time scales. An example is given in Fig. 15.

An open source software module: During year two, the extended AIM was implemented in the SEARISE framework. In this development phase, INRIA focused on three main aspects:

- Flexibility of the Surprise module in order to allow investigation of a variety of heretofore unexplored definitions of context and spatiotemporal support, with various types of features.

- Integration: Efforts have been towards the integration and evaluation of the desired saliency computation within the SEARISE framework and consideration of interaction with existing feature modules.
- Optimization of the code, since saliency computation will need to be estimated in real time.

Once these two requirements were met together with the participation of FhG, further testing and evaluation were performed, in conjunction with various components and optimization as the system developed further. This module is now available on-line in the SEARISE website.

Figure 15. Saliency determination based on a SEARISE stadium video surveillance sequence. Top: Saliency given by -log(p(x)) with response likelihood based on an estimate of responses of filters oriented in space-time over the entire scene. Bottom: The response likelihoods are instead associated with each pixel location with likelihoods based on response variation over time. Note the significant difference in what is judged most salient: In the global prior based on the current response of spatiotemporal filters, a quickly waving flag (B) is the most salient target. In the temporal case, a high degree of motion has been observed at this location over time and instead a more subtle movement of a man entering a doorway where movement is not typically observed is deemed most salient (C). There is also a change in the relative saliency of people moving along a more common pathway across conditions (A).

Hierarchical fusion of saliency maps

In SEARISE there are many different representations of what is salient that derive from the various types of features represented in the network. These representations correspond to different types of features, they are derived from several levels of the hierarchy and they have differences in scale, range and sparsity.

So, proposing a **principled means of combining these representations** into a single unified representation is crucial and necessary for the operation of the system, specifically, to tell the Smart Eyes where to look.

Of course, it is entirely possible that a multiplicative weighting of feature maps akin to those employed by [Itti and Koch 1999] is more than adequate for the purposes of this application. That said, it was difficult to predict whether this would be the case in the final composition of the system and as such, it was prudent to include within the system a flexible set of operators that allow behaviour to be adjusted in line with system behaviour.

During year two, in the spirit of the [Itti and Koch 1999] study, we proposed a means of combining saliency maps that retains the simplicity of their approach, with a much broader range of behavior and defined by a principled set of aggregation operators for which an established literature exists: the **fuzzy data aggregation** literature. This flexibility allows the nature of feature combination to be adapted to the needs of the live setup.

Overall it is possible to observe a variety of different behaviors from the operators we have explored. Some general conclusions that we took into account in the final system setup are the following. In light of the results of [Itti and Koch 1999], and given the specialized nature weighted average may be a very sensible choice. The min-like OWA operator as well as some of the t-conorms appear to be of interest. Given the sparsity of the data, the t-norms which generalize the AND operation tend to produce very sparse output and are not stable in general. It is also interesting to consider operations that involve a nonlinear combination of feature maps to observe what additional gains may be had by said operations.

An open source software module: The fusion module was developed and integrated in the SEARISE framework during year two by INRIA. This module encompass the following methods: weighted average, ordered weighted average, t-norms and t-conorms, parameterized t-norms and t-conorms.

1.3.5 Hierarchical architecture and software framework

The development of the integrated neural architecture (Fig. 16) went through several iterations and it outlines the processing hierarchy in the global video stream with output governing the fixation of the binocular cameras. The architecture has hierarchical structure with pipeline comprising several visual processing modules run in parallel threads to reach real time operations. The architecture maintains the segregation between the form (ventral) and motion (dorsal) processing streams. The low-level and mid-level processing utilises hard-wired mechanisms found in the cortical areas V1-V2-MT. these incorporate local interactions of model neural units and iterative feedback interactions between model units corresponding to V1, V2 (form pathway) and MT (motion pathway).

The high-level processing in the form and motion stream utilises flexible learning mechanisms working on subsequent hierarchical levels. Feature learning is essentially unsupervised, with only a few labelled classes added manually as task bias. Saliency map utilises statistical information derived from features with different complexity computed on different hierarchical levels. Fusion of the saliency and categorization maps on the top of the hierarchy is complemented by an attention strategy to provide video record of salient events most suitable for human visual observations.

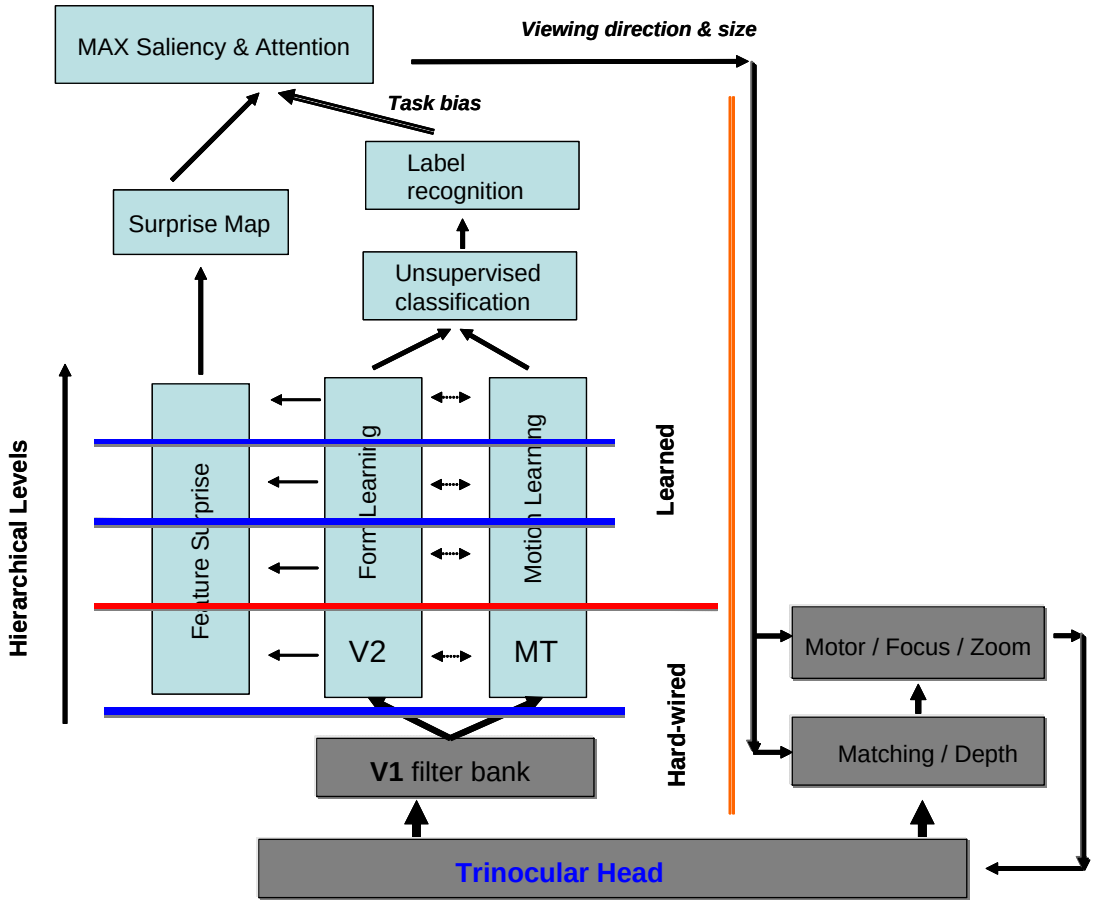


Figure 16. Hierarchical neural architecture of Smart-Eyes system. Smart-Eyes system generates two video streams, a global one and a binocular one, which are processed in real time. Detection of salient events results from the processing of the global stream shown to the left from the red line. The global stream processing is hierarchical, with hard-wired hierarchy levels corresponding to V1-V2-MT, and flexible learning processing mechanisms taking place at higher hierarchies corresponding to IT. Saliency evaluation feeds on feature with different hierarchical complexity and incorporates information about recognized events of interest as task bias. Region with maximum saliency is segmented on the combined saliency map at the top of the hierarchical architecture. Viewing direction towards currently detected most salient region and its size guide the fixation of the binocular cameras.

Software layout reflects the hierarchical neural architecture and performs the following processing steps: 1) Image capturing and pre-processing; 2) Low-level feature computation on GPU; 3) Saliency detection and post-processing; 4) Pattern recognition and task-bias; 5) Fusion of saliency map and recognition results; 6) Simulation of attention behavior; 7) Visualization, video recording and head control. Common software layout can be divided into a hierarchy of six layers (Fig. 17). Each layer comprises closely related processing modules, which are described below.

SEARISE head

The Smart Eyes head provides image streams and receives movement and lens control commands. The headControl module communicates with the PC104, passing the computed position changes to the steering unit.

Low-level GPU image processing

Pre-processing and basic feature extraction is done on the raw image data. The Bayesian Optical Flow algorithm computes motion features on the input video stream sub-sampled by factor 2 to 812 x 618 pixels. The features are smoothed in the temporal domain by an exponentially moving

average strategy with halftime of one frame (i.e. confidence of previous features is reduced by 50% after one frame). Then the resulting features are smoothed spatially by factor 3 to a size of 270 x 206 pixels. All operations are executed on GPU.

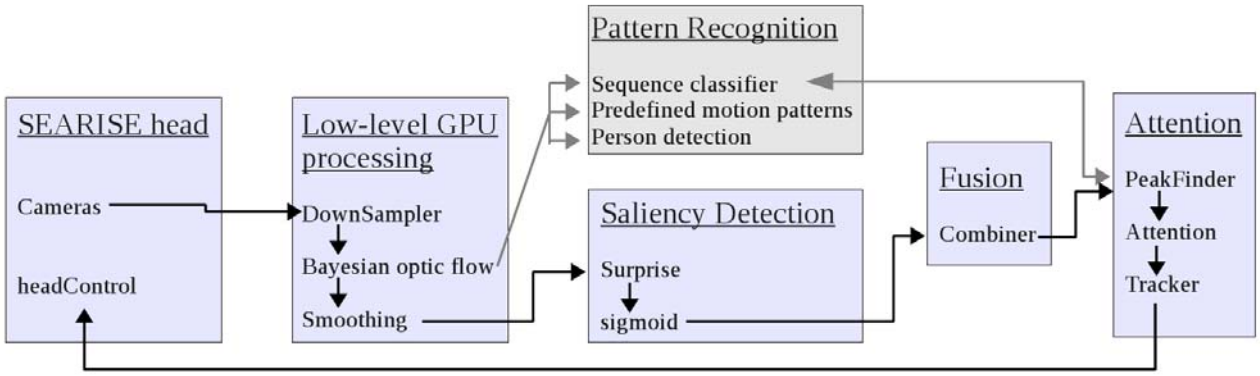


Figure 17. Software layout. Boxes indicate logical clusters and form the hierarchical layout. Each cluster comprises closely related processing modules.

Saliency detection

The surprise module learns basic motion patterns on the optical flow data. The learned motion patterns are decayed over time using an exponentially moving weighted average strategy for storage. The halftime is 6000 frames, meaning that after approximately 10 minutes pattern information is decayed by 50%. A saliency map is computed using the current motion input and the stored pattern memory.

Pattern recognition

The pattern recognition module works on motion features from the optical flow and local shape features. Depending on the task at hands, different recognition detectors are employed to classify either motions of waving flags or people shapes as individuals.

Fusion

This part fuses the resulting maps from the saliency detection and the pattern recognition module. In the long range scenario the saliency map is multiplied with the flag detection map in the predefined motion pattern case. In case of the sequence classifier, the salient regions containing flags are masked out in the saliency map. In the short range scenario regions containing persons get higher priority.

Attention

The high-level processing consumes the fused saliency map and decides which region is to be pursued by the active cameras. The module PeakFinder extracts disjunctive salient regions from the map. The attention module simulates saccadic move. Given one or more salient regions, the module guides whether to follow the current one or to perform a saccade to a new one.

The exact network depends on the scenario and the recognition module. The default behaviour is that the most salient region is extracted and passed to the Attention module. The resulting region is fed into the tracking module and then forwarded to headControl for pursuit.

For the sequence classifier, the PeakFinder the attention module first detects N most salient regions and then the sequence classifier labels each region if it contains the motion pattern of interest (i.e. waving flags). The labelled regions are updated in the saliency map (i.e. flag-containing regions are

masked out) and PeakFinder module detects the most salient region in the updated saliency map and passes it on to the Attention module.

Tracker

The Tracker module receives the position of the region that the Attention module considers most salient. The tracker uses the saliency map and the image to find this salient object again in the following frames. The new position calculated by the Tracker is sent to the control of the active camera (headControl). The explicit tracking of the salient object allows for a smooth movement of the active camera.

1.3.6 System testing and evaluation

For the **long range scenario at the ESPRIT Arena Düsseldorf**, the Smart Eyes system has been mounted on a catwalk above the tribunes. The system is located 30m above the north-east tribune. It faces the southern corner which hosts the most active fans. Average distance between the tribune and the camera system is about 110m.

To generate fixation commands for the active cameras based on the saliency and task bias maps we have worked out and implemented a certain heuristic that is triggered by the maximum saliency or task bias likelihood location (depending on which is larger, saliency or task-bias likelihood). The heuristic then ensures that this location will be fixated for a while by providing the corresponding location with a saliency boost, thus biasing the competition towards staying at this location. This boost decays over time, also becoming negative after a while which effectively implements inhibition-of-return which avoids few high-saliency locations from dominating the fixation. An example fixation path on a video stream from the Arena scenario is shown in Fig. 18.

Pattern recognition employs real time detection of waving flags. The achieved detection of rate is about 94%. Two algorithms– the sequence classifier and the classifier using pre-defined complex motion patterns have been developed and extensively tested on the Arena videos. Both algorithms produce only rare false negatives, meaning waving flags which were not detected as such. These occurrences appeared mostly at transitions to different motion sequences, for example when flags were extracted or taken down.

Fixation and tracking executed by the active camera shown to be precise enough to centre targeted salient region and to hold it in the field of view. The delay between the instance of image acquisition and the completion of computations for fixation update is less than 250ms. The delay is sufficient to follow motion of a single person at a maximum zoom level. However, the delay produces a certain lag between centre of the active camera image and the targeted salient region, which slightly decreases the quality of visual perception. Fixation in the Arena requires no depth evaluation as depth variations across the observed tribune are negligible as compared to the large distance to Smart Eyes. En example of Fixation is shown in Fig. 19.

For the **short range scenario on the city train station the** Smart Eyes system was mounted on one of the pillars supporting the station roof. The processing pipeline implements fixation of the most salient region consistently containing persons. The image analysis performs consistently at around 5.5 fps which proved sufficient for the coordination of the active cameras. Recording of video material takes place at approximately twice that rate. To deal with perspective problems arising from the short range perspective we applied a rough perspective correction. We also implemented the template-matching based fixation strategy to correct for fixation errors arising from significant depth changes in the scene, relative to the camera-to-scene distance. The latter has been implemented only in a stand alone application as the large network latency and latencies introduced by the analysis introduce problems for closed loop control and also template matching is a compute-heavy operation, taking away resources from the main analysis tasks. Fixation based on analytically

pre-computed formulas proved to be sufficient. Example images of the Smart-Eyes system performance, in particular fixation and zooming is shown in Fig. 20 (face is obfuscated to hide identity in compliance with ethical requirements on the privacy protection).

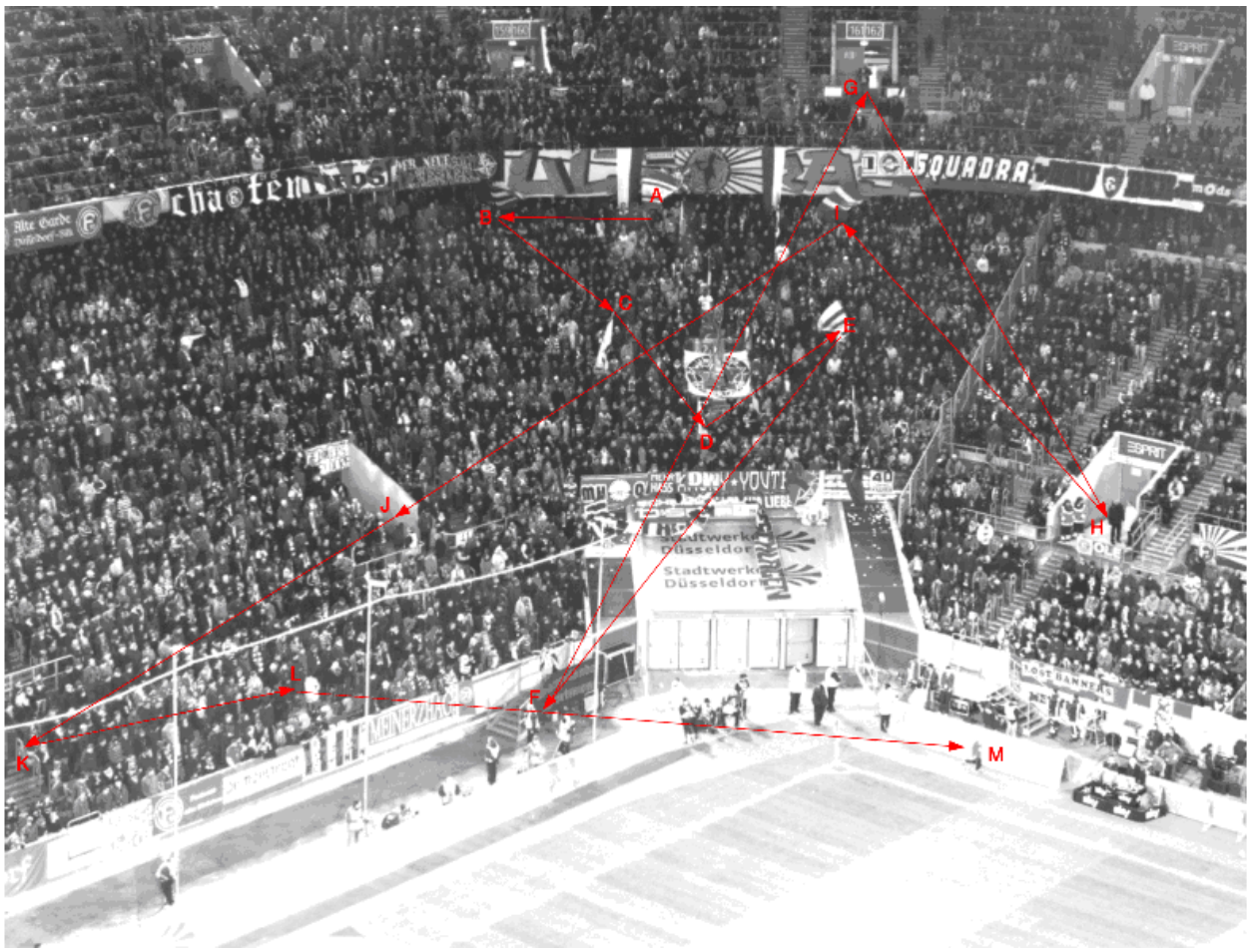


Figure 18. A fixation path generated by the heuristic on a recording of the stadium spectator ranks by the SEARISE system. Detected events: A,B,C,D,E,I: flag-waving; F: Movement of security personnel; G,H,J: new movement in entrance area; K,L: persons entering/leaving ranks; M: Player starts warming up.

Figure 19. Fixation example: The left image shows the global view and the most salient region (red box) fixated and tracked by the active camera, which image is shown on the right. The active image is acquired at a maximum zoom level.

Figure 20. Smart Eyes fixating a salient region containing several persons in the city railway station scenario shown as region within the red box in the global camera view (left) and as recorded by one of the active cameras (right).

Human psychophysics and benchmark creation

For both scenarios above we have conducted psychophysical experiments and generated appropriate benchmarks. A benchmark for behavioural pattern recognition should be comprised of a collection of videos pertinent to the application domain as defined for the SEARISE system. For the **long-range Arena scenario**, one would therefore like to use recordings from real soccer matches as benchmark data. However, this approach is problematic for two reasons: security-relevant events are rare, and legal constrains prohibit their storage beyond a short time interval.

Figure 21. Two frames from the SEARISE benchmark dataset, actor identities obfuscated. *Left*: waving crowd, not security relevant. *Right*: brawl, a security-relevant event.

To deal with these problems, we created a benchmark dataset, the ***Tübingen hooligan simulator***, by staging relevant events. We contacted officers of the Düsseldorf police in charge of stadium security at the Esprit arena to obtain the expert knowledge necessary to decide which events to include. We are particularly indebted to Polizeihauptkommissar (PHK) G. Mainda and H.J. Berg for their expert knowledge and their participation in our experiments.

Subsequently, we staged typical security-relevant and background events in a lecture theatre with a group of ≈ 10 lay actors. We repeated each event multiple times in different parts of the lecture theatre, The resulting videos were overlaid to create the impression of a larger crowd. Two frames from the videos are shown in Fig.21.

We showed these videos to the police officers, asking them for feedback with regard to realism, completeness and label correctness. The full list of security-relevant and normal events which comprise the *Tübingen hoooligan simulator* is listed in Table 2.

Normal	security-relevant
waving arms	brawl: crowd, converging and embedded
waving flags	fast dispersal
Hopping	moving up/down over seats
Swaying	pushing others (one or many)
angry gesturing	walking along filled seat rows
Sitting	vandalism against chairs
Standing	throwing objects
orderly exiting/entering	lighting and passing bengal torches

Table 2. List of normal and security relevant events for the Arena scenario.

We have obfuscated actor identities in all videos in preparation for a publication of the benchmark data.

For a more quantitative evaluation of our video set with respect to the relative contributions of saliency and expert knowledge when searching for security-relevant events, we also conducted eye-tracking experiments with the officers. We acquired a **Tobii X120** mobile eye-tracking system, largely with SEARISE funds. To construct a realistic and sufficiently difficult detection task, we built visual scenes by embedding the security-relevant events in neighbourhoods of normal events. This was accomplished by randomly filling a 7×7 grid of event patches with normal events. Spatial contiguity between patches was promoted by populating the grid with events drawn from a Markov random field with nearest-neighbour interactions. In half of the scenes thus generated, we placed a security relevant event somewhere on the grid. Two example scenes are shown in Fig. 22.

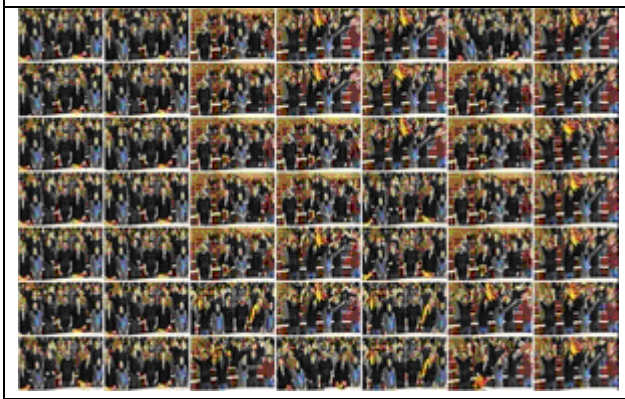
Normal background behaviour	Security-relevant event: brawl
	

Figure 22. Two examples from the eye-tracking stimulus scenes. *Left*: Scene comprised of only normal background events. Right: scene with security-relevant event (brawl), highlighted with green frame. Frame not visible during experiment.

Each scene had duration of ≈5s, 20 scenes were presented in one eye-tracking session. Presentation was halted after each scene to allow the officer to provide verbal feedback as to whether he had perceived a security-relevant event and if so, what type of event it was. These responses were recorded by us as well. Furthermore, we are also testing naïve observers on this task.

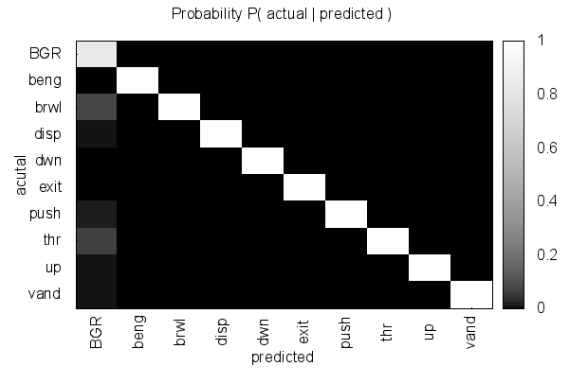
To elucidate the relative contributions of expert knowledge and low-level saliency, we computed several **gaze statistics**, including: fraction of fixation of security-relevant events, their fixation times, and time until fixation.

Result: [Endres et al. 2010b] we found only weak differences in most parameters that characterise fixation behaviour. Expert and naïve observers use a very similar fixation strategy. Our interpretation: both are driven by ***saliency*** within the first 5 seconds of this search task. We therefore decided to use saliency and classification ***hierarchically:*** the scenes are pre-processed by the saliency tracker, only tracked regions will then be classified. This approach significantly reduces the computational burden associated with classification of scenes.

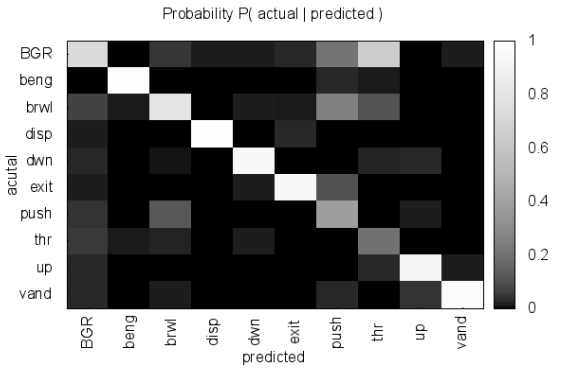
Psychophysics: behavioural data and comparison to SEARISE sequence classifier

A comparison of the classification rates and detection rates computed from the verbal responses of the subjects is shown in Fig.23.

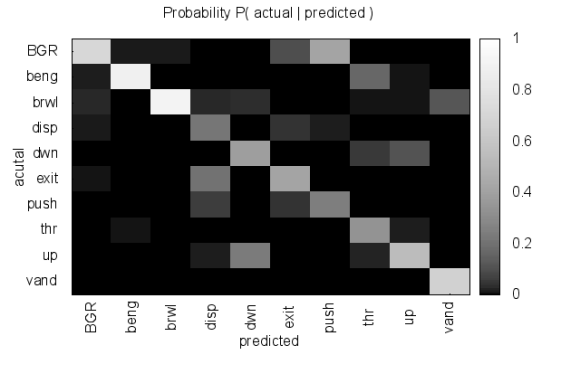
Confusion matrix: expert



Confusion matrix: naïve observer



Confusion matrix: sequence classifier



Classification and detection rates

	expert	naïve	seq. class.
classification	0.90	0.75	0.75
detection	0.90	0.79	0.90

Figure 23. Classification rates and detection rates of security-relevant events on the *Tübingen hooligan simulator*. Detection rate measures the fraction of correct classification between either security-relevant or normal background events. We compared naïve observers to the SEARISE sequence classifier module and a security expert. This expert is performing better than the average naïve observer on the majority of security-relevant events, indicating that our benchmark dataset captures events which experts are trained to detect. Event labels as above. Rates are conditional on the saliency module finding the location of the relevant event, if one was present.

Result 1: expert achieves better classification performance than naïve observer, despite high similarity in fixation patterns. The main errors committed by expert are **misses** of security-relevant events, but **hardly any false alarms**.

Result 2: SEARISE sequence classifier module is worse than human expert when trying to separate all classes, but approaches human performance when distinguishing security-relevant events from normal background events.

Long-range scenario: *task-bias modulation* on longer scenes from the Arena

The above results are conditional on the first ≈5s of our search task. We therefore wondered if experts would employ top-down *task-bias modulation* to guide their attention when watching longer scenes in a realistic setting. To address this question, we selected video sequences (2 minutes long) from recordings made with the SEARISE camera in the Arena during relevant phases of soccer matches. Relevant phases were determined by interviewing experts, see deliverable report D10.7 for details. While the subjects (both expert and naïve) watched the scenes, we recorded their gaze patterns.The subjects were asked to provide a verbal description of *why* they looked *where* and *when*. Two average expert gazemaps from different game phases are shown in Fig.24.

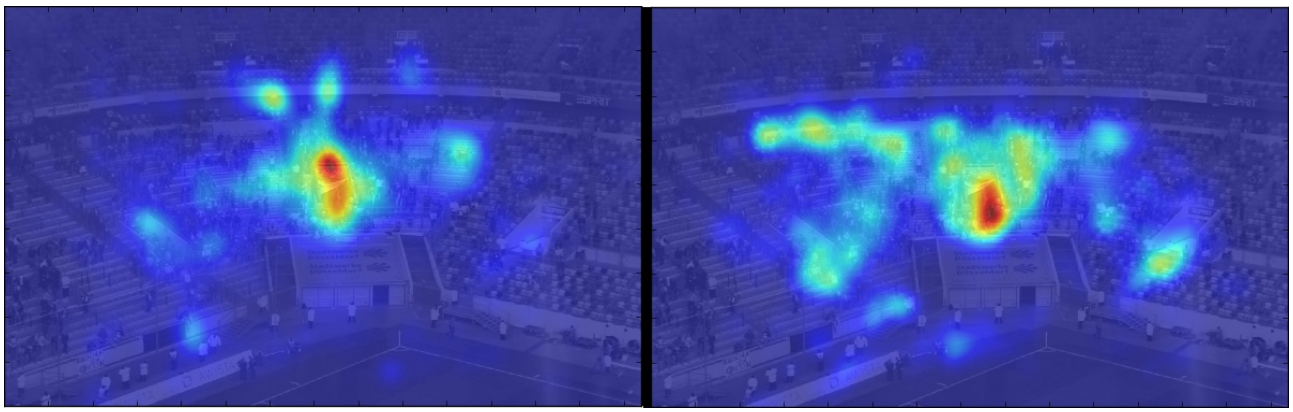


Figure 24. Average expert gaze maps from different game phases, superimposed on Arena grandstand. Gaze durations are colour-coded. Red: long/frequent fixation, blue: short/infrequent fixation. *Left:* during the game. The experts focus on the ***Ultra*** fan block, which is known to be a location of security-relevant events. The entrances are less important during this game phase. *Left:* after the game. The experts now dedicate more of their attention to the entrances, while people are exiting the grandstand.

We found a game-phase dependent spatial modulation of the gaze patterns. These results allow us to compute static saliency modulation maps (per game-phase) which can be integrated into the SEARISE system.

Short-range scenario: subway station at the Arena

The SEARISE system is also operating in a short-range mode in the city train station nearby the Arena. In collaboration with security experts from the Rheinbahn, we have compiled and staged a collection of **more than 200 instances** of security-relevant scenes, which can serve as a benchmark in this scenario. Fig. 25 shows 4 relevant scenes. The coloured frames highlight the regions which are tracked by the SEARISE attention module. Evaluation of the SEARISE system in the short-range scenario is still ongoing.

assault	vandalism
	
drug deposit under garbage bin	**urinating against lamp post**
	

Figure 25. Events staged by the SEARISE partners in the city train station scenario.

Event	relevance	relative frequency
Assault	3	0.004
Brawl	3	0.030
person on tracks	3	0.003
emergency, first-aid situation	3	0.013
molestation, threat	2	0.135
rampage	2	0.067
vandalism	2	0.001
soliciting	2	0.069
emergency stop of escalator	2	0.003
presence of drug users/dealers	1	0.264
drug consumption/dealing	1	0.021
Loitering	1	0.354
fouling(urinating, vomiting)	1	0.036

Table 3. The complete list of relevant scenes, ranked by their relevance.

1.4 Potential impacts

DIBE – Vision modules for the active trinocular system

The efficacy of neural (i.e., distributed) algorithms has usually to cope with the efficiency of the hosting (non-neural) computational architectures. Indeed, although the performances of these models were promising, they have never been *largely* employed in real-world applications. This is mainly due to their high computational cost. The major impact of the work conducted by partner DIBE concern the demonstration that efficient implementations of neural-based modules can be operative in real-time and real-word conditions. By adopting specific design strategies, it is indeed possible to implement a *neuromorphic* solution for stereo (and motion) vision problems that is characterized by an affordable computational cost. The *higher flexibility* of the neural solution does not come at the price of lower performances, being in some cases even more effective than its "non-neural" counterparts (e.g., with respect to noise resistance).

Potential impact: The GPGPU-based implementation of the distributed architecture for the computation of the full (i.e., 2D) disparity, using Nvidia CUDA Library, demonstrates that cortical-like visual modules can be employed in real-world everyday life applications. The current implementation process up to 10 512x512 pixels frames per second, by using a population of 72 binocular cells and 6 spatial scales. The use of the this module to control the binocular active movement of the SmartEyes camera systems made possible a system validation of the approach in real-world condition, to demonstrate that a bioinspired solution can be used as a systematic alternative to computer vision, working on raw images and real video sequences. Concerning the active vision system, the operative feasibility of the mapping between the fixed wide field of view and the active zoomed field of view demonstrated the possibility of using a multicamera system to emulate human-like foveations. The major features of the system are: rapid (open-loop) reactions to change the direction of gaze [with an angular velocity of about 3 deg/sec] and slower (closed-loop) vergence to 3D fixation on the basis of the computed disparity. In the worst case, the angular position control (gaze+vergence) reaches the steady state in about 3-4 sec.

Main dissemination activities: The approach and the results have been presented at Conferences and Workshops of the Computer Vision Community, such as: International Conference on Computer Vision Theory and Applications, International Conference on Cognitive and Neural Systems, International Conference on Computer Vision Systems, as well as at the 4th International Conference on Cognitive System (CogSys2010), and reported in peer-reviewed journals and edited book chapters (as detailed in the tables below).

Exploitation of results: The specific design approach followed to implement the distributed architecture demonstrated that it is possible to implement 'neuromorphic' solutions that are characterized by an affordable computational cost, to be efficiently employed in closed-loop robotic applications. Therefore, cortical-like architectures as bio-inspired structural paradigms to solve computer vision tasks, can represent a viable solution for the next-generation robot vision systems, which are capable to calibrate and adapt autonomously through the interaction with the environment. In this direction, we are planning to exploit the distributed character of processing and representation offered by our architecture to build perceptual modules where the necessary entry points for closing perception-action cycles from the very initial processing stages.

UUlm

UULM contributed knowledge in the field of biological plausible models of motion and form processing and contributed various software modules of motion and form processing, motion integration and figure-ground segregation to SEARISE. In the course of the project, various approaches have been improved (motion estimation), merged (form and motion processing) or created (depth from motion). Throughout the period, we laid increased emphasis on the capabilities

of the modules to process real-world scenarios, especially the two scenarios that are within the main focus of SEARISE.

In [Weidenbacher & Neumann, 2009], UULM proposed a recurrent model of V1-V2 interactions for the extraction of surface-related features. Humans can effortlessly segment surfaces and objects from two-dimensional (2D) images that are projections of the 3D world. The projection from 3D to 2D leads partially to occlusions of surfaces depending on their position in depth and on viewpoint. One way for the human visual system to infer monocular depth cues could be to extract and interpret occlusions. It has been suggested that the perception of contour junctions, in particular T-junctions, may be used as cue for occlusion of opaque surfaces. Furthermore, X-junctions could be used to signal occlusions of transparent surfaces.

This methodology was picked up in [Beck & Neumann, 2010], where UULM investigated interactions of form and motion in visual cortex. In the work, we present a neural model simulating parts of motion and form pathway of visual cortex. It is shown how the visual features motion, disparity, and form that are represented in a distributed way in areas V1, V2 and MT mutually interact at several levels. We suggest that here information of the form channel, namely the indication of a junction, is necessary to achieve a correct percept in the motion pathway. The model simulations reproduce psychophysical and neurophysiologic results of the chopstick illusion as well as of the barber pole illusion. The temporal course of the dominant motion percept generated by the iterative interplay between motion and form pathway is in line with data of ocular following responses in primates and humans.

In [Bouecke et al., 2010], we analyzed neural mechanisms of motion detection, integration and segregation. We utilize such principles and propose a framework for modeling neural computational mechanisms of motion in primates using biologically inspired principles. In particular, we investigate motion detection and integration in cortical areas V1 and MT utilizing feed forward and modulating feedback processing and the automatic gain control through center-surround interaction and activity normalization. We demonstrate that the model framework is capable of reproducing challenging data from experimental investigations in psychophysics and physiology. Furthermore, the model is also demonstrated to successfully deal with realistic images sequences from benchmark databases and technical applications.

Furthermore, we investigated how to improve motion estimation results. In [Ringbauer et al., 2010] we proposed a new methodology of incorporating a new feature domain into the processing cascade: In visual cortex information is processed along a cascade of neural mechanisms that pool activations from the surround with spatially increasing receptive fields. Watching a scenery of multiple moving objects leads to object boundaries on the retina defined by discontinuities in feature domains such as luminance or velocities. Spatial integration across the boundaries mixes distinct sources of input signals and leads to unreliable measurements. Partner INRIA proposed a luminance-gated motion integration mechanism, which does not account for the presence of discontinuities in other feature domains. In this contribution, we propose a biologically inspired model that utilizes the low and intermediate stages of cortical motion processing, namely V1, MT and MSTd, to detect motion by locally adapting spatial integration fields depending on motion contrast. This mechanism generalizes the concept of bilateral filtering proposed for anisotropic smoothing in image restoration in computer vision.

These contributions demonstrate that biologically inspired approaches are capable of competing with technological approaches and that they still have potential for further improvements, especially regarding the growing capability of parallel processing on consumer hardware. Here, using standard software, an equivalent increase in execution speed can not be taken for granted, and porting software to parallel processing is a nontrivial task and not always successful. Biologically inspired models on the other hand can, due to their parallel architecture, better scale with multiple execution

threads, and thus profits from the obvious trend towards parallel computing. This has been demonstrated with the implementation of the Full Neural Model using the GPU-based library. Here, we demonstrated how real-time execution of neurally inspired models becomes more feasible when parallel processing power is efficiently used.

Offering proper evaluation methodology is essential to continue progress in modeling the neural mechanisms involved in vision information processing. Before SEARISE, the evaluation of biologically inspired motion estimation models lacked a proper methodology for comparing their performance against behavioral and psychophysical data. In [Tlapale et al., 2010], we set the basis for such a new benchmark methodology based on human visual performance and designed a database of image sequences taken from neuroscience and psychophysics literature. Here, eye movements and perceived motion will serve as a reference to compare simulated and experimental data. We provide the basis for a valuable evaluation methodology to unravel the fundamental mechanisms of motion perception in the visual cortex. This database is freely available on the web together with scoring instructions.

The Algorithmic Model that was included in the SEARISE framework has also been continuously extended with findings during the period and published as a stand-alone version, including documentation. It caught attention of various groups, that now make use of the fast biologically motivated algorithm of motion estimation that was developed for SEARISE.

In addition, we see the necessity of representing and processing motion transparency regarding the SEARISE scenario in a two-fold manner: Mutual occlusion of objects and people in crowded scenes generate multimodal velocity representations in a small neighborhood which need to be segregated. Also in crowded scenes collectively moving groups of people generate multiple bands of contiguously moving patterns of possibly different widths and directions. The same core model architecture has thus been extended in order to robustly process inputs of transparent motion. Examples relevant for the project scenario are crowded scenes in which multiple motions of relatively small objects result in flow patterns of different motion directions. We argue that the ability to robustly handle transparent motion is essential for the development of general purpose vision systems that operate in real-world scenarios where mutual occlusions and semi-transparent configurations occur regularly. In [Raudies & Neumann, 2010], we proposed a model of neural mechanisms in monocular transparent motion perception in context with SEARISE.

The two SEARISE scenarios, namely the long-range and the short-range scenario, revealed different aspects of visual processing. In the long-range scenario, expected motion is of little velocity, but with high complexity. Here, a perception of transparent motion can easily occur and disparity information is of almost no use due to the limited baseline in comparison to the scene's distance. In contrast, in the short-range scenario, motion is expected to have higher velocities and more homogeneously moving regions. The disparity channel is of great interest here, as it contains clues about the scene's segmentation in depth. With partner DIBE/Unige, we showed how interaction of motion and disparity information can improve estimation of disparity in a neurally inspired network. In addition to the segmentation of the scene using disparity, we proposed a neurally inspired method of estimating a scene's depth using the motion pathway in [Tschechne & Neumann, 2011a|b]. The presence of temporal occlusions generated by surfaces hovering at different depths is evidenced by a transition from high motion energy to low motion. We propose an opponent scheme of temporal on/off interactions in which local motion energy signals from model V1 are spatially integrated by the temporally offset on/off subfields. The model was probed with artificial and real-world scenes of moving and mutually occluding object surfaces. The (dis)occlusion boundary responses together with directional motion signals determine the border-ownership direction of an occluding surface. This demonstrates that spatio-temporal figure-ground

separation can be achieved by local mechanisms at early and intermediate stages of the dorsal visual pathway

The SEARISE scenario offered a great opportunity to prove that biologically inspired mechanisms of visual processing are also applicable to real-world scenarios, and that quality of results can keep up with technical approaches. Research progress of UULM has shown this as well as the fact that results are still improvable with innovative models. UULM made far-reaching contributions to the body of knowledge in the field of cortical models and visual processing. Still, various topics that could not be finished during the project's period will be continued over the intervening months, or years. These activities will benefit from the relationships between the partners gained over the past three years.

INRIA

Within SEARISE, the two main contributions by INRIA were related to motion and saliency estimation. In this two domains, SEARISE was a great opportunity for us to confront our ideas in a real scenario, but also it allowed us to propose with SEARISE partners new ideas with a high potential impact.

Motion estimation

Within SEARISE, one of the key low-level feature was motion estimation. Since the core of the artificial cognitive visual system is a computational model of visual processing in the brain, we focused on investigating bio-inspired solutions for motion estimation. More precisely, we showed how biology can be a source of inspiration, from modelling to evaluation. Three characteristics of our work should have an impact on subsequent research in this direction.

- Together with UULM, we made substantial progress towards bio-inspired models, focusing on new features of motion integration: In proposing bio-inspired models for motion estimation, UULM had already a strong expertise in this area. SEARISE allowed to do some strong progress in this direction. UULM and INRIA started a very fruitful collaboration proposed new approaches to improve the state-of-the-art. For example, let us mention two contributions. The first contribution was to propose addition information derived from static form pathway in order to enhance the motion integration process, particularly at boundaries and, thus, to improve the quality of the computed optical flow. A solution was initially proposed in [Tlapale et al. 2010 (*Vision Research*)], and recently extended by UULM in [Ringbauer et al., 2011]. The second contribution was to consider motion integration dynamics as a new question to investigate for modellers. Looking at the literature, previous work considering bio-inspired models were based on coupled differential equations for which the steady state was only considered. In [Tlapale et al. 2010], we carefully investigated the dynamics of motion integration which was a question never addressed in the literature. We successfully reproduced the temporal dynamics of motion integration on a wide range of simple motion stimuli: line segments, rotating ellipses, plaids, and barber poles. This work should be a reference for subsequent contributions interested in motion integration dynamics.
- We proposed a mathematically well-posed model in the neural fields framework which opens new perspectives for motion analysis: In order to describe cortical activity at the population level, we proposed the neural field framework as a continuum approximation of the neural activity. Since the seminal work by [Wilson and Cowan 1972, Amari 1977], intensive research has been carried out to extend models and study them mathematically. However, up to our knowledge, there were hardly any efforts to apply this formalism to real modelling problems such as motion estimation. In [Tlapale et al. 2011, Tlapale et al. 1010b (*submitted to IJCV*)], we use this formalism for the problem of motion estimation, proving its well-posedness. Interestingly, we also showed intriguing properties of this model, such as multistability

phenomena [Rankin et al. 2011]: This will certainly have a high impact in the modelling community since it proves the richness of such formalism.

- We showed how biology can be a source of inspiration for benchmarking and we proposed a new evaluation methodology which is available on-line: Following the results shown in [Tlapale et al. 2010] where we compared our results to biological data, INRIA and UULM investigated how to set up a new kind of benchmark to evaluate models against visual system performance. This new evaluation methodology was presented at the international conferences VSS 2010 [Kornprobst et al. 2010] and Bionetics 2010 (Special Track on Bio-Inspired Machine Vision) [Tlapale et al. 2010]. The proposed standardized tools are provided on-line in order to identify the necessary and sufficient mechanisms involved in motion processing. These tools allow researchers to compare different approaches, and to challenge their models of motion processing. In term of impact, since offering proper evaluation methodology is essential to continue progress in modelling, we are convinced that this benchmark will help progress in modelling neural mechanisms in visual information processing.

Estimation and hierarchical fusion of saliency maps

In recent years, many principled probabilistic definitions for the bottom-up determination of visual saliency have been proposed. Moreover, there has been increased focus on the role of context in the determination of visual salience. While prior efforts focus mainly upon the manner in which context aids in predicting the location or presence of features associated with an object in the context of object detection or recognition, there has been little focus on the manner in which context impacts upon the purely bottom-up side of visual saliency computation. In this light, and as a major scope in SEARISE, INRIA investigated **the role of context** in the determination of salience insofar as it determines low-level stimulus driven visual salience.

The potential impact of our work comes from the novel aspects of the work we proposed. This includes expressing definitions of saliency within a **common unified framework** and exploring various notions of context including their amenability to application towards the goals of SEARISE. This also includes consideration of different spatiotemporal volumes of support upon which the determination of saliency may be made, and also in considering more general definitions of context (e.g. environmental factors) and the extent to which these might be levied for the purpose of detecting salient, important, or suspicious behaviours (see [Bruce and Kornprobst 2009]).

As a direct consequence of our contribution but more generally of the SEARISE project goal, the interest of using saliency estimation in a video surveillance application has been demonstrated. Compared to the state-of-the-art video surveillance systems, the SEARISE contribution will certainly influence future developments in this direction. These developments will be also facilitated by the two related modules (Surprise and Fusion modules) which are available on-line as open source libraries.

UniTue - – benchmarking and motion pattern classification

The main contributions of partner **UniTue** to the SEARISE effort are: creation of benchmarking and training data for the SEARISE system, psychophysical investigations to elucidate the roles of saliency and expert knowledge, and real-time motion pattern recognition capable of online learning.

Psychophysics: The detection of security-relevant events in large crowds is a difficult vision problem. We investigated differences in visual search of dangerous events between security experts and naive observers during the observation of large scenes, typically encountered on the grandstand of stadiums during soccer matches. Based on a new algorithm for the synthesis of crowd scenes with well-controlled statistical properties, subjects were eye-tracked during the observation of such scenes. Detection rates, fixation rates and times were assessed from naïve subjects and expert observers. We found only weak differences in most parameters that characterize fixation behaviour, but large differences in detection and classification accuracy of security-relevant events. This result

highlighted the importance of *saliency* for the initial guidance of visual search in a surveillance system. We also found that in longer visual searches, *top-down* task-bias knowledge modulates saliency. Furthermore, our work establishes the performance criteria which an automatic surveillance system should meet to be acceptable by expert users. This is important for future industrial applications.

Benchmarking: In the domain of motion recognition, many well-known benchmarks are overly simplistic and have little relevance from a surveillance perspective. We created benchmarking (and training) datasets with expert feedback which have controlled statistics of security-relevant events, are free of legal constraints and reflect the natural variability of such events. The benchmarking data for the **long-range scenario** have already been prepared for publication and will be made available at a suitable outlet, e.g. the next PETS workshop. The data for the **short-range scenario** are currently being post-processed. Once finished, they will be published, too.

Motion pattern classification: our real-time motion pattern classifier is built from the **varmod** library. We designed this library in collaboration with partner **Fraunhofer** with the aim of providing a flexible means for assembling a range of hierarchical Bayesian models capable of on-line learning. The library also implements *non-parametric* models, allowing for the on-line growth of new components as more data become available. Furthermore, we included efficient parallelization facilities to meet the real-time constraint. This library has been published on the SEARISE website under an open-source license, and should be useful for anyone wishing to build such models. We compared our classifier to human experts. Moreover, we also investigated machine learning approaches mimicking the way in which humans segment action streams in and investigated the impact of correct segmentation on machine recognition performance in.

Webpage: we created a webpage which summarizes our contributions to SEARISE. See http://www.compsens.uni-tuebingen.de/index.php?page=project&id=20

Fraunhofer – Smart Eyes system

Smart Eyes is a unique system in terms of its functionality: It is first operational system of its kind capable of active recognition of saliency, followed by its active fixation and recording with all these operations carried out in real time. This opens up a broad variety of applications in which Smart Eyes can offer dedicated surveillance services.

Focus on saliency

Video record of a typical monitoring camera contains mostly common events with only a few frames showing events of particular interest. Same is true for snapshots taken in a crowded place: if anything unusual happens here it takes only a small fraction of the whole scenery and can easily go unnoticed amidst numerous other activities. Smart Eyes is capable to learn what is salient in a given scene by building observation models of activities. The learned models enable the automatic detection of highly salient events in real time.

Security relevant events

Saliency is a measure of relative novelty. Therefore salient events may not be security relevant. SEARISE software can learn from a few video samples of security relevant events specified by experts. SEARISE software then uses the learned knowledge to analyze all salient events in the scene and to detect the specified security events. The software can update its knowledge about the learned events via automatic on-line learning. Additional semi-automatic learning allows incorporation of new security relevant events into the previously learned model.

At a steady pace with events

Searching for the security relevant events within a large scene is a tedious task. The search in high resolution overloads the CPU and is prone to false alarms. By reducing the search space to the

salient regions only, SEARISE software analyses events in real time while providing dramatic drop in false alarms. High resolution video of salient regions shown in real time supports onsite human monitoring. High resolution video record of the automatically detected salient regions provides indispensable evidence of their later investigation if needed.

Application areas

Computer video analysis is often challenged by an infinite variety of video appearances. Yet SEARISE software continuously learns these natural video variations thus adapting its on-line processing to ever changing illumination, weather or other conditions. This inherent flexibility allows application of the SEARISE software in very different indoor and outdoor scenarios. The software application ranges from observation of crowded public places to monitoring of secluded zones with restricted entering rights.

Easy deployment

With only a few parameters SEARISE software self-adapts to the video input. The SEARISE technology requires neither camera calibration nor sophisticated installation. Built upon the generic principles of saliency and learning, the SEARISE software is fully independent from the type of a camera and operates in different application scenarios.

Software modules and detection scenarios

SEARISE software modules are freely combined into a processing chain that is optimal for a given application scenario and a problem at hands. All software modules operate unconstrained in outdoor and indoor environment.

	Restricted access zones		Auto tunnels	Public transport infrastructures: airports, subways, malls, etc.			Sport, concert arenas
	Unusual activities	People detection	Driving irregularities	Incoherent motion	Left item	Unusual activities	Rule violations
Adaptive background segregation	●	●	●		●		
Dense motion flow	●		●	●	●	●	●
Object shape features		●			●		
automatic saliency detection	●		●	●		●	●
Object shape learning & detection		●			●		
Object tracking		●	●		●		●
Motion pattern learning & detection	●			●		●	●

Implementation details

SEARISE software is written in portable c++, has modular structure and exploits multiple levels of parallelization. The software is easily embedded by different SDKs to process video input from various cameras.

Operating system	• Windows (including **.NET**) • Mac • Linux
Hardware platform	• modern CPU • Cuda-capable nVidia GPU • multiple GPUs and/or CPUs
Run-time configuration	• Text configuration files (.json format) • Python scripting

Listed above special features of Smart Eyes simplify its adaptation to the needs of commercial video surveillance systems. Fraunhofer is keen to work on the commercialization of Smart Eyes system after the project end.

1.5 Project data and contact details

The official SEARISE project web-site: www.searise.eu.

SEARISE Logo:

Project contact details:

Dr. Marina Kolesnik

Fraunhofer Institute for Applied Information Technology,

Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Phone: +49-2241-143421

Fax: +49-2241-141506

http://www.fit.fraunhofer.de/~kolesnik

1.6 References

[Adelson Bergen (1985)] E.H. Adelson and J.R. Bergen. *Spatiotemporal energy models for the perception of motion*. J. Opt. Soc. Amer., vol. 2, pages 284–321, 1985.

[Andrade et al., 2005] Andrade E., Blunsden S., Fischer, R. *Simulation of crowd problems for computer vision.* In: first international workshop on crowd simulation (V-CROWDS'05), Lausanne, Switzerland, pp. 71-80, 2005.

[Amari, 1977] S.-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics , 27(2):77–87, June 1977

[Baker et al. (2007)] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black and R. Szeliski. *A database and evaluation methodology for optical flow*. IEEE Int. Conference on Computer Vision, 2007.

[Bartlett et al., 2002] M.S. Bartlett, J.R. Movellan and T.J. Sejnowski. *Face recognition by independent component analysis*. IEEE Transactions on Neural Networks 13(6) p. 1450-64, 2002.

[Bayerl & Neumann, 2007] Bayerl, Pierre, and Heiko Neumann. *A fast biologically inspired algorithm for recurrent motion estimation*. IEEE transactions on pattern analysis and machine intelligence 29, no. 2, 2007.

[Beck & Neumann, 2010] Beck C, Neumann H. *Interactions of motion and form in visual cortex – A neural model*. J. of Physiol. Paris 104, 61-70. 2010.

[Beck et al., 2008] Beck, Cornelia, Thilo Ognibeni, and Heiko Neumann. *Object segmentation from motion discontinuities and temporal occlusions--a biologically inspired model.* PloS one 3, no. 11, 2008.

[Bell & Sejnowski, 1996] A.J.Bell and T.J Sejnowski. *Edges are the `independent components'of natural scenes*, Advances in Neural Information Processing Systems 9, MIT press, 1996.

[Bishop, 2007] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007.

[Blank et al., 2005] Blank M., Gorelick L., Shechtman E., Irani M., Basri R. *Actions as space-time shapes.* In: the Tenth IEEE International Conference on Computer Vision (ICCV'05), pp. 1395-1402, 2005.

[Bouecke et al., 2011] Bouecke, Jan D., Emilien Tlapale, Pierre Kornprobst, and Heiko Neumann. *Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems*. EURASIP Journal on Advances in Signal Processing 2011.

[Bruce and Tsotsos, 2009] N.D.B. Bruce and J.K. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. Journal of Vision , 9(3):1–24, 2009.

[Bruce and Kornprobst, 2009] N. Bruce and P. Kornprobst. On the role of context in probabilistic models of visual saliency. In Proceedings of the International Conference on Image Processing . IEEE Signal Processing Society, 2009.

[Cardoso & Souloumiac, 1993] J-F. Cardoso and A. Souloumiac. *Blind beamforming for non Gaussian signals*, In IEE Proceedings-F, 140(6):362-370, 1993.

[Chessa et al. (2009)a] M. Chessa, F. Solari and S.P. Sabatini. *A Virtual Reality Simulator for Active Stereo Vision Systems.* International Conference on Computer Vision Theory and Applications, Lisbon 5-8 February 2009.

[Chessa et al. (2009)b] M. Chessa, A. Canessa, A. Gibaldi, F. Solari and S.P. Sabatini. *Embedding Fixation Constraints into Binocular Energy-based Models of Depth Perception.* International Conference on Cognitive and Neural Systems, Boston 27-30 May 2009.

[DeAngelis et al. (1995)] G.C. DeAngelis, I. Ohzawa and R.D. Freeman. *Receptive-field dynamics in the central visual pathways*. Trends in Neurosci., vol. 18, pages 451–458, 1995.

Endres et al, 2008] Endres D.M., Oram M.W., Schindelin J., Földiák P. *Bayesian binning beats approximate alternatives: estimating peri-stimulus time histograms*, pp. 393-400, Advances in NIPS 20, MIT Press, Cambridge, MA, 2008.

[Endres & Giese, 2009] Endres D.M., Giese M.A. *Temporal Segmentation with Bayesian Binning*. NIPS 2009 workshop on temporal segmentation, Whistler, Canada.

[Endres et al, 2010a] Endres, D M, A Christensen, L Omlor, C Beck, J D Bouecke, H Neumann, and M A Giese. *Segmentation of action streams : comparison between human and statistically optimal performance.* Journal of Vision (2010): 1-2. doi:10.1007/s10827-009-0157-3.2.

[Endres et al, 2010b] Endres D.M., Höffken M., Vintila F., Bruce N., Bouecke J.D., Kornprobst P., Neumann H., Giese M.A.: *Hooligan detection: the effects of saliency and expert knowledge*. ECVP 2010 and Perception 39 supplement, page 193.

[Endres et al, 2011] Endres D.M, Christensen A., Omlor L., Giese M. A.. *Segmentation of action streams: human observers vs. Bayesian binning*. Submitted to ICML 2011.

[Gao and Vasconcelos] D. Gao and N. Vasconcelos. Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics. Neural Computation, 21:239–271, January 2009.

[Gibaldi et al. (2009)] A. Gibaldi, M. Chessa, A. Canessa, S.P. Sabatini and F. Solari. *Reading binocular energy population codes for short-latency disparity-vergence eye movements.* International Conference on Cognitive and Neural Systems, Boston 27-30 May 2009.

[Hansen & Neumann, 2008] Hansen, Thorsten, and Heiko Neumann. *A recurrent model of contour integration in primary visual cortex.* Journal of Vision 8 (2008): 1-25. doi:10.1167/8.8.8.Introduction.

[Itti and Baldi, 2009] L. Itti and P.F. Baldi. *Bayesian surprise attracts human attention*. Vision Research, 49(10):1295–1306, May 2009.

[Itti et al., 1998] L. Itti, C. Koch, and E. Niebur. *A model of saliency-based visual attention for rapid scene analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:1254–1259, 1998.

[Itti and Koch, 1999] L. Itti and C. Koch. *A comparison of feature combination strategies for saliency-based visual attention systems*. In SPIE Human Vision and Electronic Imaging IV , pages 473–482, 1999.

[Koch and Ullman] C. Koch and S. Ullman. Shifts in selective visual attention : Towards the underlying neuronal circuitry. Human Neurobiology , 4:219–227, 1985.

[Kornprobst and Tlapale, 2010] P. Kornprobst, E. Tlapale, J.D. Bouecke, H. Neumann, and G.S. Masson. A bio-inspired evaluation methodology for motion estimation. In VSS, 2010.

[Mallot (2000)] H. A. Mallot. *Computational Vision Information Processing in Perception and Visual Behaviour*, MIT Press, 2000.

[Nestares et al. (1998)] O. Nestares, R. Navarro, J. Portilla and A. Tabernero. *Efficient spatial-domain implementation of a multiscale image representation based on gabor functions.* Journal of Electronic Imaging, vol. 7, pages 166-173, 1998.

[Oberhoff, 2011] D. Oberhoff. *Hierarchical Bayesian Image Models*. In "Object Recognition", INTECH, accepted for publication in 2011.

[Oberhoff et al.., 2011] D. Oberhoff, D. Endres, M. Kolesnik and M. Giese. *Gates for Handling Occlusion in Hierarchical Bayesian Models of Images.* Submitted to ICML 2011.

[Ohnishi & Imiya, 2008] N. Ohnishi and A. Imiya: *Independent component analysis of optical flow for robot navigation*. Neurocomputing 71(10-12): 2140-2163, 2008.

[Omlor & Giese 2007] L. Omlor, M.A. Giese. *Learning of translation-invariant independent components: multivariate anechoic mixtures*. In: M.E. Davies, C.J. James, S.A. Abdallah, M.D. Plumbley (eds): Proceedings of Independent Component Analysis and Blind Signal Separation (ICA 2007), Lecture Notes in Computer Science (4666), Springer, Berlin, 762-769, 2007.

[Omlor & Giese 2011] L. Omlor, M.A. Giese. *Anechoic blind source separation using Wigner marginals.* Journal of Machine Learning Research, in press, 2011.

[Ohzawa et al. (1990)] I. Ohzawa, G.C. DeAngelis and R.D. Freeman. *Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors*. Science, vol. 249, pages 1037–1041, 1990.

[Raudies & Neumann, 2010] Raudies, Florian, and Heiko Neumann. *A model of neural mechanisms in monocular transparent motion perception*. Journal of physiology, Paris 104, no. 1-2 (2010): 71-83, 2010.

[Rankin et al., 2011] J. Rankin, E. Tlapale, R. Veltz, P. Kornprobst, and O. Faugeras. *Multistability and bifurcations in a model of motion perception*. In Developments in Dynamical Systems Arising from the Biosciences , March 2011.

[Ringbauer et al., 2011] Ringbauer, S., Tschechne, S., Neumann, H. *Mechanisms of Adaptive Spatial Integration in a Neural Model of Cortical Motion Processing*. In: Proceedings of the International Conference on Adaptive and Natural Computing Algorithms (ICANNGA) 2011, Ljubljana, 2011.

[Royden 88] Royden, Constance S, James F Baker, and John Allman. *Perceptions of depth elicited by occluded and shearing motions of random dots*. Perception 17 (1988): 289-296.

[Samarawikrama and Sabatini (2007)] J.G. Samarawickrama and S.P. Sabatini. *Version and vergence control of a stereo camera head by fitting the movement into the Hering's law*. Fourth Canadian Conference on Computer and Robot Vision, Montreal, Canada, 28-30 May 2007.

[Schölkopf et al., 1999] B. Schölkopf, C. Burges, A. Smola (eds). Advances in Kernel Methods - Support Vector Learning. MIT Press, Cambridge, MA, 1999.

[Semmlow et al. (1998)] J. L. Semmlow, W. Yuan, and T. L. Alvarez. *Evidence for Separate Control of Slow Version and Vergence Eye Movements: Support for Hering's Law*, J. Vision Research, Vol. 38, No. 8, pp. 1145-1152, 1998.

[Tlapale et al., 2010] Tlapale, É., Kornprobst, P., Bouecke, J., Neumann, H., Masson, G. *Evaluating motion estimation models from behavioural and psychophysical data.* In Bionetics, 1-14, 2010.

[Tlapale et al., 2010b] E Tlapale, P. Kornprobst, G.S. Masson, J.D. Bouecke, and H. Neumann. *Bio-inspired motion estimation – from modelling to evaluation, can biology be a source of inspiration*? Technical Report 7447, INRIA, November 2010.

[Tlapale et al., 2010c] E. Tlapale, G.S. Masson, and P. Kornprobst. *Modelling the dynamics of motion integration with a new luminancegated diffusion mechanism*. Vision Research, 50(17):1676–1692, August 2010.

[Tlapale et al., 2011] E. Tlapale, P. Kornprobst, G.S. Masson and O. Faugeras (2011). *A neural model for motion estimation*. In M. Bergounioux (ed.), Recent Advances in Mathematical Image Processing, Springer (to appear)

[Tschechne & Neumann, 2011a] Tschechne, S., Neumann, H. *Ordinal depth from occlusion using optical flow: a neural model*. In: Proceedings of Vision Science Society Meeting 2011 (VSS), Abstract, Naples, 2011.

[Tschechne & Neumann, 2011b] Tschechne, S., Neumann, H. *Ordinal depth from kinetic occlusions - a neural model*. In: Proceedings of 15th International Conference on Cognitive and Neual Systems (ICCNS), Abstract, Boston, 2011.

[Vo 2006] Vo, B.-N., and W.-K. Ma. *The Gaussian Mixture Probability Hypothesis Density Filter*. IEEE Transactions on Signal Processing 54, no. 11: 4091-4104. doi:10.1109/TSP.2006.881190. November 2006.

[Weidenbacher & Neumann, 2009] Weidenbacher, Ulrich, and Heiko Neumann. *Extraction of Surface-Related Features in a Recurrent Model of V1-V2 Interactions*. Plos ONE, 2009.

[Wilson and Cowan, 1972] H.R. Wilson and J.D. Cowan. *Excitatory and inhibitory interactions in localized populations of model neurons*. Biophys. J. , 12:1–24, 1972.

[Yonas 87] Yonas, a, L G Craton, and W B Thompson. *Relative motion: kinetic information for the order of depth at an edge.* Perception & psychophysics 41, no. 1 (January 1987): 53-9. http://www.ncbi.nlm.nih.gov/pubmed/3822744.

2 USE AND DISSEMINATION OF FOREGROUND

2.1 Section A: Dissemination of foreground

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[1] (if available)	Is/Will open access[2] provided to this publication?
1	*Design strategies for direct multiscale and multi- orientation visual processing in the log-polar domain*	*F. Solari, M. Chessa, and S.P. Sabatini*	*Design strategies for direct multiscale and multi- orientation visual processing in the log-polar domain*	*accepted with revision*	*Elsevier*		*-*	*-*	*-*	*no*
2	*Virtual Reality to Simulate Visual Tasks for Robotic Systems*	*M. Chessa, F. Solari and S.P. Sabatini*	*Virtual Reality*		*InTech*	*Janeza Trdine, Croatia*	*2011*	*83-104*	*Book edited by: Jae-Jin Kim, ISBN: 978-953-307-518-1*	*yes*
3	*A fast joint bioinspired algorithm for optic flow and two-dimensional disparity estimation*	*Chessa M., Sabatini S.P. and Solari F.*	*Lecture Notes in Computer Science*		*Springer*	*Verlag Berlin Heidelberg*	*2009*	*184-193*		*yes*
4	*Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism*	*Tlapale*	*Vision research*	*50(17)*	*Elsevier*		*2010*	*1676--1692*	*http://www.sciencedirect.com*	*yes*
5	*Neural Mechanisms of Motion*	*Bouecke*	*EURASIP Journal*	*2011*	*Hindawi*		*2011*	*pp. 22*		*yes*

[1] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).

[2] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

	Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems		*on Advances in Signal Processing*		*Publishing Corporation*				*http://www.hindawi.com/journals/asp/2011/781561/ref/*	
6	*A neural model for motion estimation*	*Tlapale*	*Recent Advances in Mathematical Image Processing*	*To appear*	*Springer*		*2011*			*no*
7	*Interactions of motion and form in visual cortex – A neural model*	*Beck C*	*Journal of physiology*	*29*	*Elsevier*	*Paris*	*2010*	*61-70*		*Yes*
8	*A model of neural mechanisms in monocular transparent motion perception*	*Raudies, F*	*Journal of physiology*	*104*	*Elsevier*	*Paris*	*2010*	*71-83*	*doi:10.1016/j.jphysparis.2009.11.010*	*Yes*
9	*Extraction of Surface-Related Features in a Recurrent Model of V1-V2 Interactions*	*Weidenbacher, U*	*Plos ONE*				*2009*			*Yes*
10	*A recurrent model of contour integration in primary visual cortex.*	*Hansen, T*	*Journal of Vision 8*	*8*			*2008*		*doi:10.1167/8.8.8.*	*Yes*
11	*Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems*	*Bouecke, Jan*	*EURASIP Journal on Advances in Signal Processing*				*2011*			*No*
12	*Evaluating motion estimation models from behavioural and psychophysical data.*	*Tlapale, É*	*Bionetics*				*2010*	*1-14*	*ISBN 978-963-9995-22-2*	*No*
13	*Neural Mechanisms of Motion Detection, Integration, and Segregation: From Biology to Artificial Image Processing Systems*	*Bouecke, Jan*	*EURASIP Journal on Advances in Signal Processing*				*2011*			*No*
14	*Evaluating motion estimation models from behavioural and psychophysical data.*	*Tlapale, É*	*Bionetics*				*2010*	*1-14*	*ISBN 978-963-9995-22-2*	*No*
15	*Segmentation of action streams : comparison between human and statistically optimal performance*	*Endres, D*	*Journal of Vision*	*10(7)*	*ARVO*		*2010*	*807*	*doi:10.1007/s10827-009-0157-3.2.*	*Yes*
16	*Hooligan detection: the effects of saliency and expert*	*Endres,D*	*Perception*	*39*	*Pion*	*London*	*2010*	*193*	*http://www.perceptionweb.com*	*Yes*

	knowledge									
17	*Anechoic blind source separation using Wigner marginals*	*Omlor, L*	*Journal of Machine Learning research*	*In press*	*JMLR*				*http://jmlr.csail.mit.edu/*	*Yes*
18	*Visual action control does not rely on strangers-Effects of pictorial cues under monocular and binocular vision.*	*Christensen, A*	*Neuropsychologia*	*49(3)*	*Elsevier*		*2011*	*556-563*	*doi:10.1016/j.neuropsychologia.2010.12.018*	*No*
19	*Spatiotemporal Tuning of the Facilitation of Biological Motion Perception by Concurrent Motor Execution*	*Christensen, A*	*Journal of Neuroscience*	*31(9)*	*Society for Neuroscience*		*2011*	*3493-3499*	*doi:10.1523/JNEUROSCI.4277-10.2011*	*Yes (?)*
20	*Biological motion detection does not involve an automatic perspective taking*	*Christensen, A*	*Journal of Vision*	*In press*	*ARVO*		*2011*			*Yes*
21	*Facilitation of biological-motion detection by motor execution does not depend on attributed body side*	*Christensen, A*	*Perception*	*39*	*Pion*	*London*	*2010*	*18*	*http://www.perceptionweb.com*	*Yes*
22	*View-based neural encoding of goal-directed actions: a physiologically-inspired neural theory*	*Giese, MA*	*Journal of Vision*	*10(7)*	*ARVO*		*2010*	*1095*	*doi: 10.1167/10.7.1095*	*Yes*
23	*It was (not) me: Causal Inference of Agency in goal-directed actions*	*Beck, TF*	*Nature precedings*	*In press*	*Nature*		*2011*			*Yes*
24	*Hierarchical Bayesian Image Models*	*Oberhoff, D*	*Object Recognition*	*In press*	*INTECH*		*2011*			*yes*

TEMPLATE A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities[3]	Main leader	Title	Date	Place	Type of audience[4]	Size of audience	Countries addressed
1	*Workshop*	*DIBE*	*European Conference on Computer Vision – Workshop on Vision for Cognitive Tasks*	*10 September 2010*	*Hersonissos, Heraklion, Crete, Greece*	*Scientific Community*	*50*	International
2	Conference	*DIBE*	International Conference on Computer Vision Theory and Applications	5-8 February 2009	Lisbon, Portugal	Scientific Community	100	International
3	Conference	*DIBE*	International Conference on Cognitive and Neural Systems	27-30 May 2009	Boston, MA, USA	Scientific Community	50	International
4	Conference	*DIBE*	International Conference on Computer Vision Systems	13-15 October 2009	Liege, Belgium	Scientific Community	150	International
5	Conference	*DIBE*	European Conference on Visual Perception	22-26 August 2010	Lausanne, Switzerland		500	International
6	Conference	*DIBE*	4th International Conference on Cognitive Systems	27-28 January 2010	Zurich, Switzerland	Scientific Community	300	International
7	*Conference*	*INRIA*	*BIMV, Bionetics*	*1-3 December, 2010*	*Boston, MA, USA*	*Scientific Community*	*50*	*International*
8	*Conference*	*INRIA*	*ICIP*	*7-10 November, 2009*	*Cairo, Egypt*	*Scientific Community*	*500*	*International*
9	*Conference*	*INRIA*	*CVPR*	*20-25 June, 2009*	*Miami, FL, USA*	*Scientific Community*	*500*	*International*
10	*Conference*	*INRIA and UULM*	*VSS*	*7-12 May, 2010*	*Naples, FL, USA*	*Scientific Community*	*1500*	*International*
11	*Conference*	*INRIA*	*Conference New Developments in Dynamical Systems Arising from the Biosciences*	*22-26 March, 2011*	*Columbus, OH, USA*	*Scientific Community*	*50*	*International*
12	*Thesis*	*INRIA*	*Modelling the dynamics of contextual motion integration in the primate, E. Tlapale PhD*	*25 January, 2011*	*Sophia Antipolis, France*	*Scientific Community*	*30*	*International*
13	*Web*	*INRIA*	*A Bio-Inspired Evaluation Methodology for*					

[3] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[4] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias ('multiple choices' is possible.

			Motion Estimation *http://www-sop.inria.fr/neuromathcomp/public/data/motionpsychobench/*					
14	*Conference*	*UniTue*	*23rd Annual Conference on Neural Information Processing Systems*	*7-12 December 2009*	*Vancouver and Whistler, Canada*	*Scientific Community*	*1000*	*International*
15	*Conference*	*UniTue*	*European Conference on Visual Perception*	*22-26 August 2010*	*Lausanne, Switzerland*	*Scientific Community*	*500*	*International*
16	*Conference*	*UniTue*	*Vision Sciences Society 11th Annual Meeting*	*May 6-11, 2011*	*Naples, FL, USA*	*Scientific Community*	*1300*	*International*
17	*Conference*	*UniTue*	*Ninth Göttingen meeting of the German Neuroscience Society*	*March 23-27, 2011*	*Göttingen, Germany*	*Scientific Community*	*500*	*International*
18	*Conference*	*UniTue*	*Computational and Systems Neuroscience 2011*	*February 24-27, 2011*	*Salt Lake City and Snowbird, UT, USA*	*Scientific Community*	*1000*	*International*
19	*webpage*	*UniTue*	*UniTue's SEARISE page*		*http://www.compsens.uni-tuebingen.de/index.php?page=project&id=20*	*Civil Society*		*International*
20	*Conference*	*Fraunhofer*	*Int. Conference On Artificial Neural Networks (ICANN)*	*September 2009*	*Cyprus, Greece*	*Scientific Community*	*1000*	*International*
21	*Conference*	*Fraunhofer*	*Advanced Concepts for Intelligent Vision Systems*	*Aug. 22-25 2011*	*Ghent, Belgium*	*Scientific Community*	*1000*	*International*
22	*Conference*	*Fraunhofer*	*The 28th International Conference on Machine Learning*	*June 28 through July 2, 2011*	*Washington, USA*	*Scientific Community*	*2000*	*International*
23	*Trade Show*	*Fraunhofer*	*SECURITY ESSEN*	*05.10.2010 - 08.10.2010*	*Essen, Germany*	*Industry*	*1100 Exhibitors 40.541 Visitors from 42 Countries*	*International*
24	*Press release*	*Fraunhofer*	*Fraunhofer Research News 09-2010 (see 4.1)*	*September 2010*	*Internet*	*General Public*	*Open end*	*International*
25	*Press release*	*Fraunhofer*	*Fraunhofer Pressemitteilung (see 4.2)*	*13.09.2010*	*Internet*	*General Public*	*Open end*	*Germany*
26	*Radio news*	*Fraunhofer*	*DRadio Wissen (see Section 4.3)*	*2.09.2010- 13:53*	*Radio*	*General Public*	*Open end*	*Germany*
27	*Article*	*Fraunhofer*	*Funkschau (see Section 4.4)*	*2.09.2010*	*Internet*	*General Public*	*Open end*	*Germany*

28	*Article*	*Fraunhofer*	*PC WELT (see Section 4.5)*	*2.09.2010*	*Magazine*	*General Public*	*Open end*	*Germany*
29	*Article*	*Fraunhofer*	*Photonics Spectra, Trends 2011: Imaging and Vision (see Section 4.7)*	*January 2011*	*Magazine*	*General Public*	*Open end*	*International*
30	*Article*	*Fraunhofer*	*Der Spiegel 39/2010 (see Section 4.8)*	*October 2010*	*Magazine*	*General Public*	*Open end*	*German*
31	*Article*	*Fraunhofer*	*German center for research and innovation - German Innovation of the month: January 2011 Vigilant Eyes camera* http://www.germaninnovation.org/research-and-innovation/centers-of-innovation-in-germany/german-innovations	*January 2011*	*Internet*	*General public*	*Open end*	*International*
32	*TV Report*	*Fraunhofer*	*ZDF:* http://www.zdf.de/ZDFmediathek/#/beitrag/video/1177458/Smart-eye-sorgt-f%C3%BCr-Sicherheit	*November 2011*	*TV, Internet*	*General public*	*Open end*	*In German*
33	*Article*	*Fraunhofer*	*Markt & Technik (See Section 4.9)*	*22.10.2010*	*Newspaper*	*General public*	*Open end*	*In German*
34	*Release*	*Fraunhofer*	http://www.eurekalert.org/pub_releases/2010-09/f-vce092010.php	*20.09.2010*	*Internet*	*General public*	*Open end*	*International*

The Coordinator Fraunhofer had received numerous inquiries about Smart Eyes system. In response to those Fraunhofer and TML had carried out communications, organised meetings and conducted discussions with the inquiring partners and companies.

2.2 Section B: Exploitation plans

2.2.1 Part B1

No applications for patents, trademarks or registered designs have been submitted.

2.2.2 Part B2

Type of Exploitable Foreground[5]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[6]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
Commercial exploitation of R&D results	*Motion pattern recognition in real-time*	*YES*		*Semi-automatic video surveillance systems*	*Other information technology and computer service activities*	*>2012*	*none planned*	*Beneficiary: UniTue. Possible licensing to Zeiss AG*
Commercial exploitation of R&D results	*Software framework for real-time saliency detection**	*YES*		*REAL-TIME AUTOMATIC VIDEO SURVEILLANCE SOFTWAR ECOMPATIBLE WITH CAMERAS OF DIFFERENT VENDORS*	*Other information technology and computer service activities*	*>2011*	*none planned*	*Beneficiary: Fraunhofrer and TML. Optional licensing to Viseum LtD, MODI GmbH.*
Commercial exploitation of R&D results	*SMART EYES **	YES		*REAL-TIME VIDEO SURVEILLANCE SYSTEM FOR SPORT ARENAS*	*Other information technology and computer service activities*	*>2011*	*NONE PLANNED*	*Beneficiary: Fraunhofer and TML.*

***** *Software framework for real-time saliency detection:*
Purpose: Automatic real-time analysis of video acquired by video surveillance systems. Real-time support of human operator during video surveillance, in particular video surveillance of crowded places and events with high number of people.
Exploited through commercialization of the software framework or its parts and integration with commercially video cameras.
Exploitable measures are ongoing by Fraunhofer with support of TML.
Further research is needed for further software development and improvement of software performance in particular application scenarios.
Potential impact has to be evaluated.

***** *Smart Eyes:*
Purpose: Automatic real-time detection, fixation and visualization of salient events in sport arenas during football games, assisting human video surveillance by security personal.

[19] A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.
[6] A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Exploited through commercialization of Smart Eyes.
Exploitable measures are ongoing by Fraunhofer and TML.
Further research is needed to improve/tune/test the system's performance and robustness against varying conditions in different Arenas / surveyed activities.
Potential impact is being evaluated.

3 REPORT ON SOCIETAL IMPLICATIONS

A General Information *(completed automatically when **Grant Agreement number** is entered.*

Grant Agreement Number: 215866	
Title of Project: SEARISE	
Name and Title of Coordinator: Marina Kolesnik, Dr.	

B Ethics

1. Did your project undergo an Ethics Review (and/or Screening)? • If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports? Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	***NO***
2. Please indicate whether your project involved any of the following issues (tick box) :	***YES***
RESEARCH ON HUMANS	
• Did the project involve children?	
• Did the project involve patients?	
• Did the project involve persons not able to give consent?	
• Did the project involve adult healthy volunteers?	
• Did the project involve Human genetic material?	
• Did the project involve Human biological samples?	
• Did the project involve Human data collection?	
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	
• Did the project involve Human Foetal Tissue / Cells?	
• Did the project involve Human Embryonic Stem Cells (hESCs)?	
• Did the project on human Embryonic Stem Cells involve cells in culture?	
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	
• Did the project involve tracking the location or observation of people?	YES
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	
• Were those animals transgenic small laboratory animals?	
• Were those animals transgenic farm animals?	
• Were those animals cloned farm animals?	
• Were those animals non-human primates?	
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	

DUAL USE	
• Research having direct military use	0 Yes 0 No
• Research having the potential for terrorist abuse	

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator	1	
Work package leaders		5
Experienced researchers (i.e. PhD holders)	2	7
PhD Students	1	4
Other		5

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	
Of which, indicate the number of men:	

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?	○ Yes X No	

6. Which of the following actions did you carry out and how effective were they?

	Not at all effective				Very effective
❑ Design and implement an equal opportunity policy	○	○	○	○	○
❑ Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑ Organise conferences and workshops on gender	○	○	○	○	○
❑ Actions to improve work-life balance	○	○	○	○	○

○ Other:

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

○ Yes- please specify

X No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

○ Yes- please specify

X No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

X Yes- please specify Project web site, system descriptions in press, publications

○ No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

○ Main discipline[7]: 2.2

○ Associated discipline[7]: ○ Associated discipline[7]:

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	○ Yes X No	

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

X No

○ Yes- in determining what research should be performed

○ Yes - in implementing the research

○ Yes, in communicating /disseminating / using the results of the project

[7] Insert number from list below (Frascati Manual).

11c In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?	○ ○	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

X	No
○	Yes- in framing the research agenda
○	Yes - in implementing the research agenda
○	Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

○	Yes – as a **primary** objective (please indicate areas below- multiple answers possible)
○	Yes – as a **secondary** objective (please indicate areas below - multiple answer possible)
X	No

13b If Yes, in which fields?

Agriculture		Energy		Human rights	
Audiovisual and Media		Enlargement		Information Society	
Budget		Enterprise		Institutional affairs	
Competition		Environment		Internal Market	
Consumers		External Relations		Justice, freedom and security	
Culture		External Trade		Public Health	
Customs		Fisheries and Maritime Affairs		Regional Policy	
Development Economic and Monetary Affairs		Food Safety		Research and Innovation	
		Foreign and Security Policy		Space	
Education, Training, Youth		Fraud		Taxation	
Employment and Social Affairs		Humanitarian aid		Transport	

13c If Yes, at which level?

- ○ Local / regional levels
- ○ National level
- ○ European level
- ○ International level

H Use and dissemination

14. How many Articles were published/accepted for publication in peer-reviewed journals?	**24**
To how many of these is open access[8] provided?	**24**
How many of these are published in open access journals?	
How many of these are published in open repositories?	
To how many of these is open access not provided?	
Please check all applicable reasons for not providing open access:	
❑ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal ❑ lack of time and resources ❑ lack of information on open access ❑ other[9]: ……	

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	
	Registered design	
	Other	
17. How many spin-off companies were created / are planned as a direct result of the project?		**1**
Indicate the approximate number of additional jobs in these companies:		

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

❑	Increase in employment, or	❑	In small & medium-sized enterprises
❑	Safeguard employment, or	❑	In large companies
❑	Decrease in employment,	❑	None of the above / not relevant to the project
X	Difficult to estimate / not possible to quantify		

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs:	*Indicate figure:*

[8] Open Access is defined as free of charge access for anyone via Internet.
[9] For instance: classification for security project.

Difficult to estimate / not possible to quantify	❑

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

X Yes ○ No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

○ Yes X No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

■	Press Release	■	Coverage in specialist press
■	Media briefing	■	Coverage in general (non-specialist) press
■	TV coverage / report	■	Coverage in national press
■	Radio coverage / report	■	Coverage in international press
■	Brochures /posters / flyers	■	Website for the general public / internet
❑	DVD /Film /Multimedia	❑	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

■	Language of the coordinator	■	English
❑	Other language(s)		

***Question F-10*:** Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES

1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]

1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)

1.3 Chemical sciences (chemistry, other allied subjects)

1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)

1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as

geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]

4 ANNEX I: SMART EYES IN THE PRESS

4.1 Fraunhofer Research News 09-2010

Fraunhofer

Research News
09-2010 I Topic 6

Vigilant camera eye

»Goal, goal, goal!« fans in the stadium are absolutely ecstatic, the uproar is enormous. So it's hardly surprising that the security personnel fail to spot a brawl going on between a few spectators. Separating jubilant fans from scuffling hooligans is virtually impossible in such a situation. Special surveillance cameras that immediately spot anything untoward and identify anything out of the ordinary could provide a solution. Researchers from the Fraunhofer Institute for Applied Information Technology FIT in Sankt Augustin have now developed such a device as part of the EU project »SEARISE – Smart Eyes: Attending and Recognizing Instances of Salient Events«. The automatic camera system is designed to replicate human-like capabilities in identifying and processing moving images.

Like the human eye, it can, for instance, distinguish objects when observing a scene, even if the objects are moving in front of a very turbulent background. The Smart Eyes system analyzes the recorded data in real time and immediately points out salient features. »That is invaluable for video surveillance of public buildings or places«, says Dr. Martina Kolesnik, research scientist at the FIT. »In certain circumstances the capabilities of a human observer are limited. Ask someone to keep an eye on a certain stand in a football stadium and they are bound to miss many details. That same person can only carefully monitor certain sections of the whole area and will quickly get tired. That's where Smart Eyes clearly comes into its own.«

The system hardware consists of a fixed surveillance camera which covers a certain area, and two ultra-active stereo cameras. Like human eyes, these can fix on and follow various points very quickly in succession – but also zoom in on details. At the heart of Smart Eyes is innovative software that automatically analyzes the image sequences. It replicates key strategies of the human eye and brain. Taking its lead from the flow of visual images in the brain, the software has a hierarchical, modular structure. It initially ascertains the degree of movement for each pixel, thus identifying the particular active areas in the scene. From this it learns motion patterns and stores them as typical models. On the basis of these models the system then identifies events and classifies them: for instance the software can distinguish between passive spectators and fans jumping up and down. Image patterns such as empty seats or steps are also identified. The application picks out salient events and focuses on these using the active stereo cameras. Depending on the priorities set by the security experts, various events are designated as salient. The program can, where necessary, filter out objects such as flags being waved to focus specifically on other salient

events, for instance a person on the edge of the pitch. »Our image analysis software is compatible with camera systems produced by all vendors. It can be installed easily. The user doesn't have to make any adjustments«, says the researcher. The Smart Eyes system will be on show at Security Essen 2010 from October 5-8, 2010.

The Smart Eyes camera records a stand during a soccer match. The software focuses on salient events such as a person on the edge of the pitch. (© Fraunhofer FIT)

Picture in color and printing quality: www.fraunhofer.de/press

Fraunhofer Institute for Applied Information Technology FIT
Schloss Birlinghoven | 53754 Sankt Augustin, Germany | www.fit.fraunhofer.de
Contact: Dr. Marina Kolesnik | Phone +49 2241 14-3421 | marina.kolesnik@fit.fraunhofer.de
Press: Alex Deeg | Phone +49 2241 14-2208 | alex.deeg@fit.fraunhofer.de

4.2 Fraunhofer Pressemitteilung 13.09.2010

Pressemitteilung vom 01.09.2010 | 12:50
Pressefach: Fraunhofer Gesellschaft

Wachsames Kamera-Auge

Ein neuartiges Kamerasystem könnte künftig auf öffentlichen Plätzen und in Gebäuden für mehr Sicherheit sorgen: Smart Eyes funktioniert ähnlich wie das menschliche Auge. Das System analysiert die aufgenommen Daten in Echtzeit und weist sofort auf Besonderheiten und ungewöhnliche Szenen hin.
»Tor, Tor, Tor!« Der Jubel der Fans im Fußballstadion kennt keine Grenzen, der Aufruhr ist groß. Da wundert es nicht, dass das Handgemenge zwischen einigen Zuschauern vom Sicherheitspersonal unbemerkt bleibt. Jubelnde von Streitenden zu unterscheiden, ist in so einer Situation kaum möglich. Abhilfe schaffen könnten spezielle Überwachungskameras, die Auffälliges sofort entdecken und ungewöhnliche Vorkommnisse identifi zieren. Ein solches Gerät haben Forscher des Fraunhofer-Instituts für Angewandte Informationstechnik FIT in Sankt Augustin jetzt im EU-Projekt »SEARISE - Smart Eyes: Attending and Recognizing Instances of Salient Events« entwickelt. Das automatische Kamerasystem soll beim Erkennen und Verarbeiten von bewegten Bildern menschenähnliche Leistungen erreichen.
Wie das menschliche Auge kann es beispielsweise beim Betrachten einer Szene Objekte unterscheiden, auch wenn sich diese vor einem sehr unruhigen Hintergrund bewegen. Das Smart Eyes System analysiert die Videodaten in Echtzeit und weist sofort auf Besonderheiten hin. »Zur Video-Sicherheitsüberwachung von öffentlichen Gebäuden oder Plätzen ist das von unschätzbarem Wert«, sagt Dr. Martina Kolesnik, Wissenschaftlerin am FIT. »In einigen Situationen sind die Fähigkeiten eines menschlichen Beobachters begrenzt. Soll er eine Fankurve in einem Fußballstadion überwachen, entgehen ihm viele Einzelheiten. Er kann nur bestimmte Areale der Gesamtfl äche sehr aufmerksam betrachten und er ermüdet schnell. Hier sind die Smart Eyes klar im Vorteil.«
Die Hardware des Systems besteht aus einer fest installierten Übersichtskamera, die ein bestimmtes Gebiet abdeckt, und zwei ultra-aktiven Stereokameras. Diese können wie die Augen des Menschen sehr schnell nacheinander verschiedene Punkte fi xieren und verfolgen - aber darüber hinaus auch auf Details zoomen. Kern der Smart Eyes ist eine neuartige Software, die Bildsequenzen automatisch auswertet. Sie bildet wesentliche Strategien des menschlichen Seh- und Verarbeitungsapparats nach. Ähnlich den Sehströmen des Gehirns ist die Software hierarchisch und modular aufgebaut. Sie ermittelt zuerst für jeden Bildpunkt den Bewegungsgrad und identifi ziert so die besonders aktiven Areale in der Szene. Daraus werden Bewegungsmuster erlernt und als typische Modelle abgespeichert. Anhand der Modelle erkennt das System dann Ereignisse und ordnet sie ein: Beispielsweise unterscheidet die Software passive Zuschauer von aufspringenden Fans. Auch Bildmuster wie unbesetzte Stühle oder Treppen werden identifi ziert. Die Anwendung wählt markante Ereignisse aus und fokussiert diese mit den aktiven Stereo-Kameras. Je nach den von den Sicherheitsexper- ten gesetzten Prioritäten werden verschiedene Ereignisse als markant gekennzeichnet. So fi ltert das Programm auf Wunsch etwa geschwenkte Fahnen heraus, um gezielt andere Auffälligkeiten zu fokussieren, zum Beispiel eine Person am Spielfeldrand. »Unsere Bildauswertungssoftware ist zu den Kamera-Systemen aller Hersteller kompatibel. Sie lässt sich einfach installieren. Der Anwender muss keinerlei Anpassungen vornehmen«, sagt Kolesnik. Das Smart Eyes System ist auf der Messe Security Essen vom 5. bis 8. Oktober 2010 zu sehen.
Um hochauflösende Bilder herunterzuladen, klicken Sie bitte auf den Link und dann auf das Vorschaubild.
Alle Beiträge über http://www.fraunhofer.de/presse

Impressum:
Herausgeber und Redaktionsanschrift:
Fraunhofer-Gesellschaft
Franz Miller
Presse und Öffentlichkeitsarbeit
Hansastraße 27c
80686 München
Telefon: 089 1205-1333
Fax: 089 1205-7515
presse@zv.fraunhofer.de

Eingetragener Verein
Registergericht
Amtsgericht München
Register-Nr. VR 4461

4.3 DRadio Wissen 2.09.2010

DRadio Wissen

MESZ 15:29 Uhr

- A
- A
- A

- Startseite
  - Agenda
  - Natur
  - Medien
  - Globus
  - Kultur
  - Meine Zukunft
  - Spielraum
  - Blog

Wissen

Donnerstag, 2. September 2010 13:53 Uhr

Überwachungssystem analysiert Szenerien auf das Wesentliche hin

von13:53 Uhr

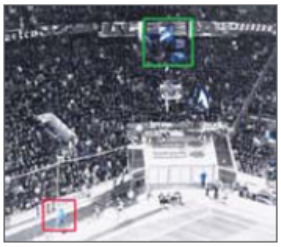
Es kann in einer Szenerie besondere Ereignisse erkennen - das neue Überwachungssystem des Fraunhofer-Instituts für Angewandte Informationstechnik in Sankt Augustin. Wie das Portal "golem.de" berichtet, ist das System "Smart Eyes" nach menschlichem Vorbild gebaut. Es kann einzelne Objekte ausmachen, selbst wenn sie sich vor einem unruhigen Hintergrund bewegen. Eine feste Kamera liefert die Übersicht über eine Szenerie und zwei bewegliche Kameras können schnell einen bestimmten Punkt anvisieren. So unterscheiden die "Smart Eyes" in einem Fußballstadion zum Beispiel sitzende von aufspringenden Zuschauern. Sie können aber auch Dinge wie unbesetzte Stühle und Treppen einordnen. Störende Elemente wie geschwenkte Fahnen filtert die Software heraus, um den Blick auf Wesentliches nicht zu verstellen.

4.4 Funkschau 2.09.2010

URL: http://www.funkschau.de/infrastruktur/news/article/wachsames_kamera-auge/34981/1d7ffb46-b670-11df-b4ba-001ec9efd5b0

02.Sep.2010

Wachsames Kamera-Auge



Frauenhofer Institut

Die Smart Eyes Kamera nimmt eine Fankurve während eines Fußballspiels auf. Markante Ereignisse wie eine Person am Spielfeldrand fokussiert die Software.

Ein neuartiges Kamerasystem könnte künftig auf öffentlichen Plätzen und in Gebäuden für mehr Sicherheit sorgen: Smart Eyes funktioniert ähnlich wie das menschliche Auge. Das System analysiert die Videodaten in Echtzeit und weist sofort auf Besonderheiten und ungewöhnliche Szenen hin. Das Smart Eyes System ist auf der Messe Security Essen vom 5. bis 8. Oktober 2010 zu sehen.

"Tor, Tor, Tor!" Der Jubel der Fans im Fußballstadion kennt keine Grenzen, der Aufruhr ist groß. Da wundert es nicht, dass das Handgemenge zwischen einigen Zuschauern vom Sicherheitspersonal unbemerkt bleibt. Jubelnde von Streitenden zu unterscheiden, ist in so einer Situation kaum möglich.

Abhilfe schaffen könnten spezielle Überwachungskameras, die Auffälliges sofort entdecken und ungewöhnliche Vorkommnisse identifizieren. Ein solches Gerät haben Forscher des Fraunhofer-Instituts für Angewandte Informationstechnik FIT in Sankt Augustin jetzt im EU-Projekt "SEARISE - Smart Eyes: Attending and Recognizing Instances of Salient Events" entwickelt. Das automatische Kamerasystem soll beim Erkennen und Verarbeiten von bewegten Bildern menschenähnliche Leistungen erreichen.

Wie das menschliche Auge kann es beispielsweise beim Betrachten einer Szene Objekte unterscheiden, auch wenn sich diese vor einem sehr unruhigen Hintergrund bewegen. Das Smart Eyes System analysiert die Videodaten in Echtzeit und weist sofort auf Besonderheiten hin.

"Zur Video-Sicherheitsüberwachung von öffentlichen Gebäuden oder Plätzen ist das von unschätzbarem Wert", sagt Dr. Martina Kolesnik, Wissenschaftlerin am FIT. "In einigen Situationen sind die Fähigkeiten eines menschlichen Beobachters begrenzt. Soll er eine Fankurve in einem Fußballstadion überwachen, entgehen ihm viele Einzelheiten. Er kann nur bestimmte Areale der Gesamtfläche sehr aufmerksam betrachten und er ermüdet schnell. Hier sind die Smart Eyes klar im Vorteil."

Die Hardware des Systems besteht aus einer fest installierten Übersichtskamera, die ein bestimmtes Gebiet abdeckt, und zwei ultra-aktiven Stereokameras. Diese können wie die Augen des Menschen sehr schnell nacheinander verschiedene Punkte fixieren und verfolgen - aber darüber hinaus auch auf Details zoomen. Kern der Smart Eyes ist eine neuartige Software, die Bildsequenzen automatisch auswertet. Sie bildet wesentliche Strategien des menschlichen Seh- und Verarbeitungsapparats nach. Ähnlich den Sehströmen des Gehirns ist die Software hierarchisch und modular aufgebaut. Sie ermittelt zuerst für jeden Bildpunkt den Bewegungsgrad und identifiziert so die besonders aktiven Areale in der Szene. Daraus werden Bewegungsmuster erlernt und als typische Modelle abgespeichert. Anhand der Modelle erkennt das System dann Ereignisse und ordnet sie ein: Beispielsweise unterscheidet die Software passive Zuschauer von aufspringenden Fans. Auch Bildmuster wie unbesetzte Stühle oder Treppen werden identifiziert.

Die Anwendung wählt markante Ereignisse aus und fokussiert diese mit den aktiven Stereo-Kameras. Je nach den von den Sicherheitsexperten gesetzten Prioritäten werden verschiedene Ereignisse als markant gekennzeichnet. So filtert das Programm auf Wunsch etwa geschwenkte Fahnen heraus, um gezielt andere Auffälligkeiten zu fokussieren, zum Beispiel eine Person am Spielfeldrand.

4.5 PC WELT 2.09.2010

Quelle	www.pcwelt.de vom 02.09.2010
Seite	0
Web-Link	http://www.pcwelt.de/start/sicherheit/sonstiges/news/2348236
Copyright	Copyright ©2006 IDG Business Verlag GmbH. All rights reserved. Alle Rechte vorbehalten

Wachsames Kamera-Auge gegen Randalierer

Ein neuartiges Kamerasystem könnte künftig auf öffentlichen Plätzen und in Gebäuden für mehr Sicherheit sorgen: Smart Eyes funktioniert ähnlich wie das menschliche Auge. Das System analysiert die Videodaten in Echtzeit und weist sofort auf Besonderheiten und ungewöhnliche Szenen hin.

"Tor, Tor, Tor!" Der Jubel der Fans im Fußballstadion kennt keine Grenzen, der Aufruhr ist groß. Da wundert es nicht, dass das Handgemenge zwischen einigen Zuschauern vom Sicherheitspersonal unbemerkt bleibt. Jubelnde von Streitenden zu unterscheiden, ist in so einer Situation kaum möglich. Abhilfe schaffen könnten spezielle Überwachungskameras, die Auffälliges sofort entdecken und ungewöhnliche Vorkommnisse identifizieren.

Ein solches Gerät haben Forscher des Fraunhofer-Instituts für Angewandte Informationstechnik FIT in Sankt Augustin jetzt im EU-Projekt "SEARISE – Smart Eyes: Attending and Recognizing Instances of Salient Events" entwickelt. Das automatische Kamerasystem soll beim Erkennen und Verarbeiten von bewegten Bildern menschenähnliche Leistungen erreichen.

Wie das menschliche Auge soll es beispielsweise beim Betrachten einer Szene Objekte unterscheiden, auch wenn sich diese vor einem sehr unruhigen Hintergrund bewegen. Das Smart Eyes System analysiert die Videodaten in Echtzeit und weist sofort auf Besonderheiten hin. "Zur Video-Sicherheitsüberwachung von öffentlichen Gebäuden oder Plätzen ist das von unschätzbarem Wert", sagt Dr. Martina Kolesnik, Wissenschaftlerin am FIT. "In einigen Situationen sind die Fähigkeiten eines menschlichen Beobachters begrenzt. Soll er eine Fankurve in einem Fußballstadion überwachen, entgehen ihm viele Einzelheiten. Er kann nur bestimmte Areale der Gesamtfläche sehr aufmerksam betrachten und er ermüdet schnell. Hier sind die Smart Eyes klar im Vorteil."

Die Hardware des Systems besteht aus einer fest installierten Übersichtskamera, die ein bestimmtes Gebiet abdeckt, und zwei ultra-aktiven Stereokameras. Diese können wie die Augen des Menschen sehr schnell nacheinander verschiedene Punkte fixieren und verfolgen – aber darüber hinaus auch auf Details zoomen. Kern der Smart Eyes ist eine neuartige Software, die Bildsequenzen automatisch auswertet. Sie bildet wesentliche Strategien des menschlichen Seh- und Verarbeitungsapparats nach. Ähnlich den Sehströmen des Gehirns ist die Software hierarchisch und modular aufgebaut. Sie ermittelt zuerst für jeden Bildpunkt den Bewegungsgrad und identifiziert so die besonders aktiven Areale in der Szene. Daraus werden Bewegungsmuster erlernt und als typische Modelle abgespeichert. Anhand der Modelle erkennt das System dann Ereignisse und ordnet sie ein: Beispielsweise unterscheidet die Software passive Zuschauer von aufspringenden Fans. Auch Bildmuster wie unbesetzte Stühle oder Treppen werden identifiziert.

Die Anwendung wählt markante Ereignisse aus und fokussiert diese mit den aktiven Stereo-Kameras. Je nach den von den Sicherheitsexperten gesetzten Prioritäten werden verschiedene Ereignisse als markant gekennzeichnet. So filtert das Programm auf Wunsch etwa geschwenkte Fahnen heraus, um gezielt andere Auffälligkeiten zu fokussieren, zum Beispiel eine Person am Spielfeldrand.

"Unsere Bildauswertungssoftware ist zu den Kamera-Systemen aller Hersteller kompatibel. Sie lässt sich einfach installieren. Der Anwender muss keinerlei Anpassungen vornehmen", sagt Kolesnik. Das Smart Eyes System ist auf der Messe Security Essen vom 5. bis 8. Oktober 2010 zu sehen. Halle 7, Stand 913.

Abbildung Smart Eyes

4.6 Bild der Wissenschaft 16.11.2010

bild der wissenschaft / 16.11.2010

Kameratechnik

Scharfer Blick ins Stadion

Ein neuartiges Kamerasystem, das ähnlich funktioniert wie ein menschliches Auge, soll für mehr Sicherheit bei Großveranstaltungen sorgen - zum Beispiel bei Fußballspielen. Das Überwachungssystem namens 'smart Eye", das Forscher am Fraunhofer-Institut für Angewandte Informationstechnik FIT in Sankt Augustin entwickelt haben, arbeitet mit einer fest installierten Übersichtskamera sowie zwei schnell beweglichen Stereokameras. Diese können ihren Blick - wie ein menschliches Auge - rasch nacheinander auf unterschiedliche Punkte heften, diese verfolgen und heranzoomen. Das Herzstück des Systems ist eine spezielle Software, die anhand bestimmter Muster, die das Programm erlernt, Menschen und unterschiedliche Objekte erkennt - selbst vor einem unruhigen Hintergrund. Bei der Analyse von Videobildern geht sie ähnlich vor wie das menschliche Gehirn beim Sehen.

Wird das intelligente Kamerasystem in einer Fußballarena auf die Zuschauertribüne gerichtet, kann es dort zuverlässig und rasch jubelnde Fans von randalierenden Zuschauern unterscheiden und auch Personen erkennen, die eine Barriere überklettern, um auf das Spielfeld zu gelangen. Damit die Bildanalyse einfacher wird, lassen sich bestimmte Objekte wie Fahnen oder Transparente automatisch ausblenden. Hat Smart Eye eine auffällige Szene erkannt, meldet es sie der Polizei.

4.7 Photonics Spectra January 2011

Trends 2011: Imaging and Vision

A Vision of the Future

BY MARIE FREEBODY
CONTRIBUTING EDITOR

Vision systems can be as important to manufacturing businesses as the manufactured products themselves. For products ranging from semiconductors to solar cells, and from pharmaceuticals to vehicles, vision systems help ensure quality and maximize production line efficiency, which means cost savings.

Global sales revenue from machine vision systems in North America was valued at $889.3 million in 2009, according to the latest market study by the Automated Imaging Association (AIA).

The semiconductor industry tops the list of users both in terms of revenue and units sold. This is followed by the automotive and wood industries. While vision systems have enjoyed a strong presence in these manufacturing areas for many years, they are now starting to catch the attention of nonmanufacturing enterprises.

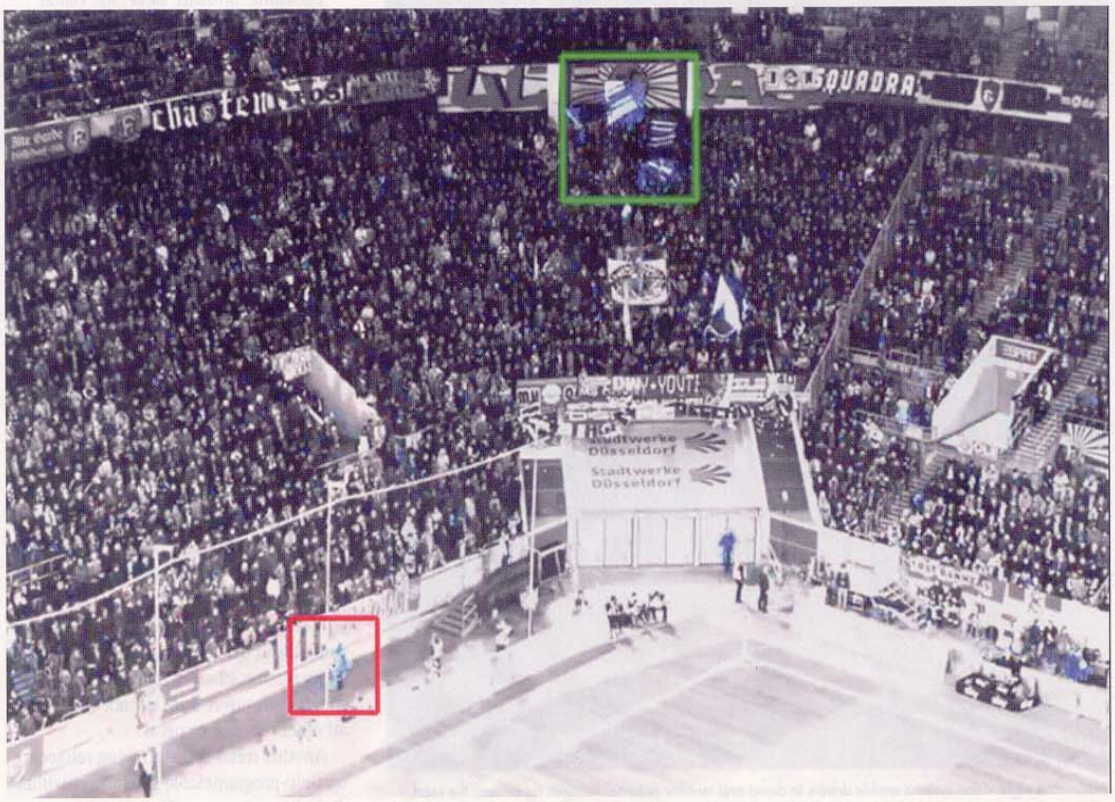
"Because of the attractiveness of the machine vision value proposition – increased efficiency, improved product quality and lower costs – machine vision technology will find success in a number of nontraditional and emerging markets," said Paul Kellett, AIA director of market analysis. "As time goes on, I fully expect machine vision companies to increasingly target nonindustrial customers – that is, to focus beyond the factory in their sales campaigns."

From economic lows to tech highs

The recent economic downturn hit the manufacturing industry hard and, as a result, the machine vision market suffered too. In fact, revenue declined by 29.4 percent from 2004 to 2009, according to AIA's study.

Even though the greater economy may be experiencing a recovery, most economists view the comeback as tentative and weak. What's more, many feel that a double-dip recession cannot be ruled out.

Despite this, Kellett is cautiously optimistic about machine vision. "While there is some evidence that the 'Great Recession' interrupted to some extent long-term trends (as customers bought down in



The Smart Eyes system picks out two possible salient events during a soccer match. Waving flags (green box) are deemed not to be security relevant, whereas the person on the edge of the pitch (red box) is determined to warrant more attention. Courtesy of the SEARISE partners.

Advanced Illumination's new expandable spot/linear array features high-brightness LEDs and an extruded aluminum housing with built-in mounting features and an efficient heat transfer design. Images courtesy of Advanced Illumination.

eroding the distinction between the camera and a relatively new product: the smart camera.

Smart cameras combine the functionality of a machine vision system with lower prices and have, not surprisingly, experienced explosive growth. Typically, a smart camera houses a camera, an image processing unit and machine vision-related programs within the same casing.

When smart cameras were invented, their capabilities were somewhat limited. Their lack of tools and processing power meant that they were generally suitable only for a narrow range of simple tasks. Typically, these early smart cameras were used for detecting the presence or absence of items but were not particularly "smart" by today's definition.

Currently, smart cameras are the fastest growing sector of the machine vision market. And such products are opening the door to some of the emerging nonmanufacturing applications of vision systems, such as biometric recognition and intelligent surveillance.

Smarter surveillance

Video surveillance has moved on from the indiscriminate capture of images to more intelligent traffic surveillance systems, for example. Smart surveillance systems can monitor the flow of traffic, alert authorities to possible accidents and identify number plates.

Perhaps the next step in intelligent surveillance is a system with a humanlike ability to learn from the environment it oversees. In the SEARISE (Smart Eyes: Attending and Recognising Instances of Salient Events) project, which comes to fruition in February 2011, a trinocular active cognitive vision system is being developed.

Unlike other approaches in video surveillance, the Smart Eyes system will be able to learn continuously from visual input. It will self-adjust to changes in the visual environment as well as concentrate its focus on salient events. Saliency is a measure of relative novelty, so a salient event is one that stands out from its surroundings.

Since not all salient events will be security relevant, the project members have developed software to help the system "learn" to focus only on security-relevant events. SEARISE software can learn from a few video samples of security-relevant events specified by experts.

The software then analyzes all salient events in the scene to detect the specified security events; it can update its knowledge about the learned events via automatic online learning. Additional interactive learning allows new security-relevant events to be incorporated into the previously learned model. These events will then be categorized according to context, and the smart camera will duly assign the majority of computational resources to the informative parts of the scene.

The aim of the project is to develop a prototype of Smart Eyes and put it to the test in real-life scenarios featuring the activity of people at varying distances from the camera. At a long-range distance capacity, the system will monitor crowd behavior of sports fans in a soccer arena. At short range, the system will be tasked with monitoring the behavior of small groups of people and single individuals.

The technology behind Smart Eyes consists of three cameras acting in unison in a coordinated "recognition loop." A wide-view global camera performs general monitoring and, once a particular event of interest is identified, active binocular stereo cameras zoom in for a closer look.

The SEARISE Project has been supported by the European Union's FP7 Programme and is made up of a team of academic and industrial partners across Europe. The smart vision cameras can be used for security applications in public places such as sports stadiums, city streets, metro systems and airports.

Driving vision into autos

Still another trend is lower cost, as subcomponents become less expensive. As a consequence, vision systems are offering increasing value and appeal in commercial markets. For example, vehicle manufacturers have discovered the benefits of thermal imaging to enhance driver vision.

Of course, for many years the military and police have used thermal imaging for night vision for security and surveillance applications. In fact, in places where constant guarding is necessary, such as airports, ports and nuclear facilities, thermal imaging will often be employed.

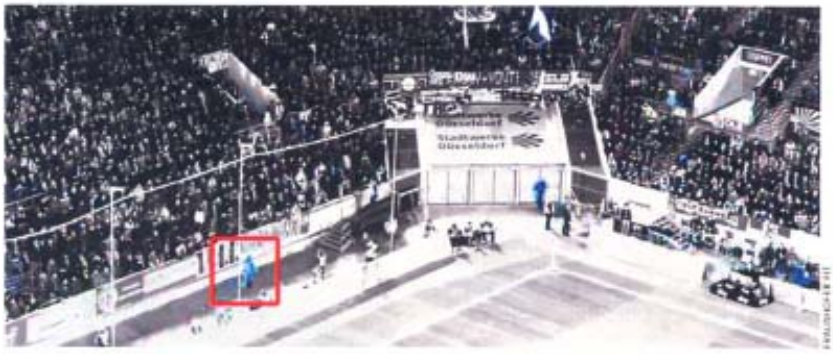
Thermal imaging specialist Flir Systems recognizes that, thanks to trends toward more compact systems, better image qual-

4.8 Der Spiegel January 39/2010 (October)

SICHERHEITSTECHNIK

Gefahr bunt markiert

Wo in einem Fußballstadion wird der Sicherheitsdienst gerade gebraucht? Auf den Bildern normaler Überwachungskameras ist das oft nur schwer auszumachen. Schlägereien oder Menschen, die plötzlich versuchen, aufs Spielfeld zu rennen, gehen im allgemeinen Durcheinander von Fahnen, Fans und La-Ola-Wellen leicht unter. Deshalb hat jetzt das Fraunhofer-Institut für Angewandte Informationstechnik in Sankt Augustin bei Bonn eine Software entwickelt, die auf den Überwachungsbildern auffällige Ereignisse gezielt erkennen und markieren kann. Dieses „Smart Eyes" getaufte System, das mit einer festen und zwei beweglichen Kameras arbeitet, analysiert zunächst die typischen – also unproblematischen – Bewegungsmuster der jeweiligen Szenerie, etwa Fahnenschwenken oder jubelnde Fans. Was sich davon abhebt, zum Beispiel eine Prügelei oder ein Einzelner, der aus der Menge ausbricht, markiert das System auf den Security-Monitoren dann farblich – und zwar in Echtzeit. Sofort richten sich zudem die beiden beweglichen, ultraschnellen Kameras auf das verdächtige Muster und liefern eine Aufnahme in besonders hoher Auflösung.



„Smart Eyes"-Monitorbild mit rot markiertem Mann am Spielfeldrand

4.9 Markt & Technik 22.10.2010

Quelle	Markt und Technik vom 22.10.2010
Seite	19
Nummer	43
Rubrik	im Fokus Bildverarbeitung

Markt&Technik
Die unabhängige Wochenzeitung für Elektronik

"Smart Eyes" können in Fußballstadien Jubelnde von Streitenden unterscheiden

Kamerasystem eifert dem menschlichen Auge nach

Ein neuartiges, "Smart Eyes" genanntes Kamerasystem könnte künftig auf öffentlichen Plätzen und in Gebäuden für mehr Sicherheit sorgen: Es funktioniert ähnlich wie das menschliche Auge. Das System analysiert die aufgenommenen Daten in Echtzeit und weist sofort auf Besonderheiten und ungewöhnliche Szenen hin.

In voll besetzten Fußballstadien bleiben kleinere Ereignisse, die schnell eskalieren können, vom Sicherheitspersonal oft unbemerkt. Jubelnde von Streitenden zu unterscheiden, ist in solchen Situationen kaum möglich. Abhilfe schaffen könnten spezielle Überwachungskameras, die ungewöhnliche Vorkommnisse identifizieren. Ein solches Gerät haben Forscher des Fraunhofer-Instituts für Angewandte Informationstechnik FIT in Sankt Augustin jetzt im EU-Projekt "SEARISE - Smart Eyes: Attending and Recognizing Instances of Salient Events" entwickelt.

Wie das menschliche Auge kann das "Smart-Eyes"-System z.B. beim Betrachten einer Szene Objekte unterscheiden, auch wenn sie sich vor einem unruhigen Hintergrund bewegen. "Das automatische Kamerasystem analysiert die Videodaten in Echtzeit", verdeutlicht Dr. Martina Kolesnik, Wissenschaftlerin am FIT. "Zur Video-Sicherheitsüberwachung öffentlicher Gebäude oder Plätze ist das von hohem Wert. In einigen Situationen sind die Fähigkeiten eines menschlichen Beobachters begrenzt: Soll er eine Fankurve in einem Fußballstadion überwachen, entgehen ihm viele Einzelheiten." Er könne nur bestimmte Areale der Gesamtfläche aufmerksam betrachten und ermüde schnell. Die Hardware des Systems besteht aus einer fest installierten Übersichtskamera, die ein bestimmtes Gebiet abdeckt, und zwei aktiven Stereokameras. Diese können wie die Augen des Menschen schnell nacheinander verschiedene Punkte fixieren und verfolgen - aber darüber hinaus auch auf Details zoomen. Kern der "Smart Eyes" ist eine neuartige Software, die Bildsequenzen automatisch auswertet. Sie bildet bestimmte Strategien des menschlichen Seh- und Verarbeitungsapparats nach.

Ähnlich den Sehströmen des Gehirns ist die Software hierarchisch und modular aufgebaut. Zuerst ermittelt sie für jeden Bildpunkt den Bewegungsgrad und identifiziert so die besonders aktiven Areale in der Szene. Daraus werden Bewegungsmuster erlernt und als typische Modelle abgespeichert. Anhand der Modelle erkennt das System dann Ereignisse und ordnet sie ein: Beispielsweise unterscheidet die Software passive Zuschauer von aufspringenden Fans. Auch Bildmuster wie unbesetzte Stühle oder Treppen werden identifiziert. Die Software wählt markante Ereignisse aus und fokussiert sie mit den aktiven Stereo-Kameras. Je nach von den Sicherheitsexperten gesetzten Prioritäten werden verschiedene Ereignisse als markant gekennzeichnet. So filtert das Programm auf Wunsch etwa geschwenkte Fahnen heraus, um gezielt andere Auffälligkeiten zu fokussieren, etwa Personen am Spielfeldrand.

"Unsere Bildauswertungs-Software ist zu den Kamera-Systemen aller Hersteller kompatibel", erläutert Kolesnik. "Sie lässt sich leicht installieren, und der Anwender muss keinerlei Anpassungen vornehmen." (ak)