

Increasing Sensor Resource Efficiency using a Proactive Sensor System

Stephan Matzka^{1,2} Yvan R. Petillot¹ Andrew M. Wallace¹ Paul Sprickmann Kerkerinck³

¹ Heriot-Watt University, School of Engineering and Physical Sciences, Edinburgh, UK

² Ingolstadt University of Applied Sciences, Institute for Applied Research, Ingolstadt, Germany

³ Audi Electronics Venture GmbH, Ingolstadt, Germany

Abstract

This paper proposes a system to direct high-resolution sensor resources by cues extracted from low-resolution data. The proposed method is highly reactive using unsupervised saliency cues, resource efficient due to the trained classifiers, and adequate to the present context.

1 Introduction

In recent years, the use of cameras and range sensors in cars has increased. Low-resolution video cameras are used for lane detection, whereas radars and laser scanner are used to detect traffic participants and obstacles. Currently, these systems are often used as stand-alone devices, e.g. a low-resolution video camera for lane detection will broadcast the detected lane trajectory, while the acquired video image is not used for any other purposes.

We propose to use low-resolution data acquired by different sensors to direct high-resolution sensor resources in an efficient manner. Our proposed method is highly reactive due to its unsupervised real-time saliency detection. Adequacy is ensured using trained classifiers and assigning contextually adequate utility functions.

2 System Overview

The proposed system processes data over various stages, beginning at sensor level and increasing both in level of abstraction and significance towards a contextual reasoning level (cf. Fig. 1). From these, object classification is computationally expensive. Moreover,

the quality of the subsequent reasoning highly depends on the quality and robustness of the data level features attributes.

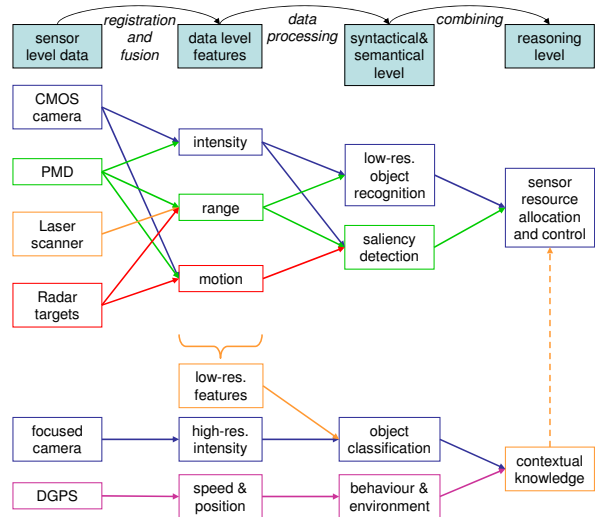


Figure 1: System overview showing the used sensors and the respective layers of sensor data abstraction.

While the transformation from sensor level information towards a contextual awareness is highly desirable in order to maximise sensor resource efficiency, it requires a total of three serial processing steps. It is obvious that an active sensor system in a rapidly changing environment such as road traffic has to exhibit a high degree of reactivity. To ensure this criterion is satisfied, an unsupervised novelty detection algorithm is performed on the low level cues in parallel.

2.1 Syntactical and Semantical Level

2.1.1 Object Recognition and Classification

Recognition and classification of objects is performed using a set of trained classifier on the high-resolution video image. We use a boosted cascade of Haar-like features to detect objects in the environment (cf. Fig. 2). This concept was proposed in [10] to detect faces in images, and has been applied to a large number of object recognition problems since. The method is computationally effective, as it discards most background regions in the first stages of the trained cascade, which allows to spend more time on regions promising to contain the desired object category. The cascade is built by subsequently adding simple Haar-like features to a stage in the cascade until it rejects a certain percentage (50% is a common value) of the background regions remaining after the previous stages. At the same time, each stage in the cascade is constrained to reject no, or only a very small amount (i.e. 0.3%) of positives.

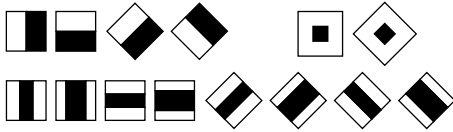


Figure 2: Haar-like features used in the trained cascades are edge features (top left), centre-surround features (top right), and line features (bottom) [10].

Haar-like features, once trained, can naturally be rescaled which is also exploited in [10] by using a representation called integral images. By splitting up the region covered by a set of features into subregions that can be reassembled to represent all used features, the integral value of every subregion only has to be determined once, saving computation time.

We trained Haar-like feature classifier cascades using OpenCV¹. Classification using the resulting cascades requires 9.12 ms per 10k pixels per class on a 2 GHz Pentium 4 processor, which is rendered it computationally expensive to run on a high-resolution image even if only for a single class.

2.1.2 Saliency Detection

In literature, saliency is often derived from the fixation patterns of the human eye which, during its pre-attentive phase, treats regions as salient, which 'pop

¹OpenCV: <http://www.sourceforge.net/projects/opencvlibrary/>

out' [9] of their surroundings (i.e. [3]). This definition follows the idea of local comparisons, evaluating the contrast between a region and its surrounding regions.

A different definition treats regions as salient, whose feature space representation is rare - at best unique - in their environment (i.e. [11]). The latter definition assumes statistical knowledge about the entire environment and determines saliency in a global context.

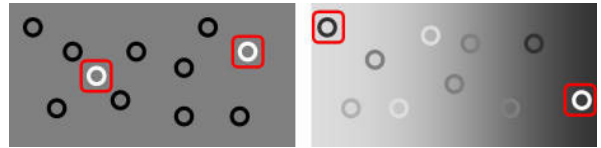


Figure 3: Saliency can emerge from both global rarity (left), and local contrast (right).

Both global and local definitions describe apparent forms of saliency, our goal is to find an algorithm that can detect both. The method presented in Itti, 2000 uses a local centre-surround approach [3], yet it also includes the notion of global rarity by dividing each feature's saliency map by its number of peaks before combining them into a single saliency map. In Walker *et al*, 1998, the Mahalanobis distance d between a local feature vector x and the environment's mean feature vector \bar{x} is computed. S represents the covariance matrix of all feature vectors x .

$$d(x, \bar{x}) = \sqrt{(x - \bar{x})^T S^{-1} (x - \bar{x})} \quad (1)$$

This distance in feature space is a good measure for a regions uniqueness and thereby global saliency.

An evaluation of both saliency detectors showed, that the centre-surround approach [3] requires 1.65ms per 10k pixels, whereas the Mahalanobis distance is computationally more expensive as it requires to calculate the inverse covariance matrix of all feature vectors (4.45ms per 10k pixels). The latter can be substantially sped up by randomly selecting a statistically significant subset of feature vectors (i.e. 2000 feature vectors, resulting in 1.96 ms per 10k pixels).

We propose a combined feature vector x_C for our saliency detection using (1), consisting of intensity I and its local derivatives, range z and translational 3-D motion which is determined using a fast motion estimation algorithm described in [5].

$$x_C = \left(x, y, z, \frac{\delta x}{\delta t}, \frac{\delta y}{\delta t}, \frac{\delta z}{\delta t}, I, \frac{\delta I}{\delta x}, \frac{\delta I}{\delta y}, \frac{\delta^2 I}{\delta x \delta y} \right)^T \quad (2)$$

3 Reasoning Level

3.1 Sensor Resource Allocation & Control

A decision making process is invoked to combine a highly reactive cues from the unsupervised novelty detection and contextual knowledge such as the safety aspect and the observability of objects and regions in the environment. In current literature, this combination is often achieved by simple means, such as multiplication [6] or summation [2] of low-level and high-level cues, or the centre of weight algorithm proposed in [4], the first two employing simple yet ineffective strategies such as winner-take-all.

An utility driven concept to satisfy two vision tasks concurrently is presented in [3], where a winner selection society is established in order to maximise different global efficiency concepts. Apart from an utility driven approach, it is interesting to investigate in a concept of context-governed bottom-up gaze concept, ensuring reactivity and contextual efficiency at the same time.

We propose a cue combination concept derived from utility theory also used in multi-agent resource allocation (cf. [1]). First, Koene *et al.*, 2007 argues that a winner-take-all approach is problematic as it disregards all attractors of lesser value. This aspect is also substantiated in [1, 7], with the latter proposing to choose the gaze direction with minimum overall saliency loss for all cues. Second, multiplicative combination is good as it is independent of scale, yet tends to annihilates a region if only one cue does not assign any saliency to it. Third, using absolute values as in [2, 4, 7] can be problematic as it requires normalisation of all cues to a common saliency metric.

All discussed gaze direction selection methods have in common, that the gaze direction is supposed to be centred upon a single object. However an image is a shareable resource (cf. [1]) as it can include more than one object and we have a fixed aspect ratio for active cameras (typically 4:3) which typically does not coincide with the aspect ratio of the object (typically 1:1 for cars, and 3:4 for lorries). Our utility optimisation scheme is able to determine the size of a region that conforms as much as possible to this optimum resolution and contains maximum relative cue values.

Evaluation showed, that centring upon an object of interest increases the chances to confirm a known traffic participants in high-resolution, whereas our method shows a higher increase in newly detected objects in high-resolution.

3.2 Contextual Knowledge

Contextual knowledge is won by combining data level features using a set of rules and constraints. In our case, the foremost context is our safety as well as the safety of other traffic participants.

The safety aspect in road traffic scenes is two-fold, as any traffic participant can be a threat or can be vulnerable to another traffic participant (including the observing traffic participant), or both. Whereas it would be possible to determine all n-n relations of all participants, we limit our scope towards a 1-n relation of our own vehicle towards other traffic participants.

For the case that our own vehicle is a car, a Failure Mode and Effects Analysis (FMEA, [8]) focused on the severity aspect is conducted. In our case, this turns out as:

- A pedestrian is *very vulnerable* and *not harmful* to a car.
- A bicycle is *very vulnerable* and *not harmful* to a car.
- A car is *vulnerable* and *harmful* to a car.
- A lorry is *not vulnerable* and *very harmful* to a car.

The protection requirement of the above participants can be defined in various manners, yet we tend towards prioritising the recognition of pedestrians due to their high vulnerability. Apart from the class of a traffic participant, its relative motion towards our own car influences the safety aspect. A traffic participant moving away from our own car is much less dangerous or vulnerable than a traffic participant moving towards it. This information is provided by the motion vectors and trajectories at data level. Besides motion, distance in itself is relevant for the safety aspect, since spatial proximity is a condition for both being dangerous or vulnerable.

Our own car's speed and global position informs us about the car's actions (such as turning, driving at 30m/s, or backing a car into a parking spot) as well as our current environment (such as urban, cross-country, or highway). This information provides us with information about the likeliness of presence of a certain traffic participant category (i.e. pedestrians are common in cities, and very rare on highways) as well as the chance to detect this traffic participant in time. The latter becomes a problem for small traffic participants like pedestrians or bicycles in fast moving environments such as highways, where it might not feasible to detect and classify a pedestrian in time.

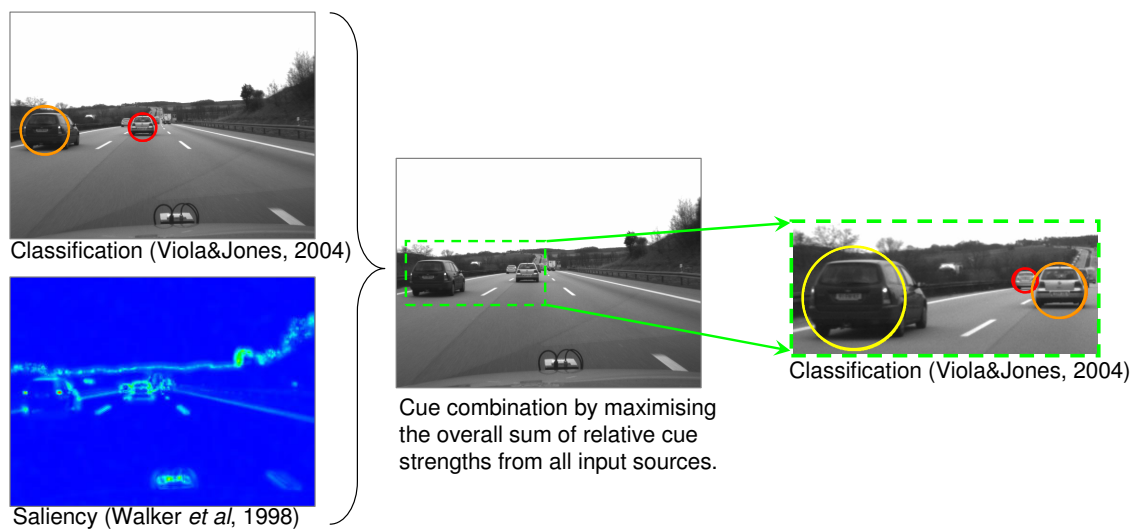


Figure 4: Cue combination method maximising overall utility. The selected region is analysed in high-resolution, confirming two previous classifications and detecting an additional car.

4 Conclusion and Future Work

We present a utility driven system to allocate and control high-resolution sensor resources based upon low-resolution cues.

The proposed system is implemented and evaluated on real-life traffic sequences and shows a substantial increase in detection of cars from 38 to 76 by 100%, of which 29 are confirmed by the high-resolution classification. For lorries, this increase is only 20% from 10 to 12 detected lorries, three of which are confirmed in high-resolution.

Evaluation of the performance on pedestrian recognition is future work as well as a detailed description of our resource allocation scheme and the extension of the latter towards a scheduling behaviour over time.

References

- [1] Y. Chevaleryre, P.E. Dunne, U. Endriss, J. Lang, M. Lematre, N. Maudet, J. Padget, S. Phelps, J.A. Rodriguez-Aguilar, and P. Sousa. Issues in multiagent resource allocation. *Informatica*, (30):3–31, 2006.
- [2] Simone Frintrop. *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search*. LNAI 3899. Springer, 2006.
- [3] Laurent Itti. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, Computation and Neural Systems, 2000.
- [4] Ansgar Koene, Jan Morn, Vlad Trifa, and Gordon Cheng. Gaze shift reflex in a humanoid active vision system. In *Proceedings of the ICVS Workshop*, 2007.
- [5] Stephan Matzka, Yvan R. Petillot, and Andrew M. Wallace. Fast motion estimation on range image sequences acquired with a 3-d camera. In *Proceedings of the British Machine Vision Conference*, volume II, pages 750–759. BMVA Press, 2007.
- [6] Vidhya Navalpakkam and Laurent Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 02, pages 2049–2056, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
- [7] Javier F. Seara. *Intelligent gaze control for vision-guided humanoid walking*. PhD thesis, Technische Universität München, 2004.
- [8] D.H. Stamatis. *Failure mode and effect analysis*. Quality Press, 2nd edition, 2003.
- [9] Anne Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156–177, 1985.
- [10] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [11] K. N. Walker, T. F. Cootes, and C. J. Taylor. Locating salient object features. In *British Machine Vision Conference (BMVC)*, pages 557–566. BMVA Press, 1998.

Institut für Angewandte Forschung
IAF

HERIOT WATT UNIVERSITY
HERIOT WATT UNIVERSITY

Audi
Audi

Increasing Sensor Resource Efficiency using a Proactive Sensor System

Stephan Matzka^{1,2}
 Dr. Yvan R. Petillot¹
 Prof. Dr. Andrew. M. Wallace¹
 Paul Sprickmann Kerkerinck³

¹ Heriot-Watt University, Edinburgh
² Ingolstadt University of Applied Sciences
³ Audi Electronics Venture GmbH, Ingolstadt

Institut für Angewandte Forschung
IAF

HERIOT WATT UNIVERSITY
HERIOT WATT UNIVERSITY

Audi
Audi

Increasing Sensor Resource Efficiency

Content

- 1 Motivation
- 2 Proactive Sensor System
- 3 Demonstration
- 4 Outlook

Stephan Matzka April 8, 2008 Slide 2

Institut für Angewandte Forschung
IAF

HERIOT WATT UNIVERSITY
HERIOT WATT UNIVERSITY

Audi
Audi

Increasing Sensor Resource Efficiency

Content

- 1 Motivation
- 2 Proactive Sensor System
- 3 Demonstration
- 4 Outlook

Stephan Matzka April 8, 2008 Slide 4


Institut für Angewandte Forschung
IAF

HERIOT WATT UNIVERSITY
HERIOT WATT UNIVERSITY


Audi
Audi

Proactive Sensor System

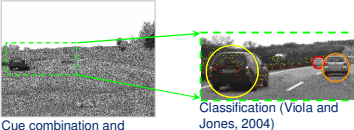
Through the Looking-Glass, and What Alice Found There




Classification (Viola and Jones, 2004)



Saliency (Walker et al, 1998)



Cue combination and maximisation.



Classification (Viola and Jones, 2004)

Stephan Matzka April 8, 2008 Slide 5

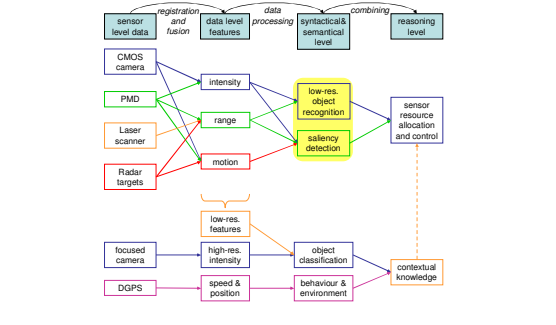
Institut für Angewandte Forschung
IAF

HERIOT WATT UNIVERSITY
HERIOT WATT UNIVERSITY

Audi
Audi

Proactive Sensor System

Schematic Overview



Stephan Matzka April 8, 2008 Slide 6

Proactive Sensor System
Unsupervised Saliency Detection

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Source Image

Saliency using a local approach (Itti, 2000)

Saliency using a global approach (Walker *et al.*, 1998)

Stephan Malska April 8, 2008 Slide 7

Proactive Sensor System
Trained Classifier

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

We use the trained classifier proposed in Viola and Jones, 2004.

Positive samples

Negative samples

Classifier cascade using simple Haar-like features.

True Positive Rate

False Positive Rate

— Lorry Cascade 14(24x32)

— Car Cascade 23(32x32)

Stephan Malska April 8, 2008 Slide 8

Proactive Sensor System
Classification Results

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

We trained three cascades for cars, lorries, and pedestrians.

Car Cascade

Lorry Cascade

Pedestrian Cascade

Stephan Malska April 8, 2008 Slide 9

Proactive Sensor System
Computational Costs I

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Computation time for image processing (in ms per 10k pixel)

- Saliency using Walker *et al.*, 1998: 1.96 ms
- Object classification using Viola and Jones, 2004: 9.12 ms

Saliency can be computed on low-resolution image (i.e. 64x48 pixel).

- Computation time for saliency: 0.60 ms

Object classification must be computed on high-resolution image (i.e. 320x240 pixel). Also, object classification must also be computed separately for every object class.

- Computation time for 3 object classes: 210.12 ms

We are still missing bicycles, motorcycles, etc.

Stephan Malska April 8, 2008 Slide 10

Proactive Sensor System
How does nature handle this?

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Attentive scrutiny <math>< 10^4</math> bits/sec

Attentional bottleneck

Optic Nerve ~3 x 10⁶ bits/sec

Retinal processing

Retinal Images > 10¹⁰ bits/sec

Optics

3-dimensional world (unlimited information)

AN INFORMATION PYRAMID

From: Itti, Rees, Tsotsos (Eds.): Encyclopedia of Visual Attention, Chapter 3, Elsevier, Oxford, 2005.

The human visual system handles this bottleneck by selecting regions of interest, which are focused and scrutinised for known objects (or categories).

The information is reduced to

$$\frac{10^4}{3 \cdot 10^{10}} = 3.3 \cdot 10^{-3} = 0.3\%$$

The required reduction is not nearly as much in our case.

Stephan Malska April 8, 2008 Slide 11

Proactive Sensor System
Schematic Overview (Reprise)

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

sensor level data registration and fusion data level processing combining reasoning level

CMOS camera intensity low-res. object recognition sensor resource allocation and control

PMD range low-res. features

Laser scanner motion saliency detection

Radar targets

focused camera high-res. intensity object classification contextual knowledge

DGPS speed & position behaviour & environment

Stephan Malska April 8, 2008 Slide 12

Proactive Sensor System
Computational Costs II

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Computation time for image processing (in ms per 10k pixel)

- Saliency using Walker *et al.*, 1998 1.96 ms
- Object classification using Viola and Jones, 2004 9.12 ms

Saliency can be computed on low-resolution image (i.e. 64x48 pixel).

- Computation time for saliency 0.60 ms

Object recognition on low-resolution image (i.e. 64x48 pixel) for 5 classes.

- Computation time for 5 object classes 14.01 ms

Object classification on focused high-resolution image (i.e. 128x96 pixel) for 5 classes

- Computation time for 5 object classes 56.03 ms

We now include bicycles and motorcycles in the classification process.

Total computation time is 70.64ms over 14 fps.

Stephan Matzka April 8, 2008 Slide 13

Proactive Sensor System
Schematic Overview (Reprise)

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Stephan Matzka April 8, 2008 Slide 14

Proactive Sensor System
Contextual Knowledge

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

We have prior knowledge based upon:

- Type of traffic environment (motorway, urban road, ...)
- Position and category of classified traffic participants (pedestrian, bicycle, car, lorry, ...)
- Behaviour of our own car and surrounding traffic participants (fast, slow, turning, ...)

An ontology-based framework will determine adequate cue combination weights and tuned recognition cascades.

Stephan Matzka April 8, 2008 Slide 15

Increasing Sensor Resource Efficiency
Content

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

- Motivation
- Proactive Sensor System
- Demonstration
- Outlook

Stephan Matzka April 8, 2008 Slide 17

Outlook
Summary and Future Work

Institut für Angewandte Forschung IAF HERIOT WATT UNIVERSITY Audi

Summary

- We propose a system to efficiently control high-resolution sensors using cues (unsupervised saliency and trained classifiers) acquired from low-resolution sensor data.
- Reducing the data from 640 x 480 pixel to 128 x 96 pixel (4%) is not as much as the human eye (0.3%) but allows to process data faster than 10 frames per second.
- Results show an increase in the number of correctly classified traffic participants (up to 100% for cars) as compared to a fixed-gaze system.

Future Work

- Extensive evaluation and publication of the proposed sensor resource allocation scheme.
- Further investigation of contextual influence on classification performance.

Stephan Matzka April 8, 2008 Slide 18

Thank you for your attention.

Increasing Sensor Resource Efficiency
using a Proactive Sensor System

Stephan Matzka, Yvan R. Petillot,
Andrew M. Wallace, Paul Sprickmann Kerkerinck

Ingolstadt University of Applied Science
Institute for Applied Research (IAF)

Stephan Matzka
Eiplanstraße 10
85049 Ingolstadt
Telefon: 0841-9348-612
E-Mail: stephan.matzka@ifti-ingolstadt.de

Stephan Matzka April 8, 2008 Slide 19