Computing visionaries

Professor Tony Lindeberg explains how the development of algorithms can improve the visual understanding of computers, enabling them to perform tasks that are currently beyond their capabilities

To begin, can you discuss your fundamental research goals?

A main goal of my research is to understand vision. Vision makes it possible for us to gather information about objects at a distance, based on the light that is reflected from them. Yet an image measurement at a single point does not, in general, reveal any useful information. Instead, the information is mediated by relative relations between image measurements at neighbouring points in the visual field. I seek to understand the properties of light patterns that make such extraction possible.

An object may appear in different ways on the retina depending on the viewing distance and direction, and the illumination conditions. Nevertheless, we perceive it as the same object. We can furthermore easily distinguish between a very large number of different objects. How we combine measurements of relevant image properties in an efficient manner to understand the image contents under the huge variations that occur in natural scenes is also of interest to me. In addition, I am involved in deriving computational theories that describe the possibilities of extracting properties of the world from visual information, as well as formulating efficient algorithms that can be implemented on a computer and in modelling biological vision.

Can you discuss the broad context from which this project has sprung?

The project has originated from our earlier work on theories of visual operations regarding scale-space theory and the application of scale-space theory for computing image features for computer vision. Previously, we have developed methods for object recognition and for recognising spatio-temporal actions based on receptive field-like image operations. The current project continues, extends and generalises this direction of research. Recent work has included performing an extensive evaluation of a large number of spatial image descriptors in terms of histograms of receptive field responses of different types and at different scales. The results achieved from this may provide a general indication of the exciting possibility of defining more discriminatory image descriptors. Object recognition and action recognition are also very active research areas internationally.

What challenges do you hope to overcome during the course of your research?

Within this project we are aiming at developing image descriptors that can handle natural imaging variations in a theoretically well-founded manner and thereby be robust to the natural variabilities that occur in real-world image data. We are also aiming at image descriptors that can be more discriminatory than their predecessors, and thereby allow for more accurate decisions at later processing stages. We hope to increase the understanding of measurement-based descriptors in terms of receptive field-like image operations, which can also lead to better algorithms.

In what ways do the kernels generated from your theories have interesting relations to biological vision?

Based on a set of structural assumptions of an idealised vision system, we have performed a mathematical characterisation of what kernels are compatible with basic symmetry properties of the environment. We have derived a set of kernels regarding both spatial (time-independent) image data and spatio-temporal image data (video). Interestingly, the scale-space kernels obtained from this theory are in strong agreement with receptive fields measured in biological vision. Indeed, from our theory we can generate receptive field profiles with qualitatively very similar properties to the linear receptive fields that have been found in the lateral geniculate nucleus (LGN) and the primary visual cortex (area V1) for cats or monkeys.

What implications might this have?

The similarity between our scale-space kernels and biological vision supports our belief that this approach should be a viable path for expressing robust computer vision operations. Specifically, it supports the use of receptive field-based image measurements for object recognition and spatio-temporal recognition. The similarity between the scale-space kernels and biologically receptive fields also indicates that biological vision seems to be close to ideal to what follows from our idealised mathematical theory. Thereby, there are very close connections between mathematical theory, biological receptive fields and the methodology we follow for image-based recognition. Our mathematical theory could also provide a valuable tool for modelling biological vision.

Visual **recognition**

While the notion of computers that can identify objects might sound somewhat futuristic, a research group is investigating the complex issues surrounding the apprehension of visual information, focusing in particular on the scale-space theory for multi-scale data handling

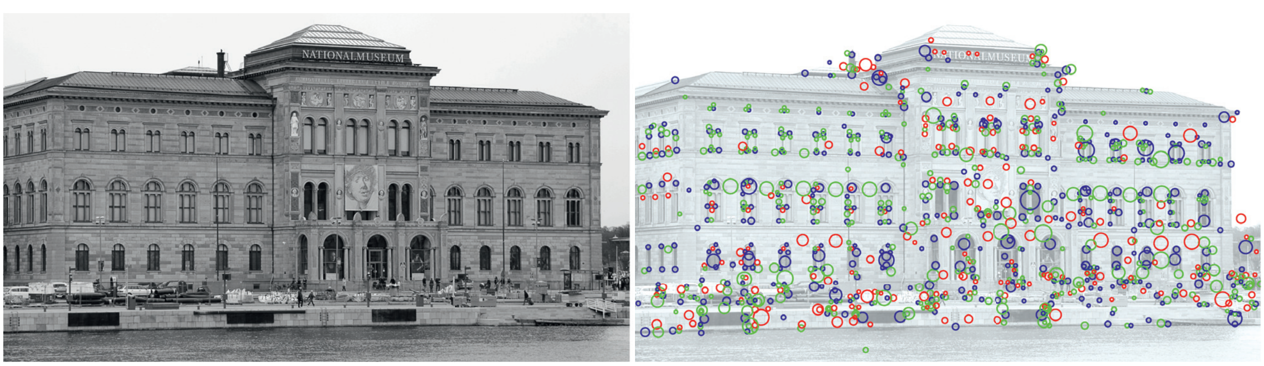
THE DIFFICULT TASK of allowing computers to deal with visual information has been the focus of a number of scientists in recent years, and breakthroughs are beginning to move the technology towards applications. With object recognition, three-dimensional reconstruction and action recognition of all viable targets, a team at the KTH Royal Institute of Technology in Stockholm is aiming at a number of uses for this technology. These include the possibility of visually guided inspection and surveillance, tasks which will be advantageous to automate in situations both dangerous and where large amounts of data must be processed for long periods of time. The possibility of safety systems for driving is another area where the technology will be able to play an important role. Automated searches of image databases will also contribute to tasks from environmental monitoring to the automated matching of medical image data. In each of these cases, complex images have to be precisely analysed, and the difficulties connected with making software that will allow computers to do this are numerous.

One of the major issues the team is trying to overcome is the development of appropriate image descriptors for objects of varying scale and structure. What makes it particularly difficult is the fact that image data is significantly affected by variations such as viewing direction, illumination conditions and, for time-varying image data, the relative motions between object and observer. With these shifting inputs, the parameters of the image descriptors, which must be sufficiently robust to deal with these natural variations in image condition, are difficult to define. Beyond this, there are problems of scale. This relates to the way in which any object may be composed of different substructures at different scales, as well as the fact that the appearance of objects is affected by the distance between the object and the observer. What is required is automatic ways of handling these variations in images. The changes caused by adjusting light conditions alone can comprise shifts in colour, affect which parts of an object are visible and which sections obscured by shadow, and impact the way in which an image is focused. What the team aims to do is produce different approaches for computing invariant image descriptors with respect to these natural image transformations, which can be placed alongside a method for combining image information over different scales. The challenge lies in translating the complex and continually variable world of images into stable, computational descriptors which can then be used.

HANDLING DIFFERENT SCALES IN IMAGE DATA

The intersection between the idealised mathematical world, defined by points and lines, and the actual images which need to be analysed, drives a great deal of the work which the team has been completing. Beyond the problems of direction, illumination and motion, multi-scale images are a notably difficult issue to deal with. Where mathematical objects can be described without issues of scale, the images that have to be analysed can depict objects at a large range of scales, and the distance of observation is extremely important to this scheme.

Furthermore, a single image can contain objects which require representation on a range of different scales. For example, a tree is a relevant concept on a scale of metres, whereas a leaf is a more natural concept on a scale of centimetres, and a forest at the scale of kilometres. Variations in scale can be handled by successive smoothing of the image data to different scales, which allows the suppression of fine-scale structures when the large scale is supposed to be the focus, thereby separating image structures that occur at different scales. The handling of such multi-scale data has been dealt with by scale-space theory. Scale-space theory was originally developed to handle such data in a sound theoretical manner, but has since evolved into a general theory of early visual operations, pertinent to other types of image variations. It has been successfully applied to a large number of the variations which are being dealt with in the work conducted by the team.



Interest points detected from an image of a building, with their scales adapted to the local image structures. At such interest points local image descriptors are computed and then used for matching and recognition.

INTELLIGENCE

IMAGE DESCRIPTORS AND SCALE-SPACE THEORY FOR SPATIAL AND SPATIO-TEMPORAL RECOGNITION

OBJECTIVES

- To develop robust, discriminatory and efficient image descriptors for object recognition, object categorisation and action recognition
- To develop theories and algorithms for robust and efficient computation of the underlying image features needed for the aforementioned tasks. Particular emphasis will be on extending current scale-space theory to the needs from recognition tasks

PARTNERS

Oskar Linde; **Lars Bretzner**, KTH, Sweden

KEY COLLABORATORS

Professor Per Roland, Karolinska Institute, Sweden

Ivan Laptev, INRIA, France

CONTACT

Professor Tony Lindeberg
Research leader

KTH Royal Institute of Technology
School of Computer Science and Communication
SE-100 44 Stockholm
Sweden

T +46 879 062 05
E tony@csc.kth.se

www.csc.kth.se/~tony/lab

TONY LINDEBERG is a Professor in Computer Science at KTH (Royal Institute of Technology) in Stockholm, Sweden. He received his MSc degree in 1987, his PhD in 1991, became Docent in 1996, and was appointed Professor in 2000. He was a Research Fellow at the Royal Swedish Academy of Sciences between 2000 and 2010. Lindeberg's research interests in computer vision relate to scale-space representation, image features, object recognition, spatio-temporal recognition, focus-of-attention and shape. He is author of *Scale-Space Theory in Computer Vision*.

KTH Computer Science and Communication

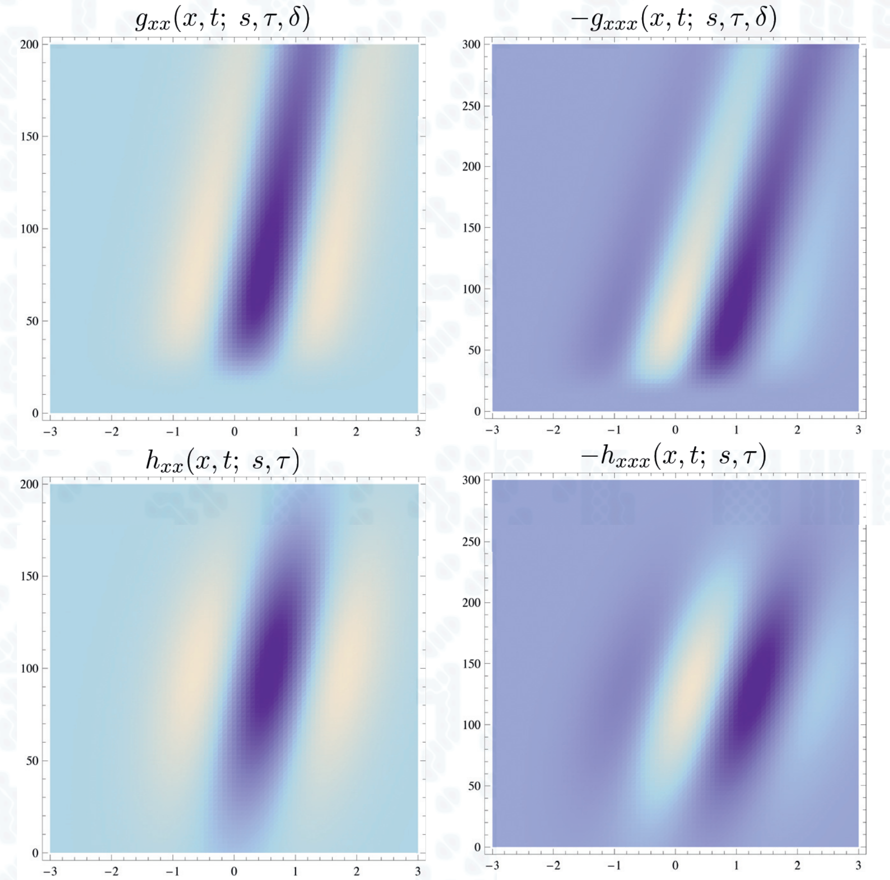


FIGURE 1. Receptive fields over space x (horisontal dimension) and time t (vertical dimension) corresponding to second- and third-order spatial derivatives, whose extent over space and time is adapted to a non-zero image velocity.

PROJECT CONTINUUM

Work completed by other researchers has demonstrated that image recognition can be performed by summarising receptive field response in terms of histograms. Working from these previous breakthroughs, the team in Stockholm, led by Professor Tony Lindeberg, has built its own projects extending this methodology: "In previous work, we developed a set of spatio-temporal image descriptors in terms of local position-dependent histograms of spatio-temporal gradients or optic flow," Lindeberg notes. What this means is that the temporal and positionally local image structures are able to be characterised around different points in space-time. This work on image-based recognition has allowed the computation of image features at different scales and of different orders, which can be combined to produce the invariants that are so essential to the work. This has allowed the team to experimentally demonstrate that these histograms can contribute significantly to image understanding. In particular, the recognition of actions and events in space and time can be likened to the recognition of objects by local spatial features, making a parallel in the processes which the software algorithms are capable of completing. The hope is that such developments will continue to drive the research towards the applications towards which the group is aiming.

CONTRIBUTION TO THE FIELD

Continuing on from their predecessors, as well as building on projects they have already completed, the team at KTH has been able to make significant contributions to the field. Notably, their previous work on automatic scale selection has enabled scale-invariant object recognition, which has now become a standard tool in the field. Furthermore, their work on affine shape adaptation and affine invariant fixed points in scale space has led in part to the general adoption of affine invariant image features. By taking these innovations and broadening them into new areas, the work has been able to continue the innovation which characterised previous breakthroughs. The experimental work conducted on histogram descriptors has led to significantly better performance than was previously possible with those tools, and is set to lead to more discriminatory descriptors for images. The research has also led to the potential for qualitative image features being used for recognition, a step forward in the production of software that can robustly apprehend images.

With Lindeberg's guidance, the team has been able to produce these excellent results, and has a number of other projects in the pipeline. Having spent 10 years as a Research Fellow at the Royal Swedish Academy of Sciences, Lindeberg is excellently placed to orchestrate the investigations into image features, object recognition and spatio-temporal apprehension. It is hoped that the team will continue to flourish under his leadership, broadening knowledge and contributing to the development of a greater understanding of image recognition.