Background and Mission

The Machine Vision Group (MVG) is renowned world-wide for its expertise in computer vision. The group has a solid record, which now spans 28 years, of scientific merits on both basic and applied research on computer vision. The mission of the group is to always pursue actual research challenges and improve the state-of-the-art methods.

MVG works as a single well-focused research group, and intensively collaborates with other groups of similar status in Europe, the USA, China and Japan. The group has been invited to take part in several European project proposals, and during the reporting period had a joint research project on face analysis and visual surveillance together with Prof. Stan Z. Li from the Institute of Automation at the Chinese Academy of Sciences. These are clear signals of our attractiveness as a distinguished partner in the global research community. Within the Seventh Framework Programme FP7, the group currently participates in a project consortium of Mobile Biometry (MOBIO), which is coordinated by the IDIAP Research Institute, Switzerland. A new project called Tabula Rasa will continue the biometry theme and the MVG has a significant role in this as well.

The research areas of the group range from generic computer vision methodologies to machine vision applications and vision systems engineering. The results of our research have been widely exploited in industry, and contract research forms a part of our activities. The current areas of research interest include texture analysis, geometric computer vision, computational imaging, face analysis, analysis of motion and human actions, vision-based human-computer interaction and energy-efficient architectures for vision computing.

Highlights and Events in 2009

During 2009, the group produced a large number of significant scientific contributions and published its key results through major conferences and journals. The group has been recently very successful in winning highly competitive funding for its long term research from the Academy of Finland. Altogether, as many as eight Academy-funded research or postdoctoral projects were running in 2009. As a highlight, the Affective Human-Robot Interaction project aims at creating leading edge solutions for natural human-robot interaction in smart environments.

The research group has established its position among the most significant research teams in the field of machine vision. One of the indicators of this is, that the Local Binary Pattern (LBP) papers of the group have received over 3300 citations in Google Scholar by the first quarter of 2010. An invited article on the LBP methodology has appeared on Scholarpedia: www.scholarpedia.org/article/Local_Binary_Pattern.

In addition, visibility outside academic forums has been achieved. The group has a tradition of presenting regularly its research projects and activities in different media. Last year, this included, for example, coverage in the leading Finnish newspaper Helsingin Sanomat, a national professional magazine Prosessori, and the local newspaper Kaleva.

The group hosts regularly visits by respected and renowned scientists from abroad. In 2009, the group had the pleasure of hosting Professor Hanan Samet from the University of Maryland, USA and Prof. Georgy Gimel’farb from the University of Auckland, New Zealand. In addition, several domestic visitors from both public and private sectors were briefed about our research activities.

The group has established active collaboration with some of the world’s leading institutions and top scientists. It has had collaboration with the University of Maryland (USA) from the early 1980’s. More recent partners include the Chinese Academy of Sciences, the Academy of Sciences of the Czech Republic, Czech Technical University in Prague, the École Polytechnique Fédérale de Lausanne (EPFL), and the consortium of the European Mobile Biometry (MOBIO) project. In 2009, MVG established contacts with the University of Tokyo, Japan and the University of Oxford, UK with postdoctoral researcher visits.

In 2009 MVG hosted visitors from the Chinese Academy of Sciences.

The group fosters international mobility to and from our unit. Four of our researchers made research visits to foreign institutions during the reporting period. The group has at-
Texture is a fundamental property of surfaces. We have de-
trated visiting postdoctoral researchers and postgraduate
students from abroad, who are affiliated with us for periods
of a couple of weeks up to several years. Three such short
visits were made to MVG in 2009 and one long visit began.
In addition, MVG recruited as many as three new post doc-
toral scientists from abroad.

The group and its members are active in the scientific com-

The professors of the group were committee members of
several major conferences and many researchers of the
group served as reviewers for various journal and confer-

The current main areas of the research are: 1) Computer
vision methods, 2) Human-centered vision systems, and 3)
Vision systems engineering.

Computer Vision Methods
The group has a long and highly successful research tradi-
tion in two important generic areas of computer vision: tex-
ture analysis, and geometric computer vision. The fundamen-
tal research in these areas has created a basis for many
novel contributions in our research on human cen-
tered vision systems and vision systems engineering. In
2009, we also started a new research area, computational
imaging, which has recently gained much attention within
the computer vision and computer graphics communities.

Texture Analysis
Texture is a fundamental property of surfaces. We have de-
veloped a novel methodology based on Local Binary Pat-
terms (LBP), which has evolved to represent a breakthrough
in texture analysis. Today, LBP is probably the most widely
used texture operator. It is frequently cited and used all over
the world. With over 430 citations in early 2010, our paper
on local binary patterns published in IEEE Transactions on
Pattern Analysis and Machine Intelligence (PAMI 2002) is,
according to the ISI Web of Knowledge, the most cited
Finnish paper published since 2002 e.g. in electrical & elec-
tronic engineering, computer science and telecommunications
categories, and the 8th most cited paper of the presti-
gious PAMI journal. The group’s more recent LBP-
based publications on face recognition (PAMI 2006) and
dynamic texture and facial expression recognition (PAMI
2007) were the 3rd and 4th most cited papers of PAMI pub-
lished since 2006 and 2007, respectively.

In 2009, our research focused on robust LBP based descrip-
tors for static and dynamic textures, and on texture based
methods for different tasks in face and activity analysis.
The purpose of Timo Ahonen’s doctoral thesis “Face and
texture image analysis with quantized filter response statistics”, published in August 2009, was the development and
analysis of photometric descriptors for the appearance of
real life images. The two application areas included in the
thesis were face recognition and texture classification. The
contributions of the thesis include a face representation
method using distributions of local binary patterns com-
pared in local rectangular regions, a face recognition method
tolerant to image blur using local phase quantization, and
three new approaches and extensions to texture analysis
based on a general framework for image description using
statistics of quantized filter bank responses. The first two
aim at increasing the robustness of the quantization pro-
cess. The soft local binary pattern operator accomplishes
this by making a soft quantization to several labels, whereas
Bayesian local binary patterns make use of a prior distribu-
tion of labeling, and aim for the one maximizing the poste-
rior probability. Third, a novel method for computing rotation
invariant statistics from histograms of local binary pattern
labels using the discrete Fourier transform was introduced.
In April 2010, Ahonen’s thesis was ranked by The Pattern
Recognition Society of Finland as the best Finnish PhD

In breast cancer classification, the dominant factor is the
histological appearance of the tissue. In collaboration with
Docent Johan Lundin from the Biomedical Informatics Group
at the University of Helsinki, we have been developing an
automatic stroma-parenchyma tissue classification system.
Automatic tissue classification could speed up, focus and
add reliability to the visual inspection of histopathological
images. Our experimental results show that joint distribu-
tion of Local Binary Pattern codes and local contrast (LBP/
C) is a strong texture descriptor for histopathological breast
images. Very promising results were achieved using the
AdaBoost algorithm and Support Vector Machine (SVM)
classifier with LBP/C features. Related to this topic, we are
also collaborating in a mobile microscopy project, which
aims to develop a device for medical diagnostics particu-
larly used in developing countries.
In collaboration with the Institute of Computing Technology of the Chinese Academy of Sciences, we have proposed a simple, yet very powerful and robust local descriptor called the Weber Law Descriptor (WLD). It is based on the fact that human perception of a pattern depends not only on the change of a stimulus (such as sound, lighting) but also on the original intensity of the stimulus. Specifically, WLD consists of two components: differential excitation and orientation. The differential excitation component is a function of the ratio between two terms: one is the relative intensity differences of a current pixel against its neighbors; the other is the intensity of the current pixel. The orientation component is the gradient orientation of the current pixel. For a given image, we use the two components to construct a concatenated WLD histogram. Experimental results on texture analysis and face detection problems provided excellent performance. A paper on this topic was accepted by the top-tier IEEE Transactions on Pattern Analysis and Machine Intelligence journal.

Dynamic textures (DT) are image sequences with visual pattern repetition in time and space, such as smoke, flames, moving objects and so on. The Local Binary Pattern histograms from Three Orthogonal Planes (LBP-TOP) method which we recently proposed is an effective descriptor for dynamic texture recognition, as shown by its success in different application problems. A problem with the LBP-TOP descriptor is that it is not rotation invariant, which makes it difficult to deal with rotation variations common in video sequences. Inspired by Local Binary Pattern Histogram Fourier features (LBP-HF), which are a novel rotation invariant image descriptor computed from discrete Fourier transforms of local binary pattern (LBP) histograms, research on extending it to acquire rotation invariant LBP-TOP descriptors was carried out with very good preliminary results.

How to segment dynamic textures is a challenging problem. In 2009, we developed significant improvements to our DT segmentation method published at ICPR 2008. We employ a new spatiotemporal local texture descriptor which combines local binary patterns with a differential excitation measure of the WLD descriptor. We also addressed the important problem of threshold selection by proposing a method for determining thresholds for the segmentation method by statistical learning. An improved criterion for merging adjacent regions was also introduced. Experimental results show that our approach provides very good segmentation results compared to the state-of-the-art.

Dynamic texture synthesis is to provide a continuous and infinitely varying stream of images by performing operations on dynamic textures. The starting point for our research was that the previous video texture method proposed by Schödl et al. (SIGGRAPH 2000) provides good quality visual results, but does not well explore the temporal correlation among frames. As an improvement, we developed a novel spatial temporal descriptor for frame description accompanied with a similarity measure. Our method considers both the spatial and temporal domains of video sequences in representation; moreover, it combines the local and global description on each spatial-temporal plane. To the best of our knowledge, this is the first utilization of texture properties in both the spatial and temporal domains in frame representation for dynamic texture synthesis. A paper on this topic was published in the International Conference on Image Processing (ICIP).

We have continued our work on the novel Local Phase Quantization (LPQ) descriptor which is based on a locally automatic tissue classification with Local Binary Pattern codes makes the visual inspection of histopathological images faster and more reliable.
computed Fourier transform phase. LPQ is insensitive to image blurring, and it has proven to be a very efficient descriptor in face recognition from blurred as well as sharp images. In our latest research, we have compared different approaches for computing the local phase information. According to our experiments, we have found that ordinary short-term Fourier transform can provide the best results in the case of blurred images. We have also improved the blur robustness of the descriptor by introducing an optimization framework that can adjust the descriptor to better tolerate different types of blur, including motion, defocus and Gaussian blur.

**Geometric Computer Vision**

Geometric camera calibration has traditionally been one of our areas of expertise. Recently, our contributions in this research area were summarized in a comprehensive review article, which was published in the Wiley Encyclopedia of Computer Science and Engineering in 2009. The review article gives an overview of camera models and calibration methods used in the field. The emphasis is on conventional calibration methods in which the parameters of the camera model are determined by using images of a calibration object whose geometric properties are known. In addition, the presented techniques are illustrated with real calibration examples in which various kinds of cameras, both omnidirectional and conventional ones, are calibrated using open source software tools developed in our group.

One recent trend in many vision problems which involve image matching, is the success of correspondence growing methods. For example, in object recognition and image retrieval, the idea of correspondence growing has been used to improve the discrimination between correct and incorrect region correspondences. On the other hand, in multi-view reconstruction, patch-based approaches, which use local match expansion, have produced good results. Our previously proposed quasi-dense wide baseline matching method is one particular growing technique which can be used as a basic tool in different recognition and reconstruction tasks. During the year 2009, we have continued our studies with quasi-dense matching and its applications. One of the recent developments is a dense and deformable motion segmentation method for wide baseline image pairs. In this work, we proposed a bottom-up motion segmentation approach which gradually expands and merges the initial matching regions into smooth motion layers and finally provides a dense assignment of pixels into these layers. Besides segmentation, the proposed method provides the geometric and photometric transformations for each layer. Hence, due to its ability to deal with deforming motions and large illumination changes, our approach has a wider range of applicability than the previous methods.

In many computer vision applications, images are degraded by motion or defocus blur, which may cause significant problems to image analysis. To overcome these problems, one can use deblurring techniques to eliminate the blur, but this is likely to result in disturbing artefacts. Another approach is to employ such features in image analysis that are insensitive to blur. In 2009, we have continued our research on blur invariant pattern recognition and registration. We have developed global image features that are based on the Fourier transform phase. In contrast to the LPQ, these features are global descriptors that can be used to geometrically align images and recognize objects that are geometrically transformed. Their main advantage is that they are invariant to centrally symmetric blur kernels including linear motion and out of focus blur. In our latest work, the results obtained with these invariants were greatly improved by using a statistical method to weight the features according to their estimated signal-to-noise ratio. When this method was applied to recognition of blurred and noisy objects, the results were improved by up to 20%.

**Examples of fish images used in the object recognition experiment. Motion blurred, noisy, and shifted versions of the same images can be recognized correctly.**

**Computational Imaging**

Computational imaging is a relatively new field of research, and it refers to the convergence of the camera and the computer. Computational cameras are able to enhance and extend the capabilities of ordinary photographic cameras by using computational techniques to process and reconstruct the image data provided by the camera sensor. In our research, we have developed methods to process the raw sensor data in order to improve the subjective quality of the resulting images by eliminating or correcting the effects of certain error sources.

Color constancy or white balance adjustment is a fundamental element in virtually all modern digital cameras. Accurate color constancy is crucial, since even small errors may cause severe deterioration in the empirical quality of the output image. Current methods for color constancy rely on general assumptions of image colors statistics (e.g. the mean color of the image is always gray), which further require information about the sensor characteristics. Although such approaches work adequately well with an average photograph, they easily make coarse errors when the preselected assumptions are not fulfilled (e.g. the mean color in an image of football field is far from being gray). The situation is even worse if accurate sensor characteristics are not available, which is commonly the case with low cost sensors.

In our research, we have applied higher level content understanding to color constancy estimation. The developed algorithm finds semantically meaningful objects from the image, whose color is known in advance. If such objects, e.g. grass or snow, are found, we can use this knowledge to make more realistic assumptions on the color statistics and sensor characteristics to be applied in color adjustments.

During the acquisition of a digital image, the camera sensor array records the spectral irradiance values for each pixel location. Raw data of this kind do not usually give a representation of the scene as a human would perceive it. They
must be processed and refined in several stages before a useful image is output to the user. However, the operations performed at each stage also contribute to altering the original properties of the sensor noise.

Since the properties of noise for a specific sensor are usually known, one can obtain a fairly accurate mathematical model which helps to improve noise reduction. This is the main advantage of de-noising raw data. There exist many algorithms for raw image de-noising, but it is not trivial to measure their objective and subjective performance, and the literature lacks exhaustive comparisons between different methods. We proposed a novel quality measure for color images which is based on known mechanism of human visual perception. We then executed an extensive session of experiments with real and artificially degraded noisy images using several state-of-the-art de-noising algorithms, and evaluated their performance with ours and other quality measures proposed in the literature.

Human-Centered Vision Systems

It is widely predicted that computing will be moving into the background, being omnipresent and invisible, and projecting the human user into the foreground. Therefore future ubiquitous computing environments should be designed to be human-centered instead of computer-centered. Computer vision will play a key role in implementing human-centered systems, for example, for natural human-computer interaction (HCI) or for identifying humans and their behavior in smart environments.

Face Analysis

In 2009, we continued the research on soft biometrics from videos, and introduced new methods based on manifold learning. Soft biometrics, including age, beard, gender, glasses, ethnicity, eye/hair color, height/weight, skin color etc., provide some vague physical or behavioral information which is not necessarily permanent or distinctive. This is very useful in many applications such as human-machine interaction, and content based image/video retrieval. Our extensive experiments on the gender and age classification problems showed that the developed manifold learning based approach yields excellent results, outperforming those of traditional methods. The idea behind the proposed approach consists of learning the intrinsic characteristics of the subjects in each class (e.g. male, female, young etc.) from the training video sequences by discovering the hidden low-dimensional nonlinear manifold of these subjects.

In collaboration with Prof. Yoichi Sato and his team from the University of Tokyo (Japan), we have also started investigating the recognition of blurred faces combining facial deblurring and LPQ, gaze estimation using appearance based methods, and audio-video correlation analysis for spoofing detection in biometric systems.

In the period 2008-2010, the Machine Vision Group is participating in the Mobile Biometry (MOBIO) project funded by the European Commission. The scientific and technical objectives of the project include robust-to-illumination face authentication, robust-to-noise speaker authentication, joint bi-modal authentication, model adaptation and scalability. The LBP method developed in MVG plays an important role in the project. In the second year of the project, we actively participated in the design and collection of a multibiometric research database that is recorded using mobile phones during the course of the project, and at later stages used for the development and evaluation of mobile biometric systems. Furthermore, MVG carried on research on face recognition methods based on local binary patterns and local phase quantization. In this research, we had collaboration with Prof. Josef Kittler’s group from the University of Surrey, UK. We have also been developing a scalable face detector, which will be considered for the final demonstration system using the Nokia N900 mobile phone.

In collaboration with Prof. Stan Li’s group, we have collected a novel Oulu-CASIA NIR-VIS facial expression database, which includes the six facial expressions from both near infrared (NIR) and visible light (VIS) imaging systems with 80 people (50 from Oulu and 30 from CASIA). We have carried out experiments on this database using LBP-TOP for illumination invariant facial expression recognition. We plan to release this database and provide the baseline results.

We have made an extensive investigation on visual speech recognition. Visual speech information plays an important role in lip reading under noisy conditions, or for listeners with a hearing impairment. We utilized local spatiotemporal descriptors to represent and recognize spoken isolated phrases based solely on visual input. Spatiotemporal local binary patterns extracted from mouth regions are used for describing isolated phrase sequences. The advantages of our approach include local processing and robustness to monotonic gray-scale changes. Moreover, no error prone segmentation of moving lips is needed. The results were published in IEEE Transactions on Multimedia.

We also developed a face tracker, which creates an adaptive color histogram model for tracked objects and applies motion estimation, model based tracking, and generic face detection for keeping track of the faces. Multiple faces can be tracked at the same time using a single camera. Occlusion
Human-Computer Interaction

Methods for organizing images automatically based on their content are becoming more and more attractive as the number of photos in personal collections is rapidly increasing, and finding particular images manually from these collections is becoming too laborious. In our research, we have developed a new generic method for organizing images and image categories. While the traditional approach for categorization is to divide the data into disjoint classes or clusters, the basic idea in our approach is to find the intrinsic structure of the image collections, and to utilize this structure to organize images. Our generic structure for describing the connections between categories creates a more natural and versatile way of organizing the images than the commonly used tree-shaped structures.

The objective of the ongoing Academy-funded project “Affective Human-Robot Interaction”, carried out jointly with Prof. Juha Röning’s Robotics Group, is to develop leading edge solutions for affective human-robot interaction (HRI) in smart environments. An intelligent robot should be able to detect and identify the user in order to personalize its services and guarantee security. Furthermore, it should recognize the user’s emotions to allow affective interaction, and be able to communicate easily with the user in order to understand given commands by recognizing speech and gestures. The robot should also learn to change its behavior according to the user and his/her emotional state.

The focus in the research during the first year of the project was in developing new methods and tools for computer vision and robot navigation to be used in integrated HRI systems. The design of an experimental HRI system was also begun.

As a part of this project, research aiming at video-realistic speech animation was also begun. An animated talking face (avatar) with realistic mouth movements and facial expressions can provide a natural way for interaction between humans and robots. In the first phase, problems of visual speech analysis were investigated and a new method based on manifold learning was developed for lip reading. Very promising results were obtained.

Analysis of Motion and Human Actions

The research efforts on detection and tracking moving objects in camera networks resulted in the development of a new algorithm able to re-recognize objects across camera views. The algorithm uses adaptive boosting and a multitude of image properties, like color, texture and shape, to learn object templates, which can then be used to track objects from one view to another. Experiments on real-world datasets suggest that tracking objects in a real-time camera network should be feasible.

In crowd scenarios, people often walk in groups. We developed a new solution to the problem of matching groups of people across multiple non-overlapping cameras. Here, we use “group” to refer to a small number of people walking in proximity. Similar to the problem of matching individuals across cameras, matching groups of people also encounters challenges such as variations of illumination conditions, poses and camera parameters. Moreover, people often swap their positions while walking in a group. We proposed using a covariance descriptor in appearance matching of group images. Experimental results demonstrate the effectiveness of the proposed method.
Human motion can be seen as a type of texture pattern. We adopted the ideas of spatiotemporal analysis and the use of local features for motion description by describing human movements with LBP-TOP dynamic texture features. The proposed description is computationally simple and suitable for various applications such as action and gait recognition. We tested our method with popular Weizmann and KTH datasets for action recognition, and an article describing the results has been accepted by the Machine Vision and Applications journal. Gait recognition results on the CMU MoBo database were published in the International Conference on Biometrics (ICB).

The ability to separate informative regions from background clutter is an essential requirement for a practical vision algorithm. Biological systems have developed to be remarkably effective in focusing their visual attention to relevant targets, unlike computer vision where background subtraction is still an unsolved problem.

Background subtraction is commonly approached by detecting moving objects against a nearly static background. Joint research on this kind of approach was carried out with Prof. Stan Z. Li’s group. The LBP operator was extended to a scale invariant local ternary pattern (SILTP) operator, which is effective for handling illumination variations, especially soft cast shadows. Excellent results were reported for different types of video sequences. A paper with the title “Modelling pixel process with scale invariant local patterns for background subtraction in complex scenes” was accepted for the top-tier IEEE Conference on Computer Vision and Pattern Recognition.

While effective in certain scenes, this kind of approach has severe problems when the scenes are dynamic. A recent approach to background subtraction in dynamic scenes is provided by saliency detectors. The motivation for this approach originates from the attention mechanism of the human visual system, which enables us to quickly focus on general salient objects without preceding training. The most recent saliency detection methods are based on phase response of the estimated Fourier transform. Although this approach is easy to compute, it provides rather poor accuracy and few possibilities to adjust the method for the current application.

In our group, we have developed a new saliency detection method which is based on semi-local feature contrast of the image. Without trading off much of the computational simplicity, we achieve far better accuracy than the existing methods, and also provide several possibilities to tune the detector according to the requirements of the application. The developed algorithm is also applicable to video sequences.

We have also developed a new method for tracking multiple objects from image sequences using detector responses as measurements. The method utilizes soft assignment to associate the detection responses to the objects tracked. Due to soft assignment, we are able to cope with inaccurate responses and inter-object occlusions. The method includes a component which combines Kalman filtering and expectation maximization (EM) algorithms to estimate the parameters of the objects tracked and to assign the measurements softly. One of the further benefits of this approach is that neither iterations nor long measurement history are needed.

**Vision Systems Engineering**

Vision systems engineering provides guidelines for methodological research, helping to identify attractive computing approaches, architectures, and algorithms for useful commercial systems. In practice, solutions from low-level image processing to even equipment installation and operating procedures are considered simultaneously. The roots of this expertise are in our industrial visual inspection studies in which we met extreme computational requirements already in the early 1980’s, and we have contributed to the designs of several industrial systems. We have also applied our expertise to applications intended for smart environments and embedded platforms such as mobile phones.

On mobile platforms, we have developed a document mosaic constructor that stitches together frames selected from video sequences. The user interactive image capturing process analyzes the displacements between video frames, measures the quality of the frames based on the motion blur and focus, and selects the suitable frames for mosaicing. The implementation includes a parallel pipeline that uses the mobile Graphical Processing Unit (GPU) integrated on the platforms, reducing the computation and power needs of the application.

**Application flow of the mobile document mosaic constructor.**
Another example of an application that takes advantage of GPU processing is the 3-D image reconstruction software developed in collaboration with VTT. The software constructs a surface map from several images of a sinusoidal pattern projected onto a wooden or metallic sample. The method analyzes the displacements between the images, and computes the phase image using an iterative technique. All the required operations of the algorithm are computed in parallel using general purpose computations on GPUs.

We have also developed a vision based system for distinguishing humans from moving work machines. The system constructed had both narrow-field and fisheye stereo cameras for capturing the image data, which made it possible to use the same system also for image-based 3-D modeling of the work machine environment. Several human detection methods using 2-D pattern recognition, motion analysis, and 3-D reconstruction were implemented and experimented with. The system was evaluated with extensive image sequences of authentic work machine environments, and it proved to be feasible.

The results of our recent body part segmentation work have been applied to unusual activity recognition in home environments. The most important event we seek to detect is falling over. The system developed works robustly when the person monitored is partially occluded or there is noise in the video data. We are developing low cost systems that are easy to install. During the year 2010 we shall begin using cameras with very thin optics. The thin optics lower the overall cost of the systems, but also introduce new challenges for the detection. The resolution of such a camera is very low. The work in the future will focus on adopting the new cameras to the system and testing the performance in real environments.

One of the promising uses for visual inspection is automated strength grading for sawn timber. The solution under development employs real-time feature extraction, classification, and the Finite Element Method (FEM) combined into an adaptive learning scheme. The idea is to launch a FEM based analysis whenever the region under inspection deviates from those modeled earlier. The 3-D model and element mesh needed in the FEM analysis are built from the data.
acquired during feature extraction, utilizing knowledge of the grain structure of wood. Various FEM analysis methods are surveyed, from almost complete reconstruction of the microstructure of wood to much faster, but less precise material parameter approaches.

Computer vision applications are characterized by both high data input rates and high computational costs. Often they need to be realized in a small package, such as a mobile communications device that may consume a limited amount of power for application processing like video coding. This prevents implementations based on conventional processors, even in the future, but hardware acceleration is mandatory. Currently, only monolithic long latency accelerators improve energy efficiency, but they are rigid and costly to design and verify, and difficult to justify for purposes which are considered marginal and computationally expensive.

In 2009, a three year Academy of Finland funded project PARadigm Shift for Embedded Computing (PARSEC) was launched, targeting automation of the tools for designing ultra-energy-efficient systems for computationally demanding purposes, such as video coding and image analysis. The research is done in cooperation with Tampere University of Technology (Prof. Jarmo Takala) and Åbo Academi University (Prof. Johan Liljus). We have demonstrated that fine grained, silicon area efficient, adaptable hardware accelerators can be employed at very low software interface overheads through deterministic multithreaded schedules. This has turned out to be much more efficient than the conventional interrupt, semaphore, and event handler mechanisms advocated by the textbooks. In essence, we are targeting a new paradigm for embedded computing and expect significant impacts in the field. Our developments include the simultaneous decoding of multiple MPEG-4 streams on shared accelerators and MIMO reception. The latter work has been carried out in cooperation with the Centre for Wireless Communications (CWC).

Research working towards fine-grained embedded multiprocessor processing continued during 2009. The previously published approach of efficient multiprocessor scheduling was extended to take into account the memory organization of the underlying platform. Another direction in the development was that of automatic multiprocessor scheduling for Reconfigurable Video Coding decoder descriptions. This approach can also be generalized to other applications that have been described in the CAL actor language.

Magnetic resonance imaging (MRI) of the human body provides 3D-imagery with various contrast options sensitive to different biological phenomena. As such MRI has become an important imaging technology both clinically and for medical sciences. Functional MRI (fMRI) is concerned with indirect imaging of brain activity. fMRI data provides 3D-video on blood oxygenation coupled with electrical activity through metabolic control over vasculature in the brain. Analysis of the 4D-domain combines image analysis with biomedical signal processing.

We have recently started to analyze factors contributing to spatial Independent Component Analysis (sICA) which is currently a widely used analysis method in fMRI. Furthermore, also temporal measures related to complexity and power distribution of signals have been investigated as a means of providing novel clinically useful contrasts based on fMRI data. The research is done jointly with Department of Diagnostic Radiology, Oulu University Hospital.

In addition to several research projects, we have begun developing a Computer Vision Algorithm Library. Over the years, our group has developed numerous algorithms and methods using different development environments and tools. Now these algorithms are being collected into a single C++ library. The intention is to make efficient implementations of the existing algorithms easily reusable for both new research and as part of practical applications like a robot’s vision system. We plan to release this software library in the future with an open source license.

**Exploitation of Results**

A clear indication of the impact of our research is that many of our papers are frequently cited. According to the ISI Web of Knowledge, the most cited paper from Finland published since 2002 in the ICT-area (electrical & electronic engineering, computer science and telecommunications categories) is our PAMI paper from 2002 (out of about 5500 papers). We have also authored three most cited papers from Oulu in the ICT area published since 1995. Altogether six of our publications are within the top ten, and nine within the top 20 in the same ranking.

Many researchers have adopted and further developed our methodologies. Our research results are used in a wide variety of different applications around the world. For example, the Local Binary Pattern (LBP) methodology is used in numerous image analysis tasks and applications, such as biomedical image analysis, biometrics, industrial inspection, remote sensing and video analysis.

The MVG has actively encouraged and supported the birth of research group spin-outs. This gives an opportunity for young researchers to start their own teams and groups. Side results are the spin-out enterprises. According to our experience their roots are especially in the strands of “free academic research”.

There are currently altogether five research based spin-outs founded directly on computer vision area. The number of the spin-outs could be extended up to thirteen when taking into account the influence of the MVG’s almost thirty-year old history and the spin-out companies’ from the spin-out research groups in the area of computer science and engineering in total.

In 2009, one of the most recent spin-outs, Visidon Ltd., released a solution for recognizing objects and faces on mobile phones in real-time. The company also provided face detection technology which supports age, gender and expression recognition for MediaTeam Oulu’s UBIOulu project. Together with Symbicon Ltd., Visidon presented a new audience measurement and reporting system, where Visidon’s face technology (age, gender, smile, activity) is integrated into the Symbicon displays and a web-based user interface is provided for easy access.
Future Goals

Research on human-centered vision systems will be further extended. For example, we are launching collaboration with Prof. Tapio Seppänen’s signal analysis team in the area of multimodal analysis of human communicative behavior, combining vision with speech and biosignal data.

The use of multimodal data has also a key role in the new European project called Trusted Biometrics under Spoofing Attacks (Tabula Rasa) in 2010-2014. MVG was invited to be a partner in this project consortium on the basis of the previous contribution the group made for the success of the Mobile Biometry project.

Among our research staff, as much as 40% are currently from abroad. Due to our excellent international reputation, we increasingly attract visitors from abroad to join us for some time, and many of our researchers are willing to make research visits to leading groups abroad. Such bilateral collaboration will bring fresh new ideas and expertise to our research.

Close interaction between basic and applied research has always been a major strength of our group. The scientific output of the group has been increasing significantly in recent years. With this we expect to have much new potential for producing novel innovations and exploitation of research results in collaboration with companies and other partners.

Personnel

| professors & doctors | 15 |
| graduate students    | 15 |
| others               | 8  |
| total                | 38 |
| person years         | 29 |

External Funding

<table>
<thead>
<tr>
<th>Source</th>
<th>EUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academy of Finland</td>
<td>514 000</td>
</tr>
<tr>
<td>Ministry of Education</td>
<td>123 000</td>
</tr>
<tr>
<td>Tekes</td>
<td>344 000</td>
</tr>
<tr>
<td>domestic private</td>
<td>238 000</td>
</tr>
<tr>
<td>international</td>
<td>74 000</td>
</tr>
<tr>
<td>total</td>
<td>1 293 000</td>
</tr>
</tbody>
</table>

Doctoral Theses


Selected Publications


Chen J, Yi D, Yang J, Zhao G, Li SZ & Pietikäinen M (2009) Learning mappings for face synthesis from near-infrared to visual...


