Background and Mission

The Machine Vision Group of the University of Oulu was established in 1981. It has achieved a highly respected position in its field, and its results have been widely exploited in industry. The objective of the group is to carry out leading edge long term research on machine vision technology.

The wide-spread use of digital cameras has already created large markets for efficient digital image processing and compression techniques. Regardless of this, the great majority of current camera systems are still blind, lacking the ability to understand the contents of the imagery taken by the cameras. The goal of computer vision (machine vision) is to make useful decisions about real physical objects and scenes based on sensed images. The field is faced with a large number of scientific and engineering problems to be solved, and expertise from many disciplines is needed, including computer science, electrical engineering, mathematics, physics, and cognitive sciences.

The focus areas of the research are: 1) texture-based computer vision, 2) geometric image and video analysis, 3) machine vision for sensing and understanding human actions, and 4) vision systems engineering. Our approach of combining world-class basic research in chosen key areas with more applied research on vision systems engineering is quite unique among the research groups in computer vision.

The group and its members are active in the scientific community. For example, in 2004 Prof. Matti Pietikäinen served as an Associate Editor of IEEE Transactions on Pattern Analysis and Machine Intelligence and Pattern Recognition journals; and members of the group were on the committees of several international conferences. Jari Hannuksela won the prize for the best M.Sc. thesis 2003 awarded by the Pattern Recognition Society of Finland for his Master’s thesis on facial feature based head tracking and pose estimation.

In November 2004, the group qualified for the second round of applications for the Finnish Centres of Excellence in the research programme 2006-2011 appointed by the Academy of Finland. The group is among 53 applicants that were selected from 143 plans of intent. It was also very significant that Prof. Janne Heikkilä was awarded quality unit funding from the University of Oulu for his project “Image analysis with geometric invariants and descriptors”.

Scientific Progress

Texture-Based Computer Vision

Image texture analysis is an important fundamental problem in machine vision. During the past few years, the group has developed theoretically and computationally simple, but very efficient nonparametric methodology for texture analysis based on Local Binary Patterns (LBP). The approach has evolved to represent a significant breakthrough, outperforming earlier methods in many applications. It is already widely used all over the world.

Recently, we proposed a novel facial representation for face recognition based on LBP features obtaining excellent results. In this approach, the face image is divided into several regions from which the LBP features are extracted and concatenated into an enhanced feature vector to be used as a face descriptor. However, it was unclear whether the high performance was due to the use of local regions (instead of an holistic approach) or to the discriminative power of LBP. We investigated this issue by comparing four different texture features in describing the appearance of local regions. The experimental results clearly showed the superiority of the LBP based approach.

Two approaches to facial expression recognition from static images were developed using LBP histograms computed over non-overlapping blocks for face description. In the first method, the Linear Programming
technique is adopted to classify seven facial expressions (anger, disgust, fear, happiness, sadness, surprise and neutral). In another approach, a coarse-to-fine classification scheme is used. Good results were obtained for the Japanese Female Facial Expression (JAFFE) database used in the experiments.

In addition to facial image analysis, we have started researching LBP based methods for other biometric identification tasks, including irises. The goal is to develop biometric identification algorithms for low-cost equipment such as mobile devices. The systems proposed so far require special instrumentation that limits their use. With the new methods, reliable identification would be possible with practically any camera.

A novel texture-based method for modeling the background and detecting moving objects from video sequences was developed. Each pixel is modeled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel. The method was evaluated against several video sequences including both indoor and outdoor scenes. It was shown to be tolerant to illumination variations, the multimodality of the background, and the introduction or removal of background objects. Furthermore, the method is capable for real-time processing.

Approaches to using LBP in content-based image retrieval were also studied. Block-based methods dividing the query and database images (or database images only) into blocks and comparing their LBP histograms were found to perform significantly better that methods based on global LBP histograms. The results for the LBP based approach were also better than those obtained with the widely used color correlogram features. Image databases taken from the Corel Image Gallery and from a German stamp database were used in experiments.

Research on visualization based learning in texture analysis was continued. Earlier, a visualization of LBP feature distributions with a self-organizing map was utilized in optical paper characterization and paper formation analysis. In the previous reporting period, the ISOMAP method was found to be promising in the visualization of high-dimensional texture feature data and in the training of a texture classifier. The research on utilizing the ISOMAP in texture analysis was continued. A visualization based approach was developed for finding the most discriminating LBP features and building a labelled training set for a texture classifier. The approach was utilized in paper texture analysis. A similar approach was also applied to wood grain texture analysis to study visually how local texture features can discriminate different grains.

The problems of texture image labelling were considered and a very promising labelling framework was developed. Textures were modeled with complementary measures including various versions of the LBP and Gabor features. Combined use of active learning, co-training, and visualization based learning was applied to feature data, enabling comprehensive, accurate, and user-friendly texture labelling. Also research concerning unsupervised learning of discriminative texture models from large data sets was initiated.

Research on the unsupervised learning and dimensionality reduction method, locally linear embedding (LLE), was continued. A new semi-supervised framework for visualizing high-dimensional data has been proposed. It is based on a combination of multiple machine learning algorithms, i.e. metric learning, intrinsic dimensionality estimation, linear feature extraction, and nonlinear feature extraction.

As a spin-off of the learning research mentioned above, the K-Local Hyperplane Distance Nearest Neighbor algorithm together with feature normalization was adapted for one of the important bioinformatics tasks - protein fold recognition. Fold recognition concerns structure (rather than sequence) of proteins and therefore it is very useful when there is a weak identity between protein sequences. Knowing protein structure often simplifies discovering protein function, which is an ultimate goal in biology and medicine. A comparison (on the publicly available dataset) with the state-of-the-art in pattern recognition - support vector machines - demonstrated the superiority of the chosen approach and encouraged further exploration.

**Geometric Image and Video Analysis**

The research on geometric image and video analysis has been focused on two main topics: 1) geometric invariants and descriptors, and 2) tracking and motion estimation. In geometric invariants and descrip-
tors, a novel statistical approach has been developed to produce invariants that are insensitive to various geometric transformations. As a result, an image transform invariant to nonlinear perspective mapping has been proposed. The affine invariant multi-scale autoconvolution (MSA) transform introduced earlier is based on the same principle. Another affine invariant transform has also been developed, which requires only a fraction of the computation needed by the MSA, but its performance is superior to, for example, affine invariant moments. The research on image invariants has resulted in several scientific articles, and one of them has been accepted by the most respected journal in the field.

The group has discovered a means of estimating geometric transformation parameters using the same statistical approach as with the invariants, enabling us to apply those techniques also for image registration. Based on this solution, a registration method has been derived using the MSA transform that is capable of extracting affine transformation parameters between two images without requiring any point or feature correspondences. A further approach for featureless image registration has been developed that is based on the shift-invariance property of the image bispectrum. As only the phase information is utilized, the method is very robust against illumination changes, occlusions and background clutter. The method is capable of estimating the parameters of the similarity model including scale, rotation and translation.

In tracking and motion estimation, a new method for computing the statistical uncertainty of the block-motion estimates has been invented. Using this method, it becomes possible to estimate global camera motion quickly and reliably. The robust estimator developed tolerates also inconsistent motion regions in the scene. A video stabilization method utilizing this concept has been demonstrated with excellent results. The same approach has also been a basis for a new user interface technology for mobile terminals such as smart phones. Our solution utilizes a camera as a motion sensor that provides the motion parameters of the device. This information can be used for controlling the user interface so that, for example, browsing is performed simply by moving the terminal with respect to some static background. This solution has been implemented on a Symbian operating system, and demonstrated with a smart phone.

The motion estimation research has resulted in a new algorithm for computing the motion vectors for the H.264 video coding standard with various block sizes. This method is an extension of the multilevel successive elimination algorithm (MSEA) originally developed for fixed sized image blocks. The group has also investigated solutions for analyzing video and detecting moving persons based on block-motion vectors. A video surveillance system monitoring room entrance has been demonstrated using the ideas developed in this research. Enhancement of surveillance video quality is another topic of research carried out in the group during the last year. As a result, solutions for real-time software based video deinterlacing, and content-adaptive noise filtering were implemented. Methods for super-resolution image enhancement and motion deblurring were also investigated.

In tracking and motion estimation, a new method for computing the statistical uncertainty of the block-motion estimates has been invented. Using this method, it becomes possible to estimate global camera motion quickly and reliably. The robust estimator developed tolerates also inconsistent motion regions in the scene. A video stabilization method utilizing this concept has been demonstrated with excellent results. The same approach has also been a basis for a new user interface technology for mobile terminals such as smart phones. Our solution utilizes a camera as a motion sensor that provides the motion parameters of the device. This information can be used for controlling the user interface so that, for example, browsing is performed simply by moving the terminal with respect to some static background. This solution has been implemented on a Symbian operating system, and demonstrated with a smart phone.

The motion estimation research has resulted in a new algorithm for computing the motion vectors for the H.264 video coding standard with various block sizes. This method is an extension of the multilevel successive elimination algorithm (MSEA) originally developed for fixed sized image blocks. The group has also investigated solutions for analyzing video and detecting moving persons based on block-motion vectors. A video surveillance system monitoring room entrance has been demonstrated using the ideas developed in this research. Enhancement of surveillance video quality is another topic of research carried out in the group during the last year. As a result, solutions for real-time software based video deinterlacing, and content-adaptive noise filtering were implemented. Methods for super-resolution image enhancement and motion deblurring were also investigated.

Local motion estimation: (a) feature block selection, (b) motion profile analysis; characteristics of the image block are reflected in the confidence measure.

**Machine Vision for Sensing and Understanding Human Actions**

Machine vision provides a unique capability for understanding human actions. It is based on non-contact, passive sensing technology, and a large area can be covered by a single sensor. Our hypothesis is that it is possible to realize a large number of useful machine vision applications by using several distributed vision sensors that each adds a new view or viewing angle. A new framework of building such systems is being studied. An intelligent room equipped with multiple cooperating vision units is used as a test environment.

Human-computer interaction is one of the key issues in the development of proactive systems. Communi-
A framework for proactive machine vision.

With the proactive system can be based on smart user interfaces where the system derives the user input from, for example, gestures or facial expressions. In our group, we have investigated two types of approaches for camera based user interfaces. In the first approach, the system tracks the facial features of the user and determines the gaze direction, which can be directly used to control the pointer of the user interface in the same way as the mouse is used in a traditional PC user interface. The other approach developed is more suitable for a mobile system with an internal camera, and it is based on global motion estimation that produces the motion parameters of the device. These motion parameters can be exploited for controlling the user interface, which enables browsing and zooming large objects that do not fit into the display as such. Both approaches have been demonstrated with real-time implementations.

An experimental system for access control based on face recognition was under development. The system opens the door of a laboratory room only for selected persons while denying access to the others. It is implemented with novel methods developed by the group, including color based face detection using a skin locus model and hierarchical filtering, model selection for view based face recognition using the locally linear embedding algorithm and K-means clustering, and texture based face recognition using local binary patterns.

Ways of using sequences of human postures to describe human activities were studied. A novel silhouette based method was developed, in which affine invariant Fourier descriptors are used to describe the human pose in a frame. A support vector machine classifier is used for recognizing the posture class and Hidden Markov Models are used for classifying posture sequences. The system was able to recognize 15 different continuous human activities in real-time with 90% accuracy.

We have considered distance education as a practical problem that could greatly benefit from the technology under development. The classroom can be actually considered a smart room that has been equipped with sensors like microphones and cameras. An experimental system is being developed, which will assist the lecturer to select the most appropriate camera sources and visual angles that will be transmitted to the students. The system makes the decisions by analyzing the images obtained from the different cameras. Currently, the system has three cameras: an overall camera, a pan-tilt camera and a document camera.

The research made in co-operation with the French research center INRIA Rhone-Alpes addressed the problem of tracking motion of articulated objects from their 2-D silhouettes gathered with a small number of synchronized cameras. The method relies on (i) building 3-D observations (surface patches) from image silhouettes using a feed-forward method and (ii) on fitting an articulated object model to these observations through minimization. The objective function measuring the discrepancy between the model and data takes into account both the scaled algebraic distance from data points to model surface and the difference in orientation between observed surface patches and normals to the model’s surface. The results from the project are very promising. The new innovation, utilizing surface normals in model fitting, increases the method’s efficiency allowing more noisy input data.

Vision Systems Engineering

Vision systems engineering bridges the gap between algorithms and applications. It guides the methodological research to directions that improve the likelihood...
of practical exploitation, finding and identifying approaches, architectures, methodologies, and algorithms that enable the building of useful real world systems. Solutions from low level image processing to even equipment installation and operating procedures are considered.

Visualization based non-supervised training methodology that is currently used in a number of industrial inspection systems on the market, from food sorting to lumber inspection. The central idea is to define the classification boundaries graphically to a 2-D projection of the multidimensional feature space, completely avoiding the tuning of numerical parameters. The non-supervised learning and dimensionality reduction techniques used include Isometric Mapping (ISOMAP), Locally Linear Embeddings (LLEs), and Self-Organizing Maps. The implemented systems employ non-segmenting image analysis techniques based on color and/or LBP histogram features.

In 2004, our group started a research project that targets the creation of a lumber strength grading methodology. This work is based on the use of LBP histogram features in combination with visualization based non-supervised training and FEM (Finite Element Method). We are also developing a solution for measuring the particle size distribution of a crushed aggregate used as the stone material in road pavements and concrete. The methodology is based on the non-segmenting approach, unsupervised classification methods, and texture features such as LBP. The goal is to find a means for implementing an on-line measurement instrument for an industrial environment. The non-supervised visualization based approach is used even for plant vitality measurement from spectrographic imagery. This work is at the research stage and aims at the utilization of photosynthesis and water content information in analyzer systems.

The dynamic ranges of the camera sensors are rather limited. As a cure we are endeavouring to create techniques that enable the capturing of good quality images from unevenly illuminated scenes. With a high frame rate camera this could be done by, for example, using alternating integration times for sequential frames. The proper integration times could be determined via content analysis, for example, by evaluating the quality of human “skin signal”. This image acquisition technology will be useful in proactive systems and visual surveillance applications.

We are also pursuing novel computer architectures to find efficient means for implementing image and video based systems. To minimize power dissipation, a problem for both performance computing and battery powered devices, hardware accelerated techniques need to be favored. Previously we have developed Number Theoretic Transform (NTT) based techniques that enable exploiting correlation and convolution theorems to cut the number of memory accesses, when compared to spatial domain computations. This work is being continued, while simultaneous multithreading techniques are studied to enable employing fine grained accelerators in an efficient manner, for instance, to implement real-time video codecs.

Exploitation of Results

The results of our research were applied to real-world problems in many projects, often in collaboration with industrial and other partners. Some examples of exploitation are described below.

In the wood inspection area, several board edgers using the visual training approach have been installed at sawmills by our industrial partner inX Systems. These have become fully operational within hours of installation unlike earlier systems which required at least a several weeks training period. Recently, inX Systems obtained the first ECVision (Excellence on Cognitive Vision Systems) Prize of the EU for its OptiGrader machine vision system which applies our research results.

The research in texture analysis and visualization methods has given birth to a new company, Intopii. The company is applying the results of the research to industrial visual inspection problems and is licensing its technology world-wide. Intopii currently employs four people full-time, and is growing steadily.

Our method for on-line paper characterization has also been on trial use in several paper mills in Finland and abroad. Our industrial partners have expressed interest in utilizing visualization based data analysis meth-
ods, like ISOMAP, in the training and further analysis of the paper texture data.

The research in video sequence processing and analysis has contributed to recent video codec products that have been integrated into mobile communications devices. Although the key role of the researchers was mostly in enabling rapid product development, the result is a convincing demonstration of the benefits of long term research.

Future Goals

We will continue to strengthen our long term research, researcher training and international activities. We will also continuously seek opportunities for exploitation of our research results by collaborating with partners from industry and other research institutions in national and international research programs and projects.

Personnel

<table>
<thead>
<tr>
<th>Classification</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>professors &amp; doctors</td>
<td>8</td>
</tr>
<tr>
<td>graduate students</td>
<td>15</td>
</tr>
<tr>
<td>others</td>
<td>11</td>
</tr>
<tr>
<td>total</td>
<td>34</td>
</tr>
<tr>
<td>person years</td>
<td>25</td>
</tr>
</tbody>
</table>

External Funding

<table>
<thead>
<tr>
<th>Source</th>
<th>EUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academy of Finland</td>
<td>142 000</td>
</tr>
<tr>
<td>Ministry of Education</td>
<td>176 000</td>
</tr>
<tr>
<td>Tekes</td>
<td>232 000</td>
</tr>
<tr>
<td>domestic private</td>
<td>209 000</td>
</tr>
<tr>
<td>EU + other international</td>
<td>4 500</td>
</tr>
<tr>
<td>total</td>
<td>763 500</td>
</tr>
</tbody>
</table>

Selected Publications


