Tracking with Structure in Computer Vision TWIST-CV

Project Proposal

14 June 2005

Proposed by:

WALTER KROPATSCH

Pattern Recognition and Image Processing Group
TU Wien
Favoritenstraße 9/1832
A-1040 Wien
krw@prip.tuwien.ac.at

1 Scientific Aspects

1.1 Introduction and Motivation

This research project proposal emerged from the close cooperation of Advanced Computer Vision GmbH (ACV) with the Pattern Recognition and Image Processing Group (PRIP), Vienna University of Technology.

Research performed at ACV on surveillance and tracking has delivered several new approaches to head detection and tracking [24, 71, 3] as well as person tracking [9, 68, 75]. One focus of this work was to use the correspondences of 2D and 3D data to reconstruct the scene in spatial and temporal dimensions. Although these approaches were successful, it turned out for many practical applications that current methods are not robust enough. In addition, for each particular application a different approach is used. Therefore, the main idea of the proposed project is to develop a general framework for tracking and stereo that is based on structure. Existing frameworks either do not use the structural approach described later, or – like the general analysis graph (GANAG) – have been developed for machine vision tasks [72].

During the work in recent research projects at PRIP (FWF Graph Pyramid and Geograph Projects), it was discovered that graph pyramid matching is a very powerful means to solve problems in computer vision. It was investigated, how image graphs can be used to preserve structure and topology. Promising initial results have been obtained for applying this methodology to tracking and reconstruction. An exploration of these methods and an incorporation into a framework for tracking and stereo should therefore yield useful advances in theory and practice.

The cooperation between ACV and PRIP within the proposed project also makes sure that different viewpoints are taken into account and that diverse backgrounds of involved people will stimulate the work. Whereas PRIP has gained a lot of experience in structural computer vision, ACV – a new research oriented company – does research on surveillance and tracking problems within its kPlus program using the technological synergies of its scientific and industrial partners. Therefore, ACV can – besides its scientific viewpoint – exploit the experiences gained during the work with companies involved within ACV.

Within the kPlus program many research issues, which should be investigated in a deeper way could be identified. Therefore, the proposed project has the aim of pursuing fundamental research emerging from application problems (not pre-competitive industrial research as defined by the kPlus program) on a number of these research problems.

1.2 Project Aims

The main goal of this project is to develop a framework that enables solutions to practical problems of computer vision using approaches that strongly use *structure*. This framework shall be applicable to the areas where structural computer vision methods are foreseen to be very useful like *Segmentation*, *Tracking* and *Stereo Vision*. Therefore we focus on the following subgoals:

1. Finding object correspondences in image sequences (Tracking): Given is a sequence of images (frames) from a stationary or moving camera. While in many computer vision applications the tasks segmentation, object detection and tracking are often solved stepwise, the underlying idea of the proposed project is to determine a *structure* within the observed scene that is tracked over time. The structure is represented e.g. in a graph or more generally graph pyramid representation of the segmented image and correspondence can be found by graph matching (see Fig. 1). The advantage of using a graph pyramid is that it would allow grouping of structures and hence simplify graph matching. As graph matching is NP-complete, it is only feasible on graphs with few nodes. Higher levels of the pyramid, containing fewer nodes, can be efficiently matched. This matching can then be used to guide the matching of lower levels of the pyramid. The graph representation allows the detection and correction of over- and undersegmentation and therefore leads to a new representation of the scene structure. In this approach the steps of segmentation, detection and tracking are solved in a novel, more integrated way.
2. Finding object correspondences in images from different view points (esp. in stereo configurations): Given are two (or more) images taken from different viewpoints at the same time instant. Standard stereo algorithms use certain features and establish stereo correspondence. This is very often supported by restricting the stereo search by considering the epipolar geometry calculated (or calibrated) in advance. Following this approach the correspondence search is very time-consuming or the calibration procedure cannot be avoided. The approach that shall be followed in the proposed project uses a structural representation e.g. a graph pyramid representation of single images and correspondence can be found again by graph matching (see Fig. 2). Structural representation shall in this respect (i) allow a faster correspondence search (ii) be more robust against single mismatches. Combining the information of the correlated image pair, the 3D structure can also be represented (e.g. by a single graph pyramid).

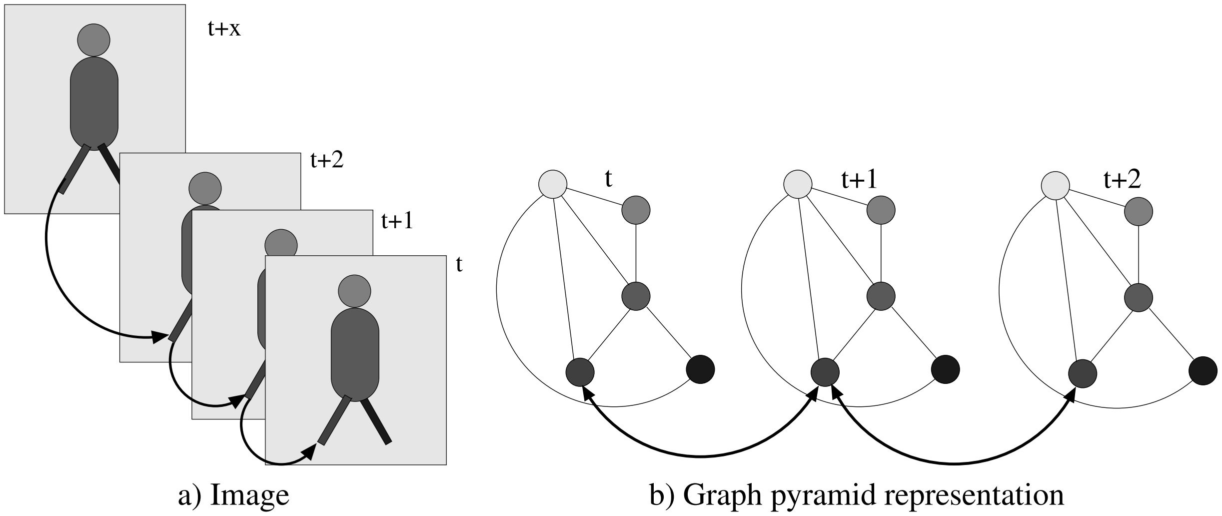


Figure 1: Representation of an image sequence (a) as a graph pyramid (b) of the 2D image structure

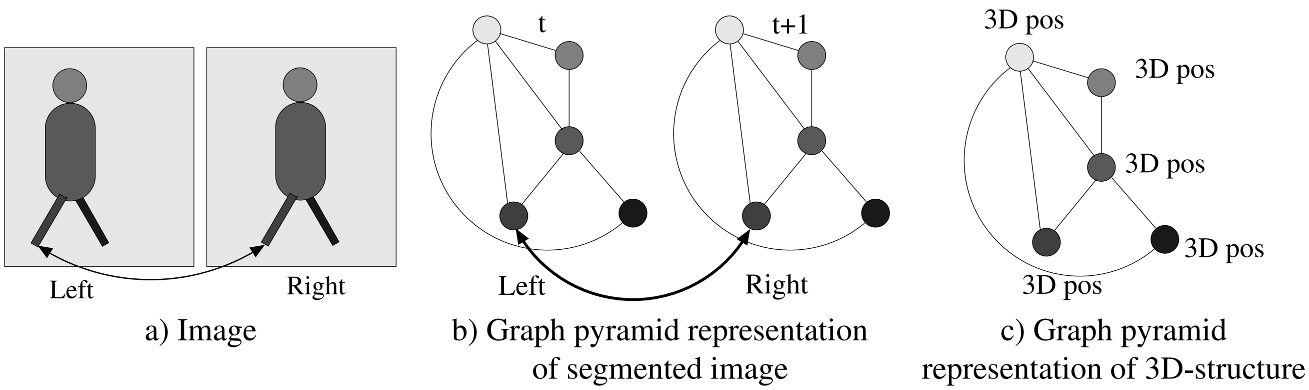
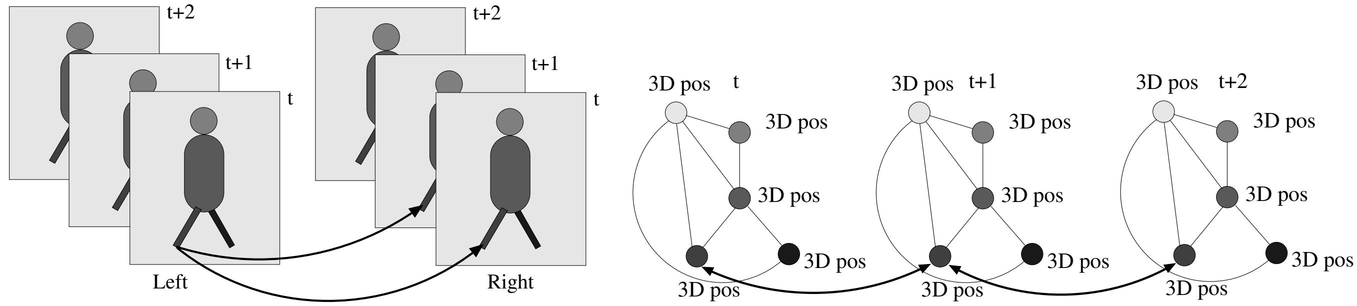


Figure 2: Representation of an image sequence (a) as graph pyramid (b) of the 2D image structure and as graph pyramid of the (c) 3D scene structure

3. Finding object correspondences in image sequences from different view points (again in stereo configurations): Given are two (or more) image sequences taken from different viewpoints. From the two (or more) images of a single time instant a structural representation (again e.g. a graph pyramid) is derived (see Fig. 3). This representation includes 3D information and is therefore able to represent the scene structure in a new way. The sequence of representations of the 3D structure is now used for establishing time correspondences within a scene. As already described in the solution of the first subgoal, object detection and tracking can be achieved robustly.

The combination of these approaches into a single framework, which has not yet been done, would simplify the solutions of many practical problems. We expect these methods to perform better, especially



a) Image b) Graph pyramid representation of 3D structure

Figure 3: (a) Representation of an stereo image sequence as sequence of graph pyramids of the (b) 3D scene structure

in terms of robustness and speed. Possible limits of such methods and corresponding questions that need to be answered are:

- What kind of influences have to be considered in case of drastic scene changes (i.e. object changes, illumination changes etc.)?
- How can an update of the scene representation (object model) be accomplished?
- Are the existing segmentation and graph matching methods under the assumption of structural search fast enough to allow real time performance?
- How can coarse-to-fine approaches be realized (e.g. using subsegmentations)?

We think that due to joint effort of PRIP and the ACV in such a basic research project practical problems would highlight the drawbacks of existing methods developed in theory and would enable to overcome these drawbacks.

The two main partners would share their expertise and produce:

- The state of the art of the tracking methods. Actual *implementation* of the best current methods is required to understand the problems inherent to real world conditions. These problems and hints should help to produce more acurate models and methods.
- A classification of objects and image sequences. Often, methods developed in basic research do not solve particular problems. It is due to a lack of classification of problems and data hypothesis, which should really be produced. It is required not only by the practitioneers but also by the researchers who want to understand quickly the current unsolved problems.
- Based on the two preceeding points, there is a gap on the methodology for measuring performances and comparing tracking methods. Usually, instead of referring to a common benchmark, researchers propose in their publications ad hoc measurements of performances, which are not

always relevant to understand the limits of the methods. Developing a *methodology* for measuring and comparing performance is a very important topic which has never properly adressed by researchers, who are often more interested in producing a new method than understanding what the limitations of the existing ones are.

- New methods based on structural object analysis. Structure has never, or seldom been used by researchers for tracking objects. This is astonishing as structure is the most important invariant.

We emphasize that the goal of this project is scientific in nature. Innovative advances made during work funded by this project will be published and will not be subject to any non-disclosure agreements.

1.3 Status of Research

The first step in our proposed approach is the representation of the image as a graph (or combinatorial map[1]). The most common way of doing this is to first segment the image and then build a region adjacency graph, although there are approaches which use graphs earlier in the segmentation process. This is discussed in Section 1.3.1.

The fact that we have motion information can already be used at this early stage to improve the segmentation. Sequences can be segmented using spatial information and motion cues. This is discussed in section 1.3.2. In the next step we make use of the richer information representation embodied by the graphs to find correspondences between subsequent images in a sequence and thereby to track objects. In Section 1.3.3, a general background to tracking in image sequences is given. The use of graphs in this task is detailed in Section 1.3.4. As graph matching is an important component of graph-based tracking, we review the current state of the art in Section 1.3.5. Finally, in Section 1.3.6, the contribution of the project partners to the current state-of-the-art is highlighted.

1.3.1 Building graphs from images

The output of any segmentation algorithm which produces regions having closed boundaries (e.g. the watershed algorithm) can be represented as a region adjacency graph (RAG) [84]. In this graph, each region is represented by a node. The edges between the nodes represent the region adjacencies. In addition, one can build an attributed graph (AG) by storing features on the nodes and edges. Features stored in the nodes can represent region characteristics (colour, texture, etc.), and features stored on the edges can represent differences in region characteristics. Stable image features, such as SIFT [26] or MSER [59] features, can also be taken to be the nodes of the graph. A problem is, however, determining

[1] Combinatorial maps are an efficient representation of graph-like structures [17].

spatial relationships between these features which lead to a useful set of edges connecting the nodes. With a segmentation, these relationships are available based on the neighbourhood relationships of the resulting regions.

Alternative approaches begin with a representation of the image as a graph, where each pixel is a node and the edges represent the "strength" with which these pixels belong to the same group, such as similarities in colour, brightness, etc. Pixels can then be grouped by making use of a Minimal Spanning Tree (MST) [31, 42], a minimum cut [94, 79, 35] or the complete linkage clustering algorithm, which reduces to a search for a complete subgraph i.e. the maximal clique [65]. The use of Markov Random Fields (MRF) has been proposed for image restoration and segmentation [38] and usually leads to NP-hard problems. The graph-based approximation method for MRF problems [16] yields a practical solution if the number of labels for the pixels is small, which limits these methods for use in segmentation.

Hierarchical structures for description of the data for segmentation purposes have been studied very early in [46]. Bister et al. [11] conclude that regular image pyramids are unsuitable as general-purpose segmentation algorithms. In [63, 47] it was shown how these drawbacks can be avoided by irregular image pyramids, the so called adaptive pyramids, where the hierarchical structure (vertical network) of the pyramid was not "a priori" known but recursively built based on the data. Meer [60] in his "consensus vison" used the concept of an irregular pyramid to produce an image segmentation. Moreover in [22, 15, 44] it was shown that irregular graph pyramids can be used for segmentation and feature detection.

1.3.2 Motion-based segmentation using graphs

Combination of motion measurement with image segmentation can result in better analysis of motion. The knowledge of spatial partition can improve the reliability of motion-based segmentation [70, 37]. Temporal tracking of a spatial partition of an image, from the motion-based segmentation, is easily done than if spatial regions are tracked individually [37]. Motion-based segmentation leads to a semantic description of the image, involving fewer and often more significant regions than a spatial segmentation. In several approaches intensity is involved at pixel level through a spatial segmentation, providing a set of regions that are handled by a region-level motion based scheme. In [27, 93] a spatial segmentation is followed by a motion-based region-merging. Given the partition at the current iteration, the adjacency graph is built and labeled on a spatial criterion, using stochastic dynamics and exploiting the desired connectivity of regions to reduce the space to be searched. The labeled graph provides an initial partition for the next iteration. Another possibility is to introduce both spatial and motion information both at pixel level. In [12] both types of constraints, along with geometrical ones are included in the same energy function in a Markovian-Bayesian scheme. Occlusion and crossings have been treated in [62], considering only a small number of regions and by tracking them independently.

Different image pyramid approaches have been used for dynamic images. Image pyramids for motion detection based on correlative measures between images is proposed in [20, 81]. They use the hierarchical algorithm to find the best estimate of the motion parameter by maximizing the correlation between images in the sequence. Pyramids are also applied for stereo matching in [83]. A spatio-temporal Laplacian pyramid has been used to decompose an image sequence and separate moving objects by their sizes and velocities in [6, 30]. In [1, 29] the authors study the spatiotemporal energy model for perception of motion. Pyramid-based approaches for motion estimation and spatiotemporal segmentation [55, 56] use the structure in a coarse to fine way. The motion is estimated by using classic 2D spatio temporal segmentation procedure at coarse levels, and the results are propagated to higher resolution levels by correcting or predicting errors.

Image pyramids have also been applied to measuring the optical flow [56], which is the basic method for recovering scene information using a moving camera (called *structure from motion*). Kropatsch [51] introduces an idea how one can explain *motion from structure* using the relation of a moving object w.r.t. the environment (background) by using the irregular graph pyramid. In [69] a regular graph pyramid and in [88] an irregular graph pyramid approach for spatio-temporal segmentation and motion estimation are used, by interlinking pyramids over consecutive frames, in order to keep relationship between regions. These methods could also track regions by following the interlinks between pyramids.

1.3.3 Tracking on image sequences

The central challenge in visual tracking systems is the determination of a single target (object, surface, contour, point) over a sequence of images or data sets. Tracking implies the necessity to cope with the variability of the object in terms of variation of pose or shape, variation of illumination and partial or full occlusion of the target. Tracking methods can be classified from an application point of view into

- Tracking with moving or stationary camera
- Tracking inside-out (sensors attached to the tracked object) or outside-in (the sensor observes the tracked object)
- Tracking single or multiple objects
- Tracking particular objects or classes of objects
- Tracking rigid or deformable objects

- Tracking whole objects or object parts
- Tracking indoor or outdoor

From the algorithmic point of view we can classify the algorithm according to the following criteria:

- Area-based or feature-based
- Underlying object model defined by various approaches of object description (e.g. 3D wire-frame model, appearance based)
- Underlying motion model (linear, non-linear, deterministic, probabilistic)
- Applied data association of detected objects to existing tracks (nearest neighbor or probabilistic data association

Usually the first step of tracking is the extraction of region of interests (blobs) which is done by background subtraction algorithm [28, 58]. Other authors use *area-based methods*, which are based on template matching [48] or are appearance based [33]. *Feature-based methods* have as main challenge the task of feature selection, that can be very unconstrained for real world applications.

In the field of tracking humans and/or human body parts, there are several methods applied. In this field there has been a lot of work done as summarized in [2], where a human is modelled as stick-figure model, as coarse volumetric model or as 2D contour model. In [82], a cinematic model is used to decompose cinematic chain structure of the human body. Others use very coarse models like ellipsoids [96]. First studies of detecting and tracking a human head using simple models have also been proposed by ACV [24, 71, 3]. However, markerless 3D tracking of humans still has several challenges such as model acquisition, occlusion, 3D data (as stated in [34]).

1.3.4 Graph-based Tracking

Using matching of graphs of two successive frames allows objects of interest to be followed along the sequence. Graph matching avoids the costly motion estimation and compensation, and moreover the graph representation permits objects (vertices) which are not visible anymore in the scene to be retained in memory (memory graphs [41]) and to be recognized correctly when they reappear. In the literature, matching is usually performed by smoothing the graph dissimilarities by means of a set of editing operations directly applied on graph [19]. In contrast, [41] does not consider differences in graphs as errors but as partition changes due to: a) a new vertex being added when a new object enters the scene, b) when the same object appears in both partitions, it can be fragmented into a different number of regions. This implies that there is no one-to-one mapping between vertices; if both partitions belong to a sequence, moving objects change their structures or even the global number of vertices due to temporal occlusion. The problem is solved in two steps [61]: the finer partitions step by splitting the regions which do not match into finer partitions and matching them, and the coarser step by merging the unmatched regions. The concept of graph matching is extended to partition sequences [41]. The method based on matching should deal with problems of changes in the neighborhood topology and temporal occlusions [41].

The authors in [37, 36] estimate a 2D motion model within each region, after the intial spatial segmentation, and the optimal motion label configuration is sought using an energy minimization approach, such that regions undergoing similar motion are given the same label. This graph, valued by motion information measured on the resulting regions, is the one used for tracking. This method copes with poorly textured areas. Insufficient intensity gradient information is available for the differential motion estimation method to supply accurate motion estimates. Involving region representation, recursive filtering and explicit formalized temporal evolution model, the tracking graph structure is obviously much simpler than spatial region graph, while involving all the useful information [37].

A compact representation of a set (collection) of AGs, called *function-described graphs* (FDGs) [77] have been introduced as an alternative to first order random graphs, but which borrow from random graphs [92] the capability of probabilistic modeling of structural and attributed informations. A similar approach was used by [49] for generic graph-based modelling from examples. A distance measure for matching AGs with FDGs is defined considering costs of the matching of vertices and edges using the edit operation approach [73], and a branch-and-bound algorithm for tolerant error matching in [76].

The approach in [14] groups features and tries to match both features and relations from different frames, similar to [52]. Image features (edge, texture ...) form the vertices of the fuzzy graph. A confidence value is given to each edge of a graph, to represent the strength of a particular relation. The tracking relies on the registration of segments of two subsequent frames by a fuzzy graph matching. In order to reduce the computational costs of graph matching, the authors do an exhaustive search over the space of possible complete pairings, first a low level tracking and ensuing hypothesis fitering using fuzzy relaxation labeling [13] is used. The method is able to cope with problems of appearing and disappearing of features (segments). It is possible to extend this method to track higher level, such as faces and their relations, by creating graphs with vertices representing these features. These graph could be created by lower lying graphs and information from images, yielding to the concept of a ‘pyramid of graphs’.

Another graph-based approach considers motion segmentation as a special instance of the more general grouping problem [80]. Each pixel in the image is treated as point living in a feature space. The features correspond to its spatio-temporal position, color, motion, texture, etc. In order to measure motion similarity, they define a motion feature vector at each pixel, called *motion profile*. Each motion profile is the probability of different displacment of each point in the image, which captures not only the direction of the motion, but also the unceratainty associated with it. To segment a motion sequence, a weighted graph is constructed by taking each pixel as a vertex, and connecting vertices by edges, which are in the spatial-temporal neighborhood of each other. The weights on edges represent the similarity between their motion profiles. Once the weighted graph is constructed, the normalized cut criterion is used to recursively partition the graph. It is shown that normalized cut is a global measure which reflects both the similarity within partitions as well as dissimilarity across the partitions [79]. This criterion is computed by solving a generalized eigenvalue system. Because of the complexity consideration one is limited in using a small spatio-temporal neighborhood which makes the method not able to deal with occlusions.

1.3.5 Graph matching

In general a (sub)graph matching problem is NP-complete. Here we give a short overview of the existing methods. A detailed overview on this topic can be found in [18]. The classical algorithm for graph and subgraph isomorphism detection is the one by Ullman [86]. Methods for error-tolerant graph matching based on the A^* search procedure are studied in [73, 78, 85]. These methods incorporate various heuristics and look-ahead techniques in order to prune the search space. All of these methods guarantee to find the optimal solution, unfortunately with exponential time and space computational cost, because of the NP-completness of the problem. It is possible to find a solution in polynomial time by using sub-optimal (approximate) methods. A wide range of algorithms are used for solving this sub-optimal matching problem, like probabilistic relaxation [50, 23, 91]; or continious optimization methods: Hopfield neural networks [32] and Kohonen maps [95]; genetic search [89, 25]; or Tabu search [90]. All of these method are based on a heuristic optimization function and therefore can easily end in local minima. Alternative approaches are based on eigenvalue decomposion [87, 54], linear programming [5, 21] or specialized work in matching trees in terms of maximum cliques [66, 67]. In [64, 66, 39] graph hierarchies of graphs are used matching. This approach is promising in terms of computation time: higher levels of the pyramids, where there are fewer nodes, can be matched first. These matches can then guide the matching of the lower levels.

1.3.6 Contribution of the project partners to the state-of-the-art

Both partners have made substantial contributions to the current state-of-the-art. These are summarised in this section.

PRIP

PRIP mainly focuses on structural methods, which include image partitioning (segmentation), encoding and construction of hierarchies of irregular partitions, tracking/matching partitions and hierarchies of partitions in image sequences.

The main idea behind current research at PRIP is that most of the information of a single image is summarised by its (topological) structure. The structure of an image can be characterized by the locations of discontinuities in the original image. These discontinuities naturally partition an image into regions of homogeneous properties. The locations of those discontinuities are not only insensitive to illumination change, but also invariant to a certain degree to continuous geometric transforms. For instance, deformable and articulated objects, while having a changing geometry, keep during movement their topological structure.

Several approaches have been used to generate a topological structures of images. At PRIP, recent research lead to the following results:

- A watershed algorithm producing a combinatorial map has been developed [57]. This method leads to a very stable segmentation algorithm that was used succesfully on image sequences provided by ACV.
- Hierarchical segmentations were produced by simple extensions on the watershed extraction algorithms. Another very promising result was produced by using a Minimum Spanning Tree (MST) [45].
- The *Redundancy Pyramid*, which takes the redundancy of structures in multiple images into account, has been used in image segmentation [58] and background subtraction [43].

ACV

One major research area of ACV is surveillance and tracking, which includes detection methods as well as extensions to the standard tracking methods like Kalman tracking, condensation tracking and mean shift tracking.

The main research topics that are currently investigated within ACV are robust multi-camera tracking, mid- and high-level motion interpretation, crowd motion detection, algorithms for moving and active cameras and model-based surveillance. The main applications addressed within the research performed at ACV are people tracking, security and safety, automotive safety and man-machine interface. The feedback from the field of practical applications inspires new fundamental approaches for innovative scientific methods.

Several approaches within the field of surveillance and tracking have been researched. Recent research led to the following results

- Framework for evaluating the performance of tracking algorithms [74, 10, 75].
- Algorithms for occlusion handling and group separation [8, 7].
- Methods for combining stereo and monocular information [4]. The third project aim of the proposed project with the strong integration of structure - shall offer an alternative to these methods.

1.4 Collaboration

The cooperation between ACV and PRIP will be based on the well established communication channels that exist due to the close cooperation within kPlus. The overall lead of the project will be Prof. Kropatsch. The WP will be split between ACV and PRIP as follows:

PRIP has currently the deeper experience on the graph methods. Therefore the WP1, which aims at developing the basic methods will be lead by PRIP and WP 1.1 - WP 1.4 will be done by PRIP with minor contributions of ACV (consulting on data formats, image classes etc). WP 1.5 will be mainly done by ACV developing different object detection methods. WP 2 will be done in common, where ACV provides and further researches the image descriptors like MSERs, SIFT etc. and PRIP provides and researches the structural aspects and the representations issues. Additionally ACV will lead the development of the framework concept as it can be based on already existing concepts. WP 3 will be lead by ACV, because ACV has the better access to real world data. However, especially in WP 3.1 and WP 3.5 contributions from PRIP are also planned. WP.4 (documentation) is performed in common.

The actual collaboration should be very productive as all required parts are present. The expected high level scientific research can be validated both on a theoretical basis and on practical one. Practical and concrete problems throw light on the drawbacks of existing theory.

It is planned that one Postdoc will work at each of the participating organisations. Thus it is planned to organise regular meetings to efficiently organize the project. This will further strengthen the ties between PRIP and ACV.

1.5 Methodology

The project is structured to attain the three fundamental research aims listed in Section 1.2. To properly valorise the theoretical developments, we will implement a rigorous benchmarking and testing of the algorithms developed. The work is divided into four sections: the first focuses on advancing the theoretical knowledge and developing novel algorithms in the field of structured representation and graph manipulation. The second incorporates this into a structured tracking and stereo framework. The third section focuses on rigorously evaluating, benchmarking and optimizing the developed algorithms. The last part includes documentation and publications. During the whole duration of the project, work and results will be published in relevant journals and presented at appropriate workshops and conferences.

1.5.1 Theoretical innovation

We will use the existing expertise on structural computer vision at PRIP as a basis on which to build the theoretical developments envisaged.

The framework that we will develop shall incorporate modules using structural computer vision for tracking, stereo and 3D tracking. Those modules include segmentation methods, graph pyramid and combinatorial map representations and graph matching methods.

The proposed project shall use a model-based approach, where the object can be described by a structured representation (graph pyramid or combinatorial map pyramid). This representation will evolve during the tracking process and will start from the structured representation of the segmentation of the tracked object. This should allow an integration of the segmentation into the tracking process, a problem which is not solved yet as most of the newer approaches also rely on proper background subtraction [53].

It may seem strange that we are proposing to use graph matching for tracking, as it is well known that general graph matching is a difficult problem. It can take rather a long time to determine a graph isomorphism or subgraph isomorphism for large graphs. It has however recently been shown that the use of hierarchical structures can speed up graph matching — one can start by matching a few vertices near the top of the hierarchy and then work one's way toward the bottom, or combine different levels of the hierarchy to force a close match. This has been demonstrated in a realtime implementation of graph matching for the tracking of a person in a video telephony application [41, 40]. Note that this was done for a constrained environment based on the assumption that only the person's head and shoulders would be visible. We intend to develop a more general framework based on this approach.

Another problem that we will consider is the initialisation of the graph-tracking algorithms. The objects to be tracked first have to be located before they can be tracked by the graph-based methods.

Additionally the project shall investigate segmentations that consider 3D information. This could be done both ways (i) using structural methods to enhance stereo correspondence methods and (ii) enhance segmentation results by considering structural correspondence.

To achieve our first and second subgoal “Tracking” and “Stereo” we will incorporate segmentation on which we will build a graph representation followed by matching. For our second subgoal “Stereo” we also aim to develop a new representation in terms of the 3D scene structure. This representation is based on the standard graph pyramid, but will also encode the depth information into the representation. To achieve our third subgoal “3D tracking” we use this new representation of the 3D structure as a basis for tracking. There we use graph matching on this new representation.

1.5.2 Benchmarking

In order to properly evaluate the algorithms developed and compare them rigororously to existing algorithms, test databases labelled with ground truth are required. A number of such databases are available to the public, for example, the database released as part of the EU Caviar project[2], but we will also test the algorithms on various kinds of data arising from real applications that are investigated in the industrial research projects within ACV. These are from the field of (i) surveillance and (ii) man-machine interfaces.

(i) In surveillance applications the system has to track persons in far or close range to the camera. The applications include occupancy detection for airbag deployment within a car, surveillance of public places, person counting and tracking in shopping malls etc. The knowledge about the situation and the camera geometry shall be incorporated in the methods using restrictions and constraints of the used models and geometric constraints. Due to partners from industry, ACV has such data available.

(ii) Man-machine interface: the interaction between the user and a technical system is today not very intuitive. The focus of the application in mind should be a simple interpreter of basic gestures and behaviours. The important issue there is proper segmentation and modelling of the person. This kind of data is easy to produce and will be recorded during the project. Alternatively available databases can be used.

[2] `http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/`

1.6 Cooperations

The PRIP group has many contacts with groups doing research in graph-based image processing. The most important contacts, with whom we already have an ongoing collaboration are:

- University of Poitiers (Prof. Dr. Pascal Lienhardt)
- University of Caen (Prof. Dr. Luc Brun)

PRIP also has international cooperations in the framework of the EU projects MUSCLE and AVITRACK.

ACV has national and international cooperations. The national cooperations are mainly with industry and universities within the kPlus program. International cooperations are within the EU projects SNOW and MUSCLE.

1.7 Workplan

The timeline in Figure 4 shows the intended workplan for the project. The workpackages are described in more detail below:

WP 1 Development and implementation of basic methods This workpackage shall develop and implement the basic tools for graph handling, from generation to manipulation and matching and it shall investigate pyramidal approaches. It is not necessary to implement the basic graph manipulation routines and data structures ourselves. They are already available in the LEDA software[3].

WP.1.1 Graph manipulation methods We will evaluate the basic graph manipulation routines implemented in LEDA. Any needed graph manipulation routines which we require but which are not available in LEDA will be implemented (making use of the data structures of LEDA). Especially graph merging, graph splitting, graph cuts and graph enlargement shall be investigated.

WP.1.2 Graph matching This workpackage shall review the graph matching techniques and implement the most promising ones which are not already included in LEDA. The efficiency of existing methods shall be investigated.

WP.1.3 Graph pyramids This workpackage shall review the graph pyramid techniques and update the current PRIP implementation to the new LEDA library. Especially the results of WPs 1.1 and 1.2 will be generalized to work also for graph pyramids.

[3] `http://www.algorithmic-solutions.com/`

Nr.	WP	Vorgangsname	Dauer
1	WP.1	Development and implementation of basic methods	380 Tage
2	WP.1.1	Graph manipulation methods	3 Monate
3	WP.1.2	Graph matching methods	5 Monate
4	WP.1.3	Graph pyramid methods	5 Monate
5	WP.1.4	Graph initialisation methods	3 Monate
6	WP.1.5	Initialisation of tracking	3 Monate
7	WP.2	Development and implementation of advanced methods	520 Tage
8	WP.2.1	Methods for representing gray/colour images	2 Monate
9	WP.2.2	Method for representing motion images	2 Monate
10	WP.2.3	Combining image representations for temporal structure	6 Monate
11	WP.2.4	Combining image representations for spatial structure	6 Monate
12	WP.2.5	Combining image representations for spatio-temporal structures	6 Monate
13	WP.3	Benchmarking + Optimization	400 Tage
14	WP.3.1	Specification and development of a benchmarking framework	2 Monate
15	WP.3.2	Tracking	4 Monate
16	WP.3.3	Stereo	4 Monate
17	WP.3.4	3D-Tracking	4 Monate
18	WP.3.5	Optimization	6 Monate
19	WP.4	Documentation	776 Tage
20	WP.4.1	Papers	38,8 Monate
21	WP.4.2	Reports	38,8 Monate

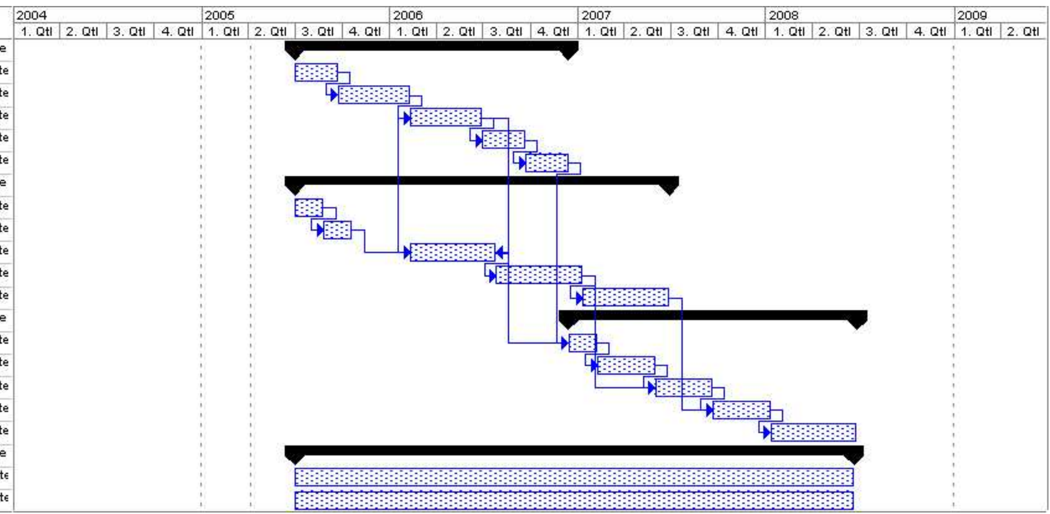


Figure 4: The intended workplan for the project.

WP.1.4 Graph initialisation This workpackage shall review the graph initialisation methods and implement some of them. Graph initialisation methods are those which are used to obtain a graph from an image.

WP.1.5 Initialisation of the tracking In order for the tracking method to function, it is necessary to find objects to track. In this WP, we will examine methods for automatically finding relevants objects to track (people, vehicles, etc. depending on the application). We will attempt to develop, as far as possible, a model-based approach, allowing the characteristics of the object found to be passed directly to a graph-based tracker.

WP 2. Development and implementation of advanced methods This workpackage is the main package of the project. Here the new methods for representing specific types of images will be investigated. The new idea of "structure" in spatial, temporal and spatio-temporal domain within the image shall be researched by combining and successively adapting scene representations within graphs and graph pyramids. In all subworkpackages different types of representations (regions, features, etc.) shall be compared.

WP.2.1 Method for representing gray/colour images This WP shall deal with single images and find efficient methods to represent specific segmentations of images.

WP.2.2 Method for representing motion images This WP shall deal with succesive images within an image sequence and find efficient methods to represent specific motion segmentation in a proper way.

WP.2.3 Combining image representations for temporal structure This WP will combine image representations for temporal structure. It will additionally select the most suitable representation and matching method that allows efficient and robust tracking in different application scenarios.

WP.2.4 Combining image representations for spatial structure This WP will combine image representations for spatial structure. It will additionally select the most suitable representation and matching method that allows efficient and robust solution to the spatial correspondence problem in different application scenarios.

WP.2.5 Combining image representations for spatio-temporal structure Based on the results of WP 2.4 and WP 2.5 the combination of spatial and temporal representations will be used to generate a spatio-temporal representation of the scene.

WP.3 Benchmarking and Optimization In this workpackage the developed methods will be compared to existing methods. Therefore mainly public available benchmarks will be used for comparision of speed, robustness and accurracy of the methods. Additionally new application scenarios will be investigated. Based on the results of the benchmark, optimization work on the methods will be done.

WP.3.1 Specification and development of a benchmarking framework We will review the benchmarking methods which are currently available and what benchmarking data needs to be collected.

WP.3.2 Tracking Benchmarking on tracking. This should be done by enhancing the existing benchmarking framework within ACV and by using public benchmarking sequences available, for example, from PETS[4].

WP 3.3 Stereo Benchmarking on stereo. It is planned to use the benchmarking framework available from Middlebury Stereo Vision page[5].

WP.3.4 3D Tracking Benchmarking on 3D-tracking. It is planned to look for public available benchmarks. If there no benchmarks available, a new framework could be established.

WP.3.5 Optimization Based on the result of WP 3.1 - 3.4 the developed algorithms will be optimized and improved if necessary.

WP.4. Documentation Documentation is an ongoing task. The work shall be documented in technical reports (WP.4.1) and published in refereed conferences and journals (WP.4.2).

2 Financial Aspects

2.1 Available Equipment

The research shall be carried out at Advanced Computer Vision Research (ACV), Vienna, in cooperation with the PRIP (Pattern Recognition and Image Processing Group), Vienna University of Technology.

[4] `http://www.cvg.cs.rdg.ac.uk/PETS2005/`
[5] `http://cat.middlebury.edu/stereo/`

The computing facilities at these institutions are sufficient for an effective research on the proposed project. The facilities include:

1. Computers running Windows and Linux operating systems.
2. State of the art image processing software (MATLAB, KHOROS).
3. Monochrome and colour video cameras and frame grabbers.

2.2 Available Personnel

- The project leader will be o. Univ.-Prof. Dipl.-Ing. Dr. **Walter KROPATSCH**. He is head of the PRIP group at the Vienna University of Technology. He has directed the FWF Joint Research Programme S70 *Theory and Applications of Digital Image Processing and Pattern Recognition* (1994–1999). He was also the head of the project *Robust and Adaptive Methods for Image Understanding* within the S70 research programme, and head of the *Graph Pyramids* (P14445-MAT) and *GeoGraph* (P14662-INF) FWF projects.
- Univ.-Ass. Dr. **Allan HANBURY** joined the PRIP group of the Vienna University of Technology in May 2002, after completing a Ph.D. degree at the Centre of Mathematical Morphology, Paris School of Mines, France. He is head of the FWF project SESAME (P17189-N04) and leader of the Benchmarking workpackage in the EU Network of Excellence MUSCLE (FP6-507752).
- At ACV, Dipl.-Ing. Dr. **Markus CLABIAN** leads projects in surveillance and classification and is responsible for the KPlus program within the second funding period. Results from Kplus-research entered the presented proposal. Continued cooperation with partners and researchers involved (Alefs, Beleznai, Rötzer Schreiber) shall guarantee the scientific value, the fast implementation and the applicability of the developed methods to real world problems. Due to their full integration into Kplus research, their contribution is limited to consultation.

2.3 Requested Personnel

Even with the outlined high personal efforts at ACV and PRIP, the goals of this project cannot be achieved without additional research personal. We therefore apply for two additional research positions at PostDoc level. This is in accordance with the high requirements and large amount of rather sophisticated previous work from different areas (computer vision, image processing, structural pattern recognition, graph matching) on which this project is built.

Two undergraduates will be employed on the basis of research grants (FB). They will support the implementation and the experimental evaluation.

Post	Name	Status	1st year	2nd year	3rd year	Sum
Postdoc (ACV)	Not Known	DV	50.240	50.240	50.240	150.720
PostDoc (PRIP)	Yll Haxhimusa	DV	50.240	50.240	50.240	150.720
Undergraduate	Not known	FB	5.280	5.280	5.280	15.840
Undergraduate	Not Known	FB	5.280	5.280	5.280	15.840
Sum			111.040	111.040	111.040	333.120

2.4 Requested Software

One LEDA source code research license[6] with subscription package for use at PRIP: €9.000 (VAT included). Buying this software will avoid having to program all the basic graph algorithms and data structures, which would require 4–6 person-months. This software can then be provided for all project participants.

2.5 Requested Equipment

We require a high-performance workstation, which will enable us to efficiently run the developed algorithms on large test databases. Processing large videos is a notoriously time-consuming process, and the use of this high-performance workstation will enable us to benchmark the developed algorithms on large test datasets in a reasonable time. The workstation will be installed at PRIP, but ACV will be granted full remote access to it.

The suggested workstation is a Dell Precision 670MT Dual Xeon 3,6GHz System, described in detail in the attached quotation. The cost is €19.542 (including VAT).

2.6 Additional Information

This proposal has not been submitted to any other funding authority.

2.7 Travel Costs

We request €14.000 for travel costs allowing the results of this research to be presented at high-level national and international conferences.

[6]See `http://www.algorithmic-solutions.com/enleda.htm` for details and latest prices.

References

[1] E. H. Adelson and J. Bergen. Spatiotemporal Energy Models for the Percetion of Motion. *J. Opt. Soc. Amer. A*, 2, 1985.

[2] J. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73:428–440, 1999.

[3] B. Alefs, M. Clabian, H. Bischof, and W. Kropatsch. Robust head detection based on feature grouping in depth slices. In *Proc. 27th Workshop Austrian Assoc. Pattern Recognition*, pages 275–280, 2003.

[4] B. Alefs, M. Clabian, H. Bischof, W. Kropatsch, and F. Khairallah. Robust occupancy detection from stereo images. In *Proc. of Int. IEEE Conf. on Intelligent Transportation Systems*, 2004.

[5] H. Almohamed. A Linear Programming Approach for the Weighted Graph Matching Problem. *IEEE Trans. on PAMI*, 15:522–525, 1993.

[6] C. H. Anderson, P. J. Burt, and G. S. van der Wal. Change detection and tracking using pyramid transform techniques. *Intelligent Robots and Computer Vision*, SPIE 579:72–78, 1985.

[7] C. Beleznai, B. Frühstück, and H. Bischof. Tracking multiple humans using fast mean shift mode seeking. In *Proc. of IEEE Workshop on Applications of Computer vision*, pages 25–32, 2005.

[8] C. Beleznai, B. Frühstück, H. Bischof, and W. Kropatsch. Detecting humans in groups using a fast mean shift prodedure. In *Proc. Workshop of the ÖAGM/AAPR*, pages 71–78, 2004.

[9] C. Beleznai, B. Frühstück, and H. Bischof. Human detection in groups using a fast mean shift procedure. In *Proc. Int. Conf. on Image Processing*, 2004.

[10] C. Beleznai, T. Schlögl, H. Ramoser, M. Winter, H. Bischof, and W. Kropatsch. Quantitative evaluation of motion detection algorithms for surveillance applications. In *Proc. Workshop of the ÖAGM/AAPR*, pages 205–212, 2003.

[11] M. Bister, J. Cornelis, and A. Rosenfeld. A critical view of pyramid segmentation algorithms. *Pattern Recognition Letters*, 11(9):605–617, 1990.

[12] M. Black. Combining Intenisty and Motion for Incremental Segmentation and Tracking Over Long Image Sequence. In *Proc. 2nd European Conference on Computer Vision*, pages 485–493, 1992.

[13] H. Borotschnig, A. Pinz, and I. Bloch. Fuzzy Relaxation Labeling Reconsidered. In *Proceedings IEEE World Congress On Computational Intelligence, FUZZ-IEEE 1998*, pages 1417–1423, 1998.

[14] H. Borotschnig, A. Pinz, and D. Sinclair. Fuzzy Graph Tracking. In *Proceedings 5th Symposium for Intelligent Robotics Systems*, pages 91–101, 1997.

[15] M. Borowy and J. Jolion. A pyramidal framework for fast feature detection. In *Proc. of 4th Int. Workshop on Parellel Image Analysis*, pages 193–202, 1995.

[16] O. Boykov, Y. Veskler and R. Zabih. Markov Random Fields with Efficient Approximations. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 648–655, 1998.

[17] L. Brun and W. G. Kropatsch. Introduction to Combinatorial Pyramids. In *Digital and Image Geometry*, pages 108–128, 2001.

[18] H. Bunke. Recent developments in graph matching. In *Proceedings of the ICPR2000*, volume 2, pages 117–124, 2000.

[19] H. Bunke and G. Allerman. Inexact Graph Matching for Structural Pattern Recognition. *Pattern Recognition Letters*, 1(4):245–253, 1983.

[20] P. J. Burt, C. Yen, and X. Xu. Multiresolution flow-through motion analysis. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 246–252, Washington, DC, 1983.

[21] C. S. C. Schellewald. Subgraph matching with semidefinite programming. In *Electronic Notes in Discrete Mathematics*, volume 12. Elsevier Science Publishers, 2003.

[22] K. Cho and P. Meer. Image Segmentation from Consensus Information. *CVGIP: Image Understanding*, 68 (1):72–89, 1997.

[23] W. Christmas, J. Kittler, and M. Petrou. Structural Matching in Computer Vision Using Probabilistic Relaxation. *IEEE Trans. on PAMI*, 17:749–764, 1995.

[24] M. Clabian, H. Rötzer, H. Bischof, and W. Kropatsch. Head detection and localization from sparse 3d data. In *Proceedings of the DAGM 2002*, pages 395–402. Springer, 2002.

[25] A. Cross, R. Wilson, and E. Hancock. Inexact graph matching using genetic search. *Pattern Recognition*, 30(6):953–970, 1997.

[26] L. David G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[27] F. Dufaux, F. Moscheni, and A. Lippman. A Spatio-temporal Segmentation Based on Motion and Static Segmentation. In *Proceeding of Second IEEE Internation Conference of Image Processing*, pages 306–309, USA, 1995.

[28] A. Elgammal, D. Harwood, and L. Davis. Nonparametric background model for background subtraction. In *Proc. of the 6th European Conference of Computer Vision*, 2000.

[29] R. C. Emerson, J. Bergen, and E. Andelson. Directionally Selective Complex Cells and the Computation of Motion Energy in Cat Visual Cortex. *Vision Research*, 32:203–218, 1992.

[30] T. Excoffier and J. M. Jolion. Spatio-Temporal Merging of Image Sequences. In *Proc. 11th Int. Conf. on MICAD*, pages 219–229, 1992.

[31] P. F. Felzenszwalb and D. P. Huttenlocher. Image Segmentation Using Local Variation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 98–104, 1998.

[32] J. Feng, M. Laumy, and M. Dhome. Inexact Matching Using Neural Networks. *Pattern Recognition in practice IV: Multiple paradigm, comparative studies and hybrid systems*, pages 177–184, 1994.

[33] C. S. Fuh and P. Maragos. Motion displacement estimation using an affine model for image matching. *Optical Engineering*, 30:881–887, 1991.

[34] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73:82–88, 1999.

[35] Y. Gdalyahu, D. Weinshall, and M. Werman. Self-Organization in Vision: Stochastic Clustering for Image Segmentation, Perceptual Grouping, and Image Database Organization. *IEEE Trans. on PAMI*, 23(10):pages 1053–1074, 2001.

[36] M. Gelgon and P. Bouthemy. A Region-level Graph Labeling Approach to Motion-based Segmentation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 514–519, 1997.

[37] M. Gelgon and P. Bouthemy. A Region-level Motion-based Graph Representation and Labeling for Tracking a Spatial Image Partition. *Pattern Recognition*, 33(4):725–740, 1999.

[38] S. Geman and D. Geman. Stochastic Relaxation, Gibbs distribution, and the Bayesian Restoration of Images. *IEEE Trans. on PAMI*, 6:721–741, 1984.

[39] R. Glantz, M. Pelillo, and W. G. Kropatsch. Matching Hierarchies of Segmentations. *Int. J. of Art. Intell. and Pattern Recognition*, accepted 2003.

[40] C. Gomila. *Mise en Correspondance de Partitions en vue du Suivi d'Objets*. PhD thesis, CMM, Ecole des Mines de Paris, 2001.

[41] C. Gomila and F. Meyer. Tracking Objects by Graph Matching of Image Partition Sequences. In *In Proceedings of the 3rd IAPR-TC15 Workshop on Graph-based Representation in Pattern Recognition*, pages 1–11, 2001.

[42] L. Guigues, L. M. Herve, and J.-P. Cocquerez. The Hierarchy of the Cocoons of a Graph and its Application to Image Segmentation. *Pattern Recognition Letters*, 24(8):pages 1059–1066, 2003.

[43] A. Hanbury, J. Marchadier, and W. G. Kropatsch. The redundancy pyramid and its application to image segmentation. In *Proceedings of the AAPR Conference 2004*, pages 157–164, 2004.

[44] Y. Haxhimusa and W. G. Kropatsch. Hierarchy of Partitions with Dual Graph Contraction. In *Proceedings of 25th DAGM Symposium*, pages 338–345, 2003.

[45] Y. Haxhimusa and W. G. Kropatsch. Segmentation Graph Hierarchies. In *Proceedings of Joint International Workshops on Structural, Syntactic, and Statistical Pattern Recognition S+SSPR 2004*, pages 343–351, 2004.

[46] S. Horowitz and T. Pavlidis. Picture Segmentation by a Tree Traversal Algorithm. *J. Assoc. Compt. Math.*, 2(23):368–388, 1976.

[47] J.-M. Jolion and A. Montanvert. The adaptive pyramid, a framework for 2D image analysis. *CVGIP: Image Understanding*, 55(3):339–348, 1992.

[48] F. Jurie and M. Dhome. Hyperplane approximation for template matching. *IEEE Trans. on PAMI*, 24(7), 2002.

[49] Y. Keselman and S. Dickinson. Generic model abstraction from examples. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 856–863, 2001.

[50] J. Kittler, W. Christmas, and M. Petrou. Probabilistic Relaxation for Matching Problems in Machine Vision. In *Proceedings 4th International Conference on Computer Vision*, pages 666–674, 1993.

[51] W. G. Kropatsch. How Useful is Structure in Motion? In D. Chetverikov and T. Szirányi, editors, *Fundamental Structural Properties in Image and Pattern Analysis*. Österr. Arbeitsgemeinschaft für Mustererkennung, 1999.

[52] S. Li. Matching: Invariant to Translation, Rotations and Scale Changes. *Pattern Recognition*, 25(2):583–594, 1992.

[53] J. P. Luck, C. Debrunner, W. Hoff, H. Quiang, and D. E. Small. Development and analysis of a real-time human motion tracking system. In *Proc. 6th IEEE Workshop on Applications of Computer Vision (WACV'02)*, 2002.

[54] B. Luo and E. Hancock. Structural graph matching using EM algorithm and singular value decomposition. *IEEE Trans. on PAMI*, 23(10):1120–1136, 2001.

[55] F. Luthon, A. Caplier, and M. Lievin. Spatiotemporal MRF Approach to Video Segmentation: Application to Motion Detection and Lip Segmentation. *Signal Processing*, 76:61–80, 1999.

[56] M. Mahzoun, J. Kim, S. Sauazaki, and K. Tamura. A Scaled Multi-grid Optical Flow Algorithm Based on the least RMS Error Between Real and Estimated Second Images. *Pattern Recognition*, 32:657–670, 1999.

[57] J. Marchadier, W. G. Kropatsch, and A. Hanbury. Homotopic transformations of combinatorial maps. In *Proceedings of the 11th International Conference on Discrete Geometry for Computer Imagery, DGCI 2003*, pages 134–146, 2003.

[58] J. Marchadier, W. G. Kropatsch, and A. Hanbury. The redundancy pyramid and its application to segmentation on an image sequence. In *Proceedings of the DAGM 2004*, pages 432–439, 2004.

[59] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, volume 1, pages 384–393, 2002.

[60] P. Meer, D. Mintz, A. Montanvert, and A. Rosenfeld. Consensus vision. In *AAAI-90 Workshop on Qualitative Vision*, pages 111–115, 1990.

[61] F. Meyer. Graph based morphological segmentation. In *2nd IAPR-TC-15 Workshop on Graph-based Representation*, pages 51–60, 1999.

[62] F. Meyer and P. Bouthemy. Region-based Tracking Using Affine Motion Models in Long Image Sequences. In *Proceedings of CVGIP: Image Understanding*, volume 60(2), pages 119–140, 1994.

[63] A. Montanvert, P. Meer, and A. Rosenfeld. Hierarchical image analysis using irregular tesselations. *IEEE Trans. on PAMI*, 13(4):307–316, 1991.

[64] J.-G. Pailloncy, W. G. Kropatsch, and J.-M. Jolion. Object Matching on Irregular Pyramid. In *14th International Conference on Pattern Recognition*, volume II, pages 1721–1723. IEEE Comp.Soc., 1998.

[65] M. Pavan and M. Pelillo. Dominiant Sets and Hierarchical Clustering. In *ICCV03*, 2003.

[66] M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching Hierarchical Structures Using Association Graphs. *IEEE Trans. on PAMI*, 21(11):1105–1120, 1999.

[67] M. Pelillo, K. Siddiqi, and S. W. Zucker. Many-to-many matching of attributed trees using association graphs and game dynamics. In *Proc. 4th Int. Workshop Visual Form*, 2001.

[68] H. Ramoser, T. Schlögl, M. Winter, and H. Bischof. Shape-based detection of humans for video surveillance. In *Proc. IEEE Int. Conf. on Image Processing*, 2003.

[69] J. Rodriguez, C. Urdiales, A. Bandera, and F. Sandoval. A multiresolution spatiotemporal motion segmentation technique for video sequences based on pyramidal structures. *Pattern Recognition Letters*, 23:1761–1769, 2002.

[70] M. G. Ross. Exploiting texture-motion duality in optical flow and image segmentation. Master's thesis, Massachusetts Institute of Technology, 2000.

[71] H. Rötzer, I. Choi, M. Clabian, H. Bischof, and W. Kropatsch. Head tracking with a condensation algorithm for close range surveillance applications. In *Proc. 27th Workshop Austrian Assoc. Pattern Recognition*, pages 149–156. 2003.

[72] R. Sablatnig. Increasing flexibility for automatic visual inspection: The general analysis graph. *Machine Vision and Applications*, 12:158–169, 2000.

[73] A. Sanfeliu and K. Fu. Adistance Measure Between Attributed Relational Graphs for Pattern Recognition. *IEEE Trans. on Sytems Man and Cybernet.*, 13:353–362, 1983.

[74] T. Schlögl, B. Wachmann, W. Kropatsch, and H. Bischof. Evaluation of people counting systems. In *Proc. Workshop of the ÖAGM/AAPR*, pages 49–53, 2001.

[75] T. Schlögl, C. Beleznai, M. Winter, and H. Bischof. Performance evaluation metrics for motion detection and tracking. In *Proc. of ICPR*, 2004.

[76] F. Serratosa, R. Alquèrez, and A. Sanfeliu. Efficient Algorithms for Matching Attributed Graphs and Function-described Graphs. In *Proceedings of the ICPR2000*, pages 867–872, 2000.

[77] F. Serratosa, R. Alquèrez, and A. Sanfeliu. Function-described Graphs for Modelling Objects Represented by Sets of Attributed Graphs. *Pattern Recognition*, 37(2):781–798, 2003.

[78] L. Shapiro and R. Haralick. Structural Description and Inexact Matching. *IEEE Trans. on PAMI*, 3:504–519, 1981.

[79] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. In *Proceedings IEEE Conference Computer Vision and Pattern*, pages 731–737, 1997.

[80] J. Shi and J. Malik. Motion Segmentation and Tracking Using Normalized Cuts. In *Proceedings IEEE Conference Computer Vision*, pages 1154–1160, 1998.

[81] S. Song, M. Liao, and J. Qin. Multiresolution Image Motion Detection and Displacement Estimation. *Machine Vision and Application*, 3:17–20, 1990.

[82] Y. Song, X. Feng, and P. Perona. Towards detection of human motion. In *Proc. of IEEE CVPR*, volume 1, pages 810–817, 2000.

[83] K. Tate and Z. Li. Multiresolution Range-guided Stereo Matching. In *Sensor Fusion III: 3D Perception and Recognition*, volume 1383, pages 491–502, 1992.

[84] A. Trémeau and P. Colantoni. Regions adjacency graph applied to color image segmentation. *IEEE Trans. on Image Processing*, 9(4):735–744, 2000.

[85] W. Tsai and K. Fu. Error-correcting Isomophisim of Attributed Relational Graphs for Pattern Analysis. *IEEE Trans. on Sytems Man and Cybernet.*, 13:757–768, 1979.

[86] J. R. Ullmann. An Algorithm for Subgraph Isomorphism. *Journal of the Association for Computing Machinery*, 23(1):31–42, 1976.

[87] S. Umeyama. An Eigendecomposition Approach to Weghted Graph Matching Problem. *IEEE Trans. on PAMI*, 10:695–703, 1998.

[88] G. Valencia, J. Rodriguez, C. Urdiales, A. Bandera, and F. Sandoval. Spatiotemporal video and motion estimation through irregular pyramids. *Pattern Recognition*, 36:1445–1447, 2003.

[89] Y. Wang, K. Fan, and J. Horng. Genetic-based Search for Error-correcting Graph Isomorphism. *IEEE Trans, on PAMI*, 27:588–597, 1997.

[90] M. L. Williams, R. C. Wilson, and E. R. Hancock. Deterministic Search for Relational Graph Matching. *Pattern Recognition*, 32:1255–1271, 1999.

[91] R. C. Wilson and E. R. Hancock. Structural Matching by Discrete Relaxation. *IEEE Trans on PAMI*, 19(6):634–648, 1997.

[92] A. Wong, J. Constant, and M. You. Random Graphs. In *Syntactic and Structural Pattern Recognition: Theory and Applications*, pages 197–234, 1990.

[93] L. Wu, J. Benois-Pineau, P. Delagnes, and D. Barba. Spatio-temporal segmentation of image sequences for object-oriented low bit image coding. *Signal Process. and Image Commun.*, 3(5):625–638, 1996.

[94] Z. Wu and R. Leahy. An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation. *IEEE Trans. on PAMI*, 15(11):1101–1113, 1993.

[95] L. Xu and E. Oja. Improved Simmulated Annealing, Boltzman Machines and Attributed Graph Matching. In L. Almeida, editor, *Lecture Notes in Computer Science*, volume 412, pages 151–161, 1990.

[96] T. Zhao, R. Nevatia, and F. Lv. Segmentation and tracking of multiple humans in complex situations. In *Proceedings of the CVPR*, pages 194–201, 2001.