BRNO
UNIVERSITY
OF TECHNOLOGY

FIT

FACULTY
OF INFORMATION
TECHNOLOGY

Dissertation Thesis Topic

Content Analysis of Distributed Video Surveillance Data for Retrieval and Knowledge Discovery

Supervisor

Doc. Ing. Jaroslav Zendulka, CSc.

2007

Ing. Petr Chmelař

Acknowledgements

My thanks go to all colleagues, friends and my family, especially to J. Zendulka, M. Švéda and A. Ženíšková.

Contents

1 Introduction

A video surveillance activity has dramatically increased over the past five years [Tri05]. It was caused mainly by the escalating amount of surveillance cameras as a request to the terrorist peril. However, all of the image understanding and risk detection is left to human. An automated system for visual event detection and indexing can reduce the burden of continuous concentration on monitoring and increases the effectiveness of information reuse by the security, police, emergency and firemen. Additionally – in contrast to the (relatively) cheap technical equipment, the work of security personnel is very expensive in developed countries.

In the last decade also the machine vision research has (in some aspects) emerged to practical applications running on present computers in real time. There are special applications that can track [CLEAR] and count people in single camera's field of view [Axis] or detect a left luggage [PETS06]. Some other commercially successful applications deal for instance with traffic [Camea], face [COGNI] or other biometrics [L1id] detection, recognition and identification. The output of such applications is a (textual) features' and semantics' annotation in a form of a relational [VACE07] or XML data [MP704], potentially connected to an alarm. An example of such process including some other (prior) knowledge is in Fig. 1.1.

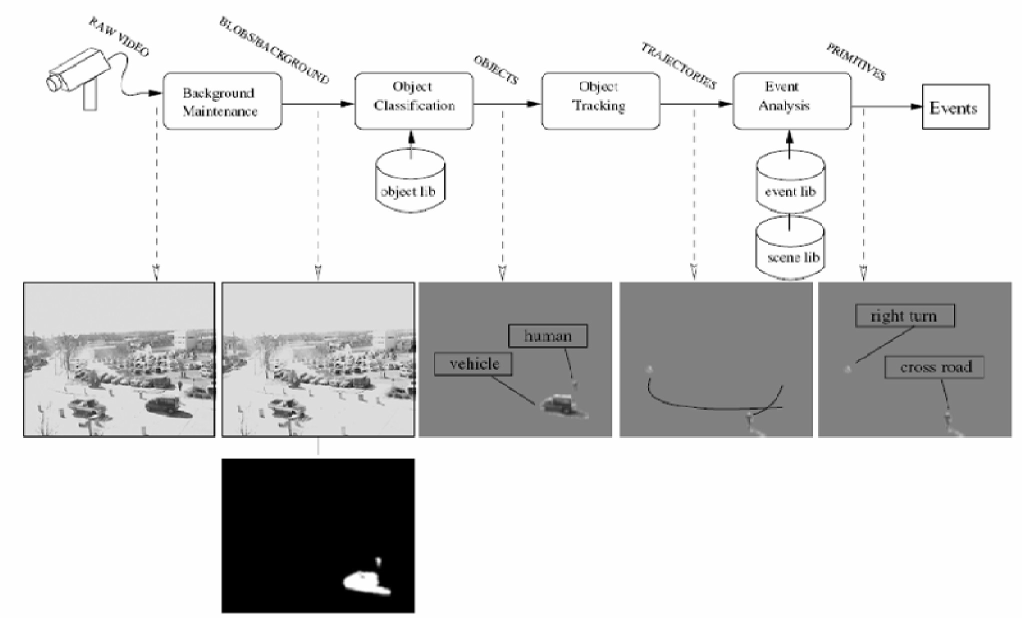

Fig. 1.1 Stages of spatio-temporal segmentation and event analysis [Ers04].

Nevertheless, the security of cities, transportation ports and companies, schools, bureaus, prisons, casinos, office buildings and hotels, banks, sport arenas or shopping centers require an additional operation – it is the retrieval of interesting information from the captured surveillance multimedia data. This is especially important for a (very) fast retrieval of captured records whenever a special event occurs – both to track the person involved in a crime and the analysis of behavior for an automated detection of special events including (false) fire alarms.

Additionally, an application of data mining in databases containing information about objects and their spatio-temporal location might prevent such events. But it is a “Grand Challenge or the X-Prize of data mining” [Pia06] crossing the semantic gap between the multimedia data and semantics. Thus European Union supports projects like is the [CARETAKER] in which the Faculty of Information Technology (FIT), Brno University of Technology participates.

The use of database technology including the data analysis and knowledge mining is usual in (almost) all branches of business, government and science. In contrast to that, the application of database technology together with (higher-level) results of machine vision is quite rare (Fig. 1.1 is a concept). Researchers have made great steps both in machine vision and database technology. Despite there are many articles addressing the problem of object (human) identification, tracking and behavior discovery, it is still a great problem handling real-world surveillance data with serious precision [Tes05]. The (distributed) database has to manipulate a large amount of incomplete, noisy, spatio-temporal (streaming) multimedia data and metadata, as described in the following chapter.

Therefore, the main goal of my dissertation thesis is to identify and fulfill gaps between the content analysis of many concurrent surveillance records, its’ (standard) spatio-temporal description, an efficient retrieval and mining such data. The thesis objectives are discussed in chapter three.

2 Report on State of the Art

The report is divided into three parts. First chapter deals with content analysis of surveillance video records – it is the machine vision part. Chapter 2.2 covers the storing, manipulation and retrieval of multimedia data and related metadata – the database technology part. Third chapter focuses on statistical and nontrivial analysis of extracted data – it is the knowledge discovery part.

2.1 Content analysis

Machine vision basics [Son99], [Jäh99] were discovered in last four centuries. Mathematicians Bayes, dealt with probability expectations, and Gauss established basic mathematics of geometry, optics and (physical) measurements. Both were inspired by Newton's work – the three laws of motion and optical basics.

During the beginnings of information and computer science (30' – 60' of 20th century), many researchers has defined the practical (engineering) base of machine vision – digital (signal) processors, computers and sensors. There were also developed necessary algorithms for universal and signal computation as well as the artificial intelligence – neural networks, support vector machines [Vap63], the Kalman Filter [Kal60] or Hidden Markov Models [Bau70].

Unfortunately, there was not enough powerful hardware necessary for a practical application of machine vision before 70'. The beginning has dealt with optical (color based) segmentation and shape (letter) recognition. In 80' researchers has begun to analyze the texture [Hec05] and motion in video [Hil83]. Research interests have thus migrated from static image based analysis to time-varying picture based dynamic monitoring and analysis [Tri05], which is widely described in chapter 2.1.4.

The computer vision has become widely popular in 90'. Since that time, most of the theoretical principles has been revised [KoV95] and confronted to the reality of required applications and concurrent computational equipment [Cog07]. Most learning methods were criticized to be inadequately precise [Sol06] or computationally too complex (e.g. non-naïve Bayesian regression). For discussion on machine learning and pattern recognition see chapter 2.1.3.

Earlier work on moving pictures analysis dealt mostly with single stationary cameras, but the recent trend is toward distributed multi-camera systems [Ngu03], [Jav03]. Such systems offer several advantages over single camera systems – multiple overlapping views for obtaining 3D information and handling occlusions [Tri05], multiple non-overlapping cameras for covering wide areas, and active pan-tilt-zoom (PTZ) cameras for observing object details. This is discussed in chapters 2.1.2 and in 2.1.5.

The process of analysis surveillance video content is in literature [Pol03], [CeS95], [Tri05] divided into following steps:

- Detection of moving regions and image segmentation.
- Localization of objects.
- Feature extraction features and object recognition.
- Tracking in real world coordinate system.

We concentrate on these topics more in detail, because machine vision has to supply the best possible information about analyzed surveillance data. The structure of the rest of the chapter 2.1 is derived from the above topics.

2.1.1 Preprocessing and Segmentation

The object detection problem can be seen as the problem of subtraction of (potentially interesting) objects or its parts from the background. These objects are segmented [Han06] 2D regions (in computer graphics sometimes called blobs [Tri05]) based on some low-level features as color, texture or motion [Son99]. But there are many problems, usually concerning the segmentation threshold – high causes object's separation, contrary the low threshold causes merging (more) objects together with background.

Many researchers [PETS06], [VSAM00] accomplish the initial task by a common and computationally inexpensive method that generates a background image using several frames of video. The background can be then easily subtracted and objects are sometimes nicely segmented. But there has been a lot of criticism [Neu84] on the background subtraction, especially at Video group at FIT. There are many new approaches how to model the background [Oli00], [Ram07], [PETS06], but it still suffers from environmental changes that cannot be easily removed in the preprocessing step. It can be outdoor lighting, such as shadows or weather changes (even clouds) or in case of the dynamic (moving) textures [Soa01], [CeS95] – trees in wind or waves on the sea are blowing, not talking about escalators or moving pan-tilt-zoom (PTZ) cameras.

Because the optimal algorithm doesn't exist, researchers deal the video segmentation task using 2D motion flow (see chapter 2.1.4) with acceptable, experimentally determined ad-hoc thresholds [Son99], [OCV06]. MPEG group [MP402] set the size of segments to 16x16 pixels, for instance, which is also suitable for the determination of PTZ camera shift [Mat05], if the stationary points are explicitly determined. Some tries to avoid the segmentation step at all using fast classification algorithms like is the WaldBoost [Soch05].

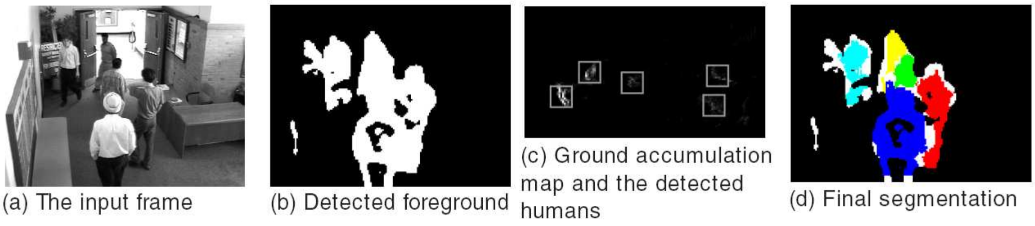
(a) The input frame
(b) Detected foreground
(c) Ground accumulation map and the detected humans
(d) Final segmentation

Fig. 2.1 a) An illustration of preprocessing using stereo vision [Zha05]

Sometimes also shadows cause problems in object's segmentation [Tri05], therefore another branch of computer vision deal with its removal [Fin06].

Another problem follows. Usually, there is a trouble with color and illumination variance [Che06] not only within multiple cameras of distributed surveillance system. However the most bothering problem is the fact that an object (almost always) looks differently in different images and the same one might look very differently, which is escalating in case of non-rigid human bodies, not talking about many occlusions in crowd analysis [And06]. This topic will be discussed further in the remaining of chapter 2.1.

At last but not least, the preprocessing is responsible of image compression [MP402], which is necessary for the long-term record storage, whereas the video (stream) content analysis is done only once and the process requires as large and clear image as possible, without any additional noise and fragments.

2.1.2 Camera Calibration

Camera calibration is a process that allows us to set the parameters in the projective matrix *P* [Son99]. The matrix provides a linear mapping from the image plane coordinates *X* to the 3D (real world) coordinate system *X'*. There’s nothing about science, but clear mathematics. The projective matrix is of the form:

Eq. 2.1

$$\begin{bmatrix} w'x' \\ w'y' \\ w' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix} \begin{bmatrix} wx \\ wy \\ w \end{bmatrix}, \text{ respective } X' = PX$$

Where *w’* and *w* are normalization coefficients, such as meters and pixels. A perspective transform maps lines into lines but only lines parallel to the projection plane remain parallel. A rectangle is mapped into an arbitrary quadrilateral. Therefore, the perspective transform is also referred to as four-point mapping [Jah99], [Son99], by which optimization is the matrix determinated in case of knowing correspondence of at least four points. The system then measures the horizontal locations and heights of people from the silhouettes by triangulation [Tri05]. Some note on the (four) points of real world – they should be hard-mounted so that even when (PTZ) camera moves or shakes, the calibration isn’t affected.

But there is more about the single camera calibration – it is necessary for measuring distances between different cameras. It is extremely important to have calibrated whole system, not to detect one object as multiple with an insufficient probability or more objects as one [Jav03].

The other calibration is the (prior) probability estimation of spatio-temporal occurrence of objects leaving field of view of one camera and after passing a blind area, appearing in the second [Tri05]. It helps much also in the identification problem, as described in chapter 2.1.5.

The last type of calibration is the color and luminance settings – objects vary a lot acquired by various cameras [Che06].

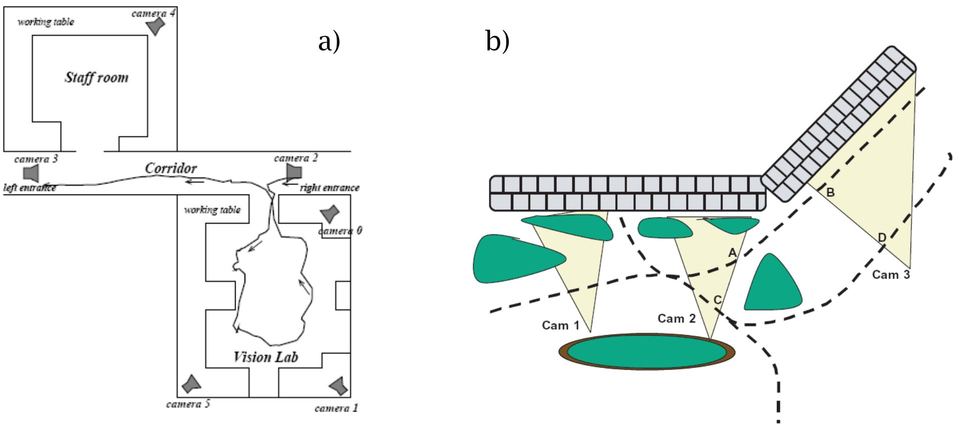

Fig. 2.2 a) An illustration of a trajectory as a result of human tracking in indoor camera system of overlapping and non-overlapping field of view [Ngu03]. b) Camera setup with non-overlapping view field and blind regions [Jav03].

2.1.3 Classification

Classification is a data analysis technique that can be used to extract models describing and distinguishing known data classes to predict even unknown data objects [Han06], [Sch01]. This is commonly used in intelligent decision making or it can provide better data understanding. Whereas classification labels categorical (discrete, unordered) data, regression predicts continuous values. Because classification builds data models based on known data classes, it is not suitable for applications where we can't predetermine learning data set.

In machine learning, vision and pattern recognition, we concern classification of hidden parameters, which is affected by information loss, illumination noise and segmentation errors what can be understood as a noise in a communication channel, as in Fig. 2.3.

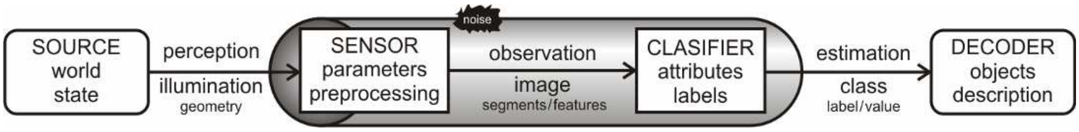

Fig. 2.3 Classification channel in machine vision [Ch06bc].

The visual information and included prior knowledge of the sensed world are the main resources of considered application. The image acquisition and preprocessing is followed by the object recognition and its' pose estimation. The Bayesian common base suits best for the problem illustration.

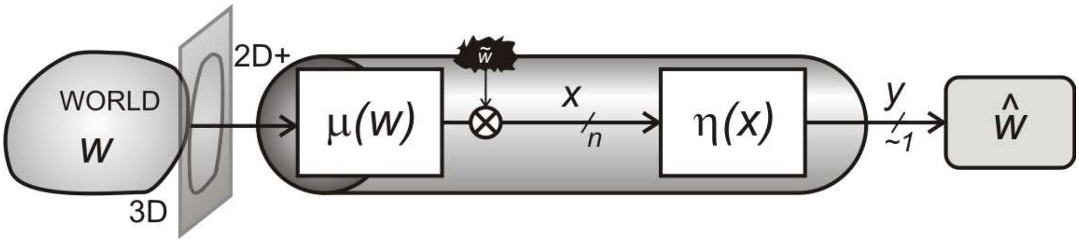

Fig. 2.4 Formal illustration of classification [Ch06bc].

In Fig. 2.4 parameter $w \in W = \{w_1, w_2, \dots w_k\}$ is a hidden state or class that cannot be observed directly, $x \subseteq X = \{x_1, x_2, \dots x_n\}$ are measurements data (information still hidden) we can get about the real world. In machine vision it is usually a result of preprocessing and segmentation of a sensor data, represented by function $\mu(x)$, the encoder. For instance if w is an *apple*, $x = \mu(x)$ could be *(round, red)*. We cannot be sure because the channel is encumbered by a white noise w. Thus in very general it holds: *complete information = observable information + information loss.*

The problem requires a decoder – mapping $\eta(x)$ between X and Y that is often called classification or parameter estimation function. We presume the result of classification y $\in Y = \{y_1, y_2, \dots y_k\}$ is an optimal estimation of class c or parameter θ corresponding to y, informally $y \sim c \sim \theta_c$ of the w. We use Bayes theorem to solve the problem:

Eq. 2.2 $$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

There $P(Y)$ is the prior information having no information about $P(X)$, which is the prior marginal probability, acting as a normalizing constant and can be counted as the sum of all mutually exclusive hypotheses $\sum_X P(X \mid y_i)P(y_i)$. $P(X \mid Y)$ is likelihood or probability (distribution) given by the system or training. Finally $P(Y \mid X)$ is the posterior probability, the conditional probability of Y is derived from X. Within this terminology, the theorem can be rephrased as the normalized likelihood multiplied by prior probability and it provides a method for adjusting degrees of belief of new information. Therefore we determine the highest posterior probability as:

Eq. 2.3 $$\eta^*(x) = \arg\max_x P(y \mid x) = \arg\max_x P(x \mid y)P(y)$$

The optimal decision rule is called Bayesian classifier. It is known as Maximum a posteriori or MAP classifier [Ch06bc]. In case of regression the likelihood is presumed to be the normal Gaussian distribution [Sch01].

In the literature [Sch01], [Han06], [Sol06] are described many classification techniques, such as by back-propagation (neural nets), boosting and decision (tree) induction [Han06] or kernel methods (such as SVM [Chm06mm]).

But what features can be used to the classification? It defines the Moving Picture Experts Group in MPEG-7 [MP704] for instance. There are descriptors of features based on color (scalable histograms, Dominant Colors, or Color Layout), shape (Region Shape, Contour, 3D model) or texture (Homogeneous Texture or Edge Histogram). Texture features are similar to the frequency-based features such as the Fast Fourier Transform, Discrete Cosine Transformation coefficients or (cascades of) Haar features [Son99].

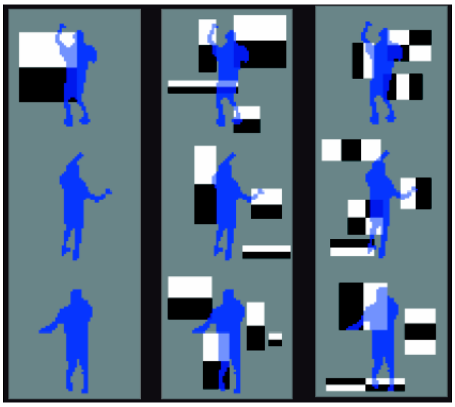

Fig. 2.5 AdaBoost classification of the Haar features of the human shape [Ren05].

On these bases, using image pixels as features are classified rigid objects, as faces are often presumed to be [Cit07]. Pedestrian detection is more difficult, because people can show widely varying appearances when the limbs are in different positions, and are clothes with many different colors and types. Although, classification methods method can learn the high variability in the pedestrian class if there is a good example set. Pedestrians are detected usually using features extracted from their (color) appearance [Jav05], contour [Ren05], [Har99] or rigid body parts [Ram07], [Pol03].

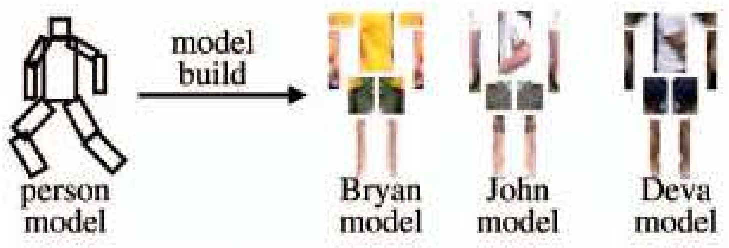

Fig. 2.6 Classification by learning the appearance of human parts [Ram07].

Because of the need of a large amount of training data to be trained to obtain sufficient models for computationally complete functions such as sinus function, there is some criticism on the presented machine learning techniques [Sol06]. Thus new techniques have been currently developed for recursive functions [Sch03]. For all that, we will cope with (experimentally) proved recursive Bayesian and Gaussian approach.

2.1.4 Tracking

Tracking visual targets in real coordinates has emerged as a central problem in surveillance applications. There have been (too) many articles in past three decades on tracking problems and techniques of which detailed description and comparison is provided by Cedras and Shaw [CeS95] who discusses also some ad-hoc human tracking problems. Even [Neu84] Neumann in 1984 summarized the state of the art that it is behind scope of this work. Most techniques deal with 2D tracking of 3D object and its reconstruction. Neumann has been concerning the optical flow algorithms, establishing vectors connecting the same pixels or regions within consecutive images [Son99], or solving the point-distribution references [Mat05]. That is also suitable for moving pictures preprocessing and compression (MPEG video codec [MP402]). Neumann's problem was the lack of experiments and tracking discontinuity, which has been solved in image plane by Horn and Schunck [HoS81], by presuming the smoothness of moving.

The same does also the Kalman Filter [Kal60] in general. It is an optimal [Sor70] and effective (Fast Kalman Filter [OCV06]) prediction-corrective recursive estimator that correct imprecise state *x* (continuous position and velocity) that cannot be observed directly, because it is encumbered by hidden Gaussian noise *w*. The filter produces visible output *y* that is a simple linear observation (first-order Gauss-Markov process), encumbered by noise *v*, as in Fig. 2.7:

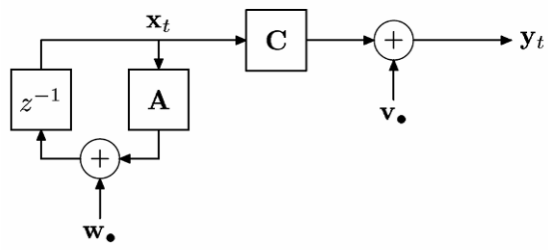

Fig. 2.7 The Kalman Filter process illustration (without the correction input).

In the figure, *A* is a state-transition matrix and *C* is the observation matrix (*x* and *y* are vectors). The Kalman Filter has two distinct phases: Predict and Correct. The predict phase uses the estimate from the previous cycle (z^{-1} is a delay) to produce an estimate of the current state. In the update phase, measurement information from the current state is used to refine this prediction to get a more accurate estimate [Kal60].

The Kalman Filter is analogous to the Hidden Markov Model [Row99], [Bau70] that uses recursively the Bayesian estimation to get the proper observation from discrete hidden state given the previous state [Rab89]. Although there is some criticism on the Kalman Filter, especially to the occlusion handling while tracking multiple objects (as noticed Cedras and Shaw [CeS95]), it is usually the same principle that is proposed to solve the occlusion problem - a HMM modification in [Ngu06] or Monte Carlo Markov Chains in [Kha05].

Promising technique for tracking multiple objects should be also the SwissTrack project [Cor06] based on the Open Computer Vision Library [OCV06]. OpenCV also contains several algorithms for tracking based on background subtraction, optical flow [HoS81] and the Kalman Filter. In case of surveillance human tracking, many approaches on the same base are described in literature such as [Oli00], [Mar03], only some of them deal with tracking body silhouette [Ren05], [Har99] (e.g. not only the bounding box) and few articles are combining tracking multiple people's body parts [Ram07], [Pol03] and their authors claim, that it provides the most accurate results.

The problem of occlusion while tracking multiple objects can be solved using multiple cameras with overlapping field of view – the object tracks the camera in which view field is not occluded [Ngu03], [Tri05]. The matching of objects between overlapping cameras can be solved using homography and calibrated cameras [VSAM00], [Hay05], [Mat05] as in the chapter 2.1.2. In case of sufficient number of cameras (three and more) the system can completely reconstruct the 3D human body [Ren05], [Zim06].

But surveillance systems require a distributed array of cameras that offer wide area monitoring. That implies the use of many non-overlapping fields of view with blind areas, such as the Video Surveillance and Monitoring Distributed Interactive Video Array [Tri05] or the initial work [Cai99], [Ket99], [Orw99], [VSAM00]. Although many researchers address these problems [Jav03], [Ngu03] there is still no robust technique for handover – passing tracked objects between sensors, where the objects temporarily appears.

2.1.5 Identification

The identification problem is central problem of handling uncertain data. Identification maps a known quantity (ID) to an unknown entity (object features), which is somehow similar to the classification (chapter 2.1.3).

The (re)identification is supposed to be the one of the most challenging problems of machine vision [Ngu03], [Zha05], together with behavior analysis and event recognition [Ers04]. Researchers are thus working on the feature extraction for better (likelihood) matching [Jav05], [Ram07], others are investigating tracking algorithms [Ngu06], tracking objects through wide areas and its geometric [Rah04] and probabilistic calibration [Mak04]. Thus follows a simple explanation of the identification problem.

At a certain time, assume we have n objects O_{ki} where $k_i \in N^+$ is the global ID of O_{ki}. We use aspecial ID k_0 to denote a virtual new object. We also have m unresolved tracks T_j. The problem is to compute an optimal global ID $d_i \in N$ for each object and track. Unlike in the case of non-overlapped cameras, the global IDs of different tracks are not exclusive, i.e., it is possible that multiple unresolved tracks have the same global ID. Due to the large number ($(n + 1)^m$) of possibilities of considering the tracks jointly, Zhao in [Zha05] decided to compute the optimal global ID for each track T_j independently as the following:

Eq. 2.4 $$d^* = \arg\max_d P(d \mid O_d, T_j)$$

Bayes theorem is used to solve the equation [Zha05]. The result can match track T_j either to an object O_d or a new object according to its features and the probability of object's track. Once a track is resolved, it is added to the object it belongs to and used to update its features. This arises into two questions: when and how to find such object.

The first question is similar to the Wald's problem [Sch99] because we make late decisions, based on the probability we know the label. This is suitable also in the database three-value logic [Sil02], it replies *NULL* on a queried class that is not satisfied enough precisely (if not forced to). The second question on how to find an object is the same answer – to query objects within database. Hence, the following chapter deals with the surveillance database technology.

2.2 Retrieval

The goal of machine vision systems, depicted in previous chapter, is to find trajectories of people in the scene, as robustly as it is possible. Others, dealing with event recognition, further analyze objects and the trajectory data [Tri05], [Ers04]. That is the area where not only computer vision researchers are interested in, but there are involved also multimedia database technology experts [Kos03], [Fur04]. Usually, researchers from each team don't pass the research area too deeply [Sub98], for instance database ones use the machine vision as a black box providing some (noisy) data [Bil05] or use only some low-level features [Kuch04].

The advantage using database paradigms instead of the computer vision is the ability of information manipulation in a comprehensive way. Additionally, database ensures the consistence, integrity, concurrency, data availability and security, which is necessary in surveillance applications. It is supported by a half century research and practical assessment [Sil02] and a standardization process that has proceeded in the last three decades, represented by the SQL [Eis04] and especially the SQL Multimedia and Application Packages (2nd edition, [WIS07]) at the beginning of 21st century. Although it doesn't define the video querying, it defines the spatial and data mining applications. The temporal operations (part History) will be added soon [WIS07]. This is well suitable for storing and analysis of tracks [Kuc04], objects and events on higher level of abstraction. Contrary the only multimedia part – Still Image is not properly applicable in the context of video surveillance although it provides similarity search based on some low-level features (color distribution, localization and texture) [Kuch04]. The main database vendors however support various kinds of media data, support the SQL/MM standard and extend its functionality [Ora06], [Chu06], [IBM06].

A universal interface of machine vision and database area might be provided by the Multimedia Content Description Interface (or MPEG-7), the ISO/IEC standard developed by MPEG (Moving Picture Experts Group) [MP704]. MPEG-7 is describing the multimedia content and supports interpretation of the information meaning, which can be passed onto or accessed by a device or computer. It contains a set of visual and audio descriptors based on low level (color, shape, frequency) and high level (objects in images, speech) of abstraction, description schemes and related systems. The complete cycle of creation, distribution and consumption of multimedia data (e.g. JPEG, MPEG-4) and related metadata (e.g. MPEG-7) in distributed multimedia (database) system is covered by the upcoming standard called Multimedia Framework or MPEG-21. Its principles, like the unique (multimedia) Digital Item, are described in [MP2102].

The design of database (system) for a visual surveillance, supporting querying, analysis and mining (i.e. not a simple video repository) will have to deal with the following [Fur04], [Kos03]:

- Object modeling including its spatio-temporal location.

- Data querying, browsing and retrieval.
- Storage techniques for streaming multimedia data and related metadata.
- Other design aspects such as physical design, indexing and performance.

The rest of the chapter deals with these topics separately.

2.2.1 Metadata Modeling

Visual surveillance provides huge amounts of two types of data – video records and information about its content. This chapter deals with the metadata part only, for raw data see the chapter 2.2.3. A video record have assigned metadata too – when it was recorded; by which camera and some other (technical character) e.g. the pan-tilt-zoom camera parameters. The surveillance system should have information about camera's location (that's why the calibration is necessary) and some other information (hardware dependent, [Jäh99]).

The surveillance video contains information about physical objects, such as humans, animals, luggage, shopping carts and wheel carts, automobiles and some others. Every object has its label and identity (if known), features and spatio-temporal location. It can be both hard parameters, assigning specific x, y, t value, or somehow flexible, using relative positions, soft margins, windows [Fur04] or not yet computed and other inexact (fuzzy) parameters [Abo03].

Optical object features, like MPEG-7 color, shape, texture, face descriptors [MP704] and other visual appearance models, are discussed in chapter 2.1.3. In addition to the graphical information, the system might handle some other one – the (class) label, identity (if checked in), biometry (face, handprint, fingerprint, iris, and retina), anthropometric measures (height, width) and from other sensors (weight, radiation). Some researchers think they can identify a human also based on their gait [Wan03]. Not only has that posed the problem of handling uncertain data in video surveillance. For instance the result of object classification is a probability density: each class label takes an existential probability and the sum of all probabilities is 1. Moreover also the existence of an object might be uncertain [Dai04].

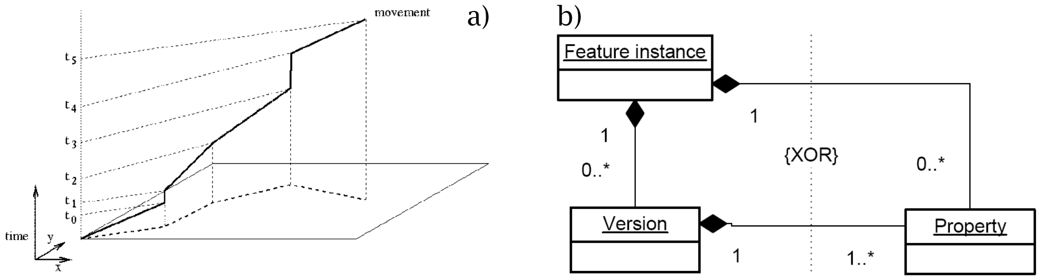

Fig. 2.8 a) An illustration of objects' spatio-temporal location [Bra04]. b) A concept of a temporal substitution [Kuc04].

The value of flexible spatio-temporal location is not always obvious, nor is the representation. The continuous change of (classical) spatial data [She03] in surveillance systems is for most researchers too complex in contrast to the spatio-temporal points and graphs (random sampling) that don't reflect the continuous nature of movement and brings another uncertainty [Che04]. Therefore there is not a standard yet [Kuc04]. Horwath [Hor06] presents an exhaustive survey that might provide the spatio-temporal representation for the understanding and interpretation of activities that take place in a typical wide-area surveillance application.

The MPEG group defines four motion descriptors in MPEG-7 [MP704]. It is the Camera Motion, e.g. for PTZ camera it is - panning (horizontal rotation), tilting (vertical rotation), zooming (change of the focal length). The Motion Activity is the amount of motion (instead of PIR). Parametric Motion is of regions within images, including motion-based segmentation. The Motion Trajectory of an object is a simple, high level feature, defined as the localization, in time and space, of one representative point of this object.

MPEG-7 defines also the Region Locator (a box or polygon in video) and the Spatio Temporal Locator that describes spatio-temporal regions in a video sequence, such as moving object regions, and provides localization functionality. The main application of it is media, which displays the related information when the designated point is inside the object. Another application is object retrieval by checking whether the object has passed particular points. The Spatio Temporal Locator has been added for surveillance purposes [MP704]. The semantics of motion are outlined in following chapters.

2.2.2 Querying

Efficient query formulation, execution, and query optimization for multimedia data is still an open problem [Fur04]. Although principles are well known (such as query by example [Sub98]), the practical realization falls behind. For instance the similarity search compares feature vectors (signatures) that were already extracted from the media, to the vector of a query (example). The relevance depends on their (Euclidean) distance [Han06]. Currently the only similarity search is based on some low-level image features [Ora06]. It is applicable together with keyword-based queries, which has been proven to be completely unsatisfactory, especially in automated surveillance system [Fur04].

Fortunately, we have defined a metadata that are extracted from the raw surveillance video data in advance. It is the (relational) object's description or its alternative in XML from the [MP704] and the spatio-temporal data. In case of following kinds of data we needn't do anything special, see [Eis04] for relational data, [Cat00] for object-oriented, [Eis02], [Fur04] for MPEG-7 XML and [She03], [Kuch04] for spatial data. But the situation is not so straight forward. Presume we put a similarity query necessary for identification. The database cannot process the query in general. Presume the join operation – how to join the track to an object? How the database recognize the most probably (continuous) trajectory? Therefore we (most probably) need a special algebra [Atn01] or more, complex queries [Fur04] handled by special internal procedures or a middle-ware. That's why the retrieval of multimedia data is sometimes called multimedia mining [MDM06].

The task of querying multimedia metadata becomes even more interesting in the moment we want to query temporal data [WIS07], that are not already determined and existentially uncertain (spatial) data [Dai05], spatio-temporal data [Hod04], [Bra04] which is more usual in air traffic control or weather prediction [Fin07]. These spatio-temporal data are uncertain. Although querying this kind of knowledge is still in its beginnings, Cheng claims it is possible [Che04].

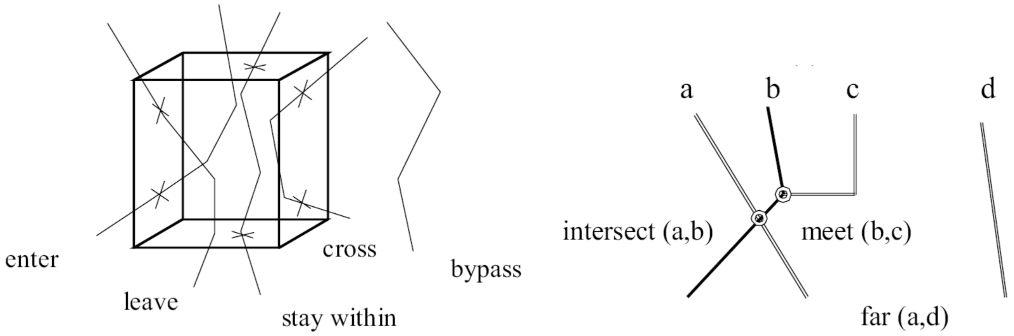

Fig. 2.9 An illustration of objects' spatio-temporal semantics and relations [Bra04].

What can be retrieved, when and what is the result? It can be either (static) metadata extracted before insertion, video records or live streams. There are some query examples [Fur04], [Sub98]. In case of video record it is e.g. `select records where (location tonight) and (luminance > 50) from the video repository`. In case of stream it might be e.g. `select a stream from a camera where a motion (> 10)`. This type of query is called continuous, it might give results continuously and if some motion in surveillance area appears in a future, the query display the new stream on a monitor. This differs a lot from nowadays understanding of querying [Kos03]. Next type is a query on metadata e.g. `select sum(person id) from person where (a person location in today location)` … where the informal `a person location in today location` means the temporal location (within some interval), anywhere in the system. Finally a combined query might look like `select records from repository where meet(a person location, person Cyril location) and (a person location in today location)`. Other queries might ask for a probability of an event or temporal changes of some features.

However, these queries are not yet (without any modification) executable on current database systems [Ora06]. But it can be performed by stored procedures or middleware. Thus researchers deal with query optimization and indexing [Aga00], [Fur04], which is necessary especially in a load of distributed continuous queries [Zho05].

2.2.3 Storage

Data storing is strongly influenced by the data. Due to the duality of surveillance data – video streams/records and the meta information, the system have to deal with both, ideally in a similar way for querying. This is supposed to be hard, although not impossible [Ora06], besides we do it in practice as students' projects in the Advanced Database Systems course at FIT.

On the other side, database experts don't want to embed streaming solutions, so they borrow it from media companies such as [Rea06], which provide repositories with all demands on continuous media creation and retrieval. So there arises a question, whether it is necessary to store and manage under transactional control (compressed) video, in various formats such as MPEG [MP402] or a series of JPEG images [JP200] from multiple cameras that have to be continuously stored and retrieved over (IP-based) networks using multiple protocols [RCo06], or without the transaction control.

Therefore it is more common that the compressed video is stored in a repository [Rea06] and database system handles only pointers to the repository and the metadata [Chu06], [Ora06]. There are only two requirements on such system. It is the standard interface – e.g. database fully controls the repository or there is a multimedia middleware that communicate with both. The second requirement is the random-access memory (or its simulation) for querying. Thus the physical storage on tapes and DVDs as in [Sub98] is suitable only for backup.

I have mentioned pointers in the last paragraph. Of course, pointers in relational databases don't exist, but it has been the spatial data as well as the multimedia that required the object-oriented approach [Cat00], [Chm06mm]. There is hard to find purely relational systems today [Sil02]. Object-relational database systems provide the basic object-oriented operations and data structures [Eis04] that are necessary for handling the multimedia content and its' description – based either on the spatial [She03], spatio-temporal [Kuc04] or MPEG-7 XML data format [Eis02].

2.2.4 Architecture Design

The architecture of database system is greatly influenced by the underlying computer system on which the database system runs. The nature of camera systems is that they have usually several nodes – departments of the state police, city (quarter) police, transportation companies and services or independent companies as banks, sport or shopping centers. The second reason for the use of distributed architecture is the overall performance of the system [Zho05]. It needs a strong parallelism [Spe97] it is simply hard to imagine that one computer can manipulate hundreds to thousands different camera streams [Kos03]. Not talking about disasters such as fire, flood, earthquake or a terrorist attack.

The case study on surveillance systems has been done by Detmold [Det06], although he didn't notice the retrieval, he calls for an integration of building access and elevator control systems and the other sensors such as PIR or smoke, because of the need of scalability, availability, evolvability and integration to a decision-making surveillance system. That's why there's a need for distributed database system handling geographically or administratively distributed data. The distributed architecture however deals with some general aspects that must the proposed system meet. It is the heterogeneity, data fragmentation, replication and especially reliable query and transaction processing due to the required availability of data and overall robustness [Sil02].

Moreover the distributed surveillance system has to deal with special problems such as a transportation of huge amounts of (stream) data created, stored and required continuously on different places and bureaus (e.g. subway and police) in addition to the metadata. These aspects are to be accomplished by the MPEG-21 Multimedia Framework [MP2102], [Kos03], although it doesn't deal with the practical realization.

Fortunately there are (commercially) active areas – voice over ip [RCo06] and the on-line distribution of (entertainment) multimedia content as TV on demand [Rea06]. Database technologies together with the content delivery and communications have emerged in many practical applications including the (real-time) distributed multimedia database systems [Chu06], [Ora06]. For more readings on multimedia databases and related area I offer the study material [Ch06mm] (in Czech). And the problem of a video surveillance system supporting queries is to combine those principles together, so that there will be possible to analyze the retrieved data further.

2.3 Knowledge Discovery

The term knowledge discovery in data means an interactive and iterative process including data preparation, integration, data mining and presentation [Han06]. The main sub process is the data mining, it is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [Han06].

Although it seems, that the term multimedia data mining means something different. Not only Han, also [MDM06] extends the mining by the similarity search or other automated techniques for e.g. event detection, optical character recognition or speech analysis. That's because it uses the same machine learning techniques [Sch01] for classification, segmentation or pattern analysis. I agree, but it is also possible to distinguish [Fuj05]:

- Information retrieval or querying (see chapter 2.2.2).
- Data analysis and statistics, e.g. counting people [Bil05].
- Data mining as the non-trivial process such as finding frequent and rare patterns.

In this viewpoint it is possible to fill in the knowledge discovery in surveillance video data with methods for automatic detection, extraction and report of high interesting objects, patterns, and trends from the visual content [VACE07].

Less formally said, we think of data mining also as an analysis of behavior for an automated detection of desired special events. In [Bod03], [Pap06] are summarized the popular tasks. It can be either mining the (human) trajectory properties of an object as is entering restricted zone or door, moving erratically and loitering, moving against traffic, when (many) people are running or crowds behave somehow differently i.e. breach or bypass some perimeter. The second group is related to human appearance – lying [Pap06], sitting on a floor, waving, or multimodal (such as screaming). And the third group relates to the interaction of humans and others objects e.g. animals, luggage [PETS06], baby carriage and wheel chairs, shopping carts and automobiles. In this group falls also the left or thrown object detection [Pap06] or detection of unknown objects.

Based on my experience (talking to police and firemen), there are some other noticeable events – the peoples' interactions (pickpockets, dealers) [Oli01], detecting vandalism such as tagging (leaving a graffiti), breaking shop windows, lamps dustbins (aggressive behavior) and similar that is more common than terrorist attacks, for which the security personnel have no "pattern definition". The last notes relate to the fire prevention and detection smoke or flames, which is strongly connected to other sensors (smoke) and especially the detection of false alarms, (112) calls or contrary the help with (immediate) determination of its causes for the security personnel, emergency and firemen.

The methods of detection such patterns are not stated clearly. They need either an exact definition of a pattern (thrown object) or a dataset where the pattern occurs to be learned, which is hard (to obtain) as much as the (fully) unsupervised learning provide insufficient results. In brief we can mine following types [Han06] of knowledge:

- Classification of an object (as in the chapter 2.1.3).
- Spatio-temporal pattern or sequence classification using Kalman Filter (2.1.4) or Hidden Markov Models [Bau70] based on decomposition of basic topological features [Ers04], [Pap06] and similarly based on Gaussian Mixture Models [Mar04], [Guh06].
- Characterization (of some class, or event) e.g. for additional learning or for traffic analysis as [Guh06].
- Clustering and outliners using Gaussian Mixture Models [Mar04] (the crowd in [Pap06]) or Principal Component Analysis [Smi02], [Bil05].

Although there are some interesting papers, the practical application of such techniques is rather limited and results are dealing with the problem of the semantic gap between (syntax) low-level features and (semantic) high-level concepts [Pia06].

3 Dissertation Thesis Objectives

The expected intent of my dissertation thesis (Content Analysis of Distributed Video Surveillance Data for Retrieval and Knowledge Discovery) can be summarized as:

- Proposition of a (Kalman Filter) method's modification for detection, identification and tracking many objects (focusing on pedestrians) using distributed camera grid with (usually) non-overlapping field of view.
- Proposition of probabilistic data model for storing and efficient retrieval of objects' description and its' spatiotemporal location.
- Proposition of a technique for analysis of objects and its' behavior.
- A theoretical justification, experimental validation and comparison of the proposed methods will be provided.

A brief explanation of objectives follows.

3.1.1 Content Analysis

There's much to be done in machine vision, but that is behind scope of this work. The only thing I would like to concern is the Kalman Filter or an equivalent optimal technique such as Hidden Markov Models. It has to cope with global models (not only one camera) and produce precise information to the database in both cases of tracking multiple objects by one camera as well as one object tracked by multiple cameras.

The classification is planned to be based on object appearance modeling as a combination of some initial model and the incremental (over) learning similar to the [Ram07], that I have proposed more than a year ago [Ch06is]. The over learning can be simplified to be a statistics of feature vectors obtained over whole period, in which the object appears in the system. This will be added as a free descriptor to the MPEG-7 scheme [MP704] and used together with standard descriptors.

3.1.2 Retrieval

The goal of developing database for retrieval of syntax and semantic-based is an efficient handling of (MPEG-7 XML) metadata provided by the machine vision module. Thus it is profitable to extract some key descriptors (e.g. semantics and the location) and index it in a common way (as a spatial data).

The appearance-based classification should provide late decisions (handling *NULL* values), but suppressed on demand of the database (client), for example if it is necessary to track a person within the system. The idea is, if there's a (112) emergency call, the service get on monitors all unusual and interesting situations the system can discover (using the knowledge discovery). Thus there should be an interruption of continuous monitoring queries for a while.

3.1.3 Knowledge discovery

There is a potential of application of most current trends in mining of general data (transactional, relational) as well as temporal and spatial data or its streams [Han06] that has not been (optimally) applied in the context of video surveillance yet, because of the impossibility of efficient querying.

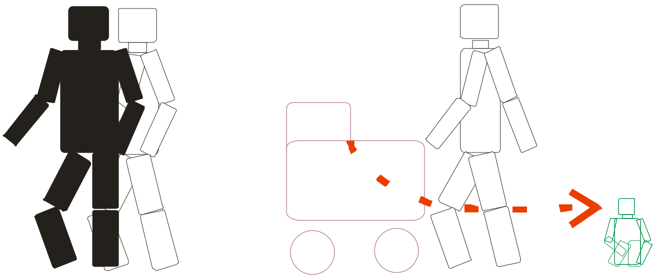

Fig. 3.1 An illustration of “interesting” events that may alert security personnel.

3.1.4 Technical details

First note on the development is the module architecture, because of the need of an arbitrary module change and (re)test for evaluation of proposed techniques, as it is usual at other universities [Pap06] and necessary for efficient publication (generation).

Second is the use of most up-to-date existing technologies instead of its re-invention. It will be done using existing tools like Open Computer Vision Library [OCV06] pro preprocessing and filtering, Yet Another Learning Environmet (YALE) [Mie06] for classification, Oracle Database [ORA06] or PostgreSQL [Pos06] as a database engine, and the MPEG-7 eXperimental Model [MP704] and the Helix DNA Server and Client [Rea07] for multimedia description, compression and delivery. Also techniques developed at FIT might be integrated. The program will be developed as an (GPL) Open-Source middleware for the architecture illustrated in Fig. 3.2.

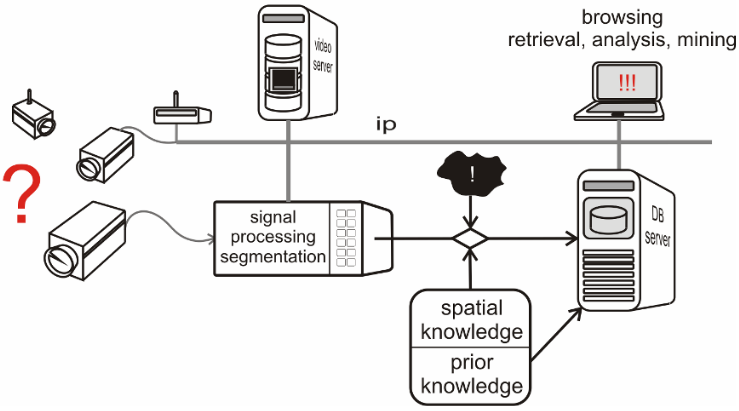

Fig. 3.2 Illustration of the surveillance system architecture.

The last note is on moral [Sew01] and legal [COE07] aspects of proposed system, that will not identify people (and faces) except the volunteers.

References

[Abo03] ABONYI, J. *Fuzzy model identification for control.* 273 p. 2003. ISBN 0-8176-4238-2.

[Aga00] AGARWAL, A.K. - ARGE, L. – ERICSKON, J. Indexing moving points. Proc. Of ACMPODS 2000 conference. 2000.

[And06] ANDRADE, E. L. – BLUNSDEN, S. - FISHER, R. B. Hidden Markov Models for Optical Flow Analysis in Crowds. *18th International Conference on Pattern Recognition.* 2006.

[Atn01] Atnafu, S. - Brunie, L. – Kosch, H. Similarity-Based Algebra for Multimedia Database Systems. *IEEE.* 2001.

[Bau70] Baum, L. E. et at. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164-71, 1970.

[Bil05] BILIOTTI, D. – ANTONINI, G. – THIRAN, J. P. Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people in video sequences. Proc. of the IEEE Workshop on Motion and Video Computing. 2005.

[Bod03] BODOR, R., JACKSON, B., PAPANIKOLOPULOUS N. Vision-Based Human Tracking and Activity Recognition [online]. 2003. [cit. 2006-09-10] http://mha.cs.umn.edu/Papers/.

[Bra04] Brakatsoulas, S. – Pfoser, D. – Tryfona, N. Modeling, Storing and Mining Moving Object Databases. *Proc. of the International Database Engineering and Applications Symposium.* 2004.

[Cai99] CAI, Q. – AGGARWAL, J.K. Tracking human motion in structured environments using a distributed-camera system. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 1999.

[CeS95] CEDRAS, C. – SHAH, M. Motion-Based Recognition: A Survey. *IVC.* March 1995. [cit. 2007-01-16] http://citeseer.ist.psu.edu/cedras95motionbased.html.

[Ch06bc] CHMELAŘ, P. Bayesian Concepts for Human Tracking and Behavior Discovery. *Student EEICT 2006* [online]. VUT Brno, 2006. [cit. 2006-10-10] http://www.feec.vutbr.cz/EEICT/2006/sbornik/03-Doktorske_projekty/07-Informacni_systemy/03-chmelarp.pdf.

[Ch06is] CHMELAŘ, P. Human Tracking Concepts. *Seminar of the Information Systems.* 2006. [cit. 2007-01-19] http://www.fit.vutbr.cz/events/view_event.php.en?id=924.

[Ch06mm] CHMELAŘ, P. *Multimediální databáze* [in Czech, online]. 2006. [cit. 2007-01-10] http://www.fit.vutbr.cz/~chmelarp/pdb/mmdbs%200009.pdf.

[Che04] CHENG, R. – KALASHNIKOV, D. V. – PRABHAKAR, S. Querying imprecise data in moving object environments. *IEEE Trans. On Knowledge and Data Engineering.* 2004.

[Che06] CHENG, E. D. – MADDEN, C. – PICCARDI, M. Mitigating the Effects of Variable Illumination for Tracking across Disjoint Camera Views. *Proc. of the IEEE Int. Conf. on Video and Signal Based Surveillance*. 2006.

[Chu06] Chuckwalla, Inc. *Chuckwalla v5 Overview* [online]. 2006. [cit. 2006-10-19] http://chuckwalla.com/products.asp.

[COE07] Council of Europe. *Data Protection* [online]. 2007. [cit. 2007-01-19] http://www.coe.int/T/E/Legal_affairs/Legal_co-operation/Data_protection/.

[Cog07] Cognex Corporation. *Our History* [online]. 59 p. 2007. [cit. 2007-01-16] http://www.cognex.com/corporate/history.asp.

[Cor06] N. Correll, G. Sempo, Y. Lopez de Meneses, J. Halloy, J.-L. Deneubourg, and A. Martinoli. SwisTrack: A Tracking Tool for Multi-Unit Robotic and Biological Systems. *In IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2006.

[CoV95] Cortes, C. – Vapnik, V. Support-Vector Networks*, Machine Learning*, 20. 1995.

[Dai05] DAI, X. et. al. Probabilistic Spatial Queries on Existentially Uncertain Data. *SSTD 2005*. LNCS. 2005.

[Det06] DETMOLD, H. Et al. Middleware for video surveillance networks. *ACM Proc. of the international workshop on Middleware for sensor networks.* 2006.

[Dev06] DEWAN, M. – HAGER, G. D. Toward Optimal Kernel-based Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* - Volume 1, 2006.

[DTI00] Consumer Affairs Directorate of the UK Department of Trade and Industry. *Childata/Adultdata/Older Adultdata: The Handbook of Child/Adultdata/Older Adultdata Measurements and Capabilities - Data for Design Safety.* 1985/1998/2000.

[Eis02] EISENBERG, Andrew – MELTON, Jim. *SQL/XML is Making Good Progress* [online]. 2002. [cit. 2006-06-19] http://www.sigmod.org/record/issues/0206/standard.pdf.

[Eis04] EISENBERG, Andrew et al. SQL:2003 Has Been Published [online]. 2004. [cit. 2006-06-19]. http://www.sigmod.org/sigmod/record/issues/0403/E.JimAndrew-standard.pdf.

[Ers04] Ersoy, I. – Bunyak, F. – Subramanya, S. R. A Framework for Trajectory Based Visual Event Retrieval. *Proceedings of the IEEE International Conference on Information Technology: Coding and Computing (ITCC'04)*. 2004.

[Fin07] FINKENSTÄDT, B. – HELD, L. – ISHAM, V. Statistical methods for spatio-temporal systems. Chapman & Hall/CRC. 2007. 286 p. ISBN 1-58488-593-9.

[Fuj05] Fujitsu Laboratories. *Multimedia Retrieval / Multimedia Mining [*online]. [cit. 2007-01-10] http://jp.fujitsu.com/group/labs/downloads/en/business/activities/activities-4/fujitsu-labs-bikm-003-en.pdf

[Fur04] FURTH, B. – MARGUES, O. *Handbook of Video Databases : Design and Applications*. CRC Press, 2004. 1211 p. ISBN 0-8493-7006-X.

[Guh06] Guha, P. et al. Surveillance Video Mining. *Proceedings International Conference on Visual Information Engineering*. 2006.

[Han01] HAN, J., KAMBER, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2001. ISBN 1-55860-489-8.

[Har99] HARITAOGLU, I. – HARWOOD, D. – DAVIS, L. S. Hydra: Multiple People Detection and Tracking Using Silhouettes. *10th Int. Conf. on Image Analysis and Processing*. 1999.

[Hay05] HAYET, J. B. et al. A Modular Multi-Camera Framework for Team Sports Tracking. *IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance*. 2006.

[Hin03] HINNER, J. Detekce a rozpoznávání obliceju osob a jejich identifikacni vyznam, *Kriminalistika* [in Czech, cit. 2006-09-10]. 1/2003. Available from: http://www.mvcr.cz/casopisy/kriminalistika/2003/03_01/hinner.htmlhttp://rfc.net/rfc3261.html.

[Hod04] MOKHTAR, H. – SU, J. Universal Trajectory Queries for Moving Object Databases. *2004 IEEE International Conference on Mobile Data Management.* 2004.

[Hop82] HOPFIELD, J. *Neural networks and physical systems with emergent collective computational abilities*. PNAS 79, 2554. 1982.

[Hor06] HORWATH, R. J. Spatial Models for Wide-Area Visual Surveillance: Computational Approaches and Spatial Building-Blocks. *Artificial Intelligence Review*. vol 23, no. 2. 2005.

[HoS81] HORN, B. K. P. – SCHUNCK, B. G., Determining Optical Flow. *Artificial Intelligence*, Vol 17. 1981.

[IBM06] IBM Corporation. DB2 AIV Extenders [online]. 2006. [cit. 2006-10-19] http://www-306.ibm.com/software/data/db2/extenders/aiv/.

[Jäh99] JÄHNE, Bernd – HAUSSECKER, Horst – GEISSLER, Peter. *Handbook of Computer Vision and Applications*. Academic Press, 1999. 3 vol. (624, 942, 894 p.). ISBN 0-12-379770-5.

[Jav03] JAVED Omar et al. Tracking Across Multiple Cameras With Disjoint Views. *Proc. of the Ninth IEEE International Conference on Computer Vision*. 2003.

[Jav05] JAVED, O. – SHAFIQUE, K. – SHAH, M. Appearance Modeling for Tracking in Multiple Non-overlapping Cameras. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005.

[JP200] MARCELLIN, M. W. An Overview of JPEG-2000. *IEEE Data Compression Conference*. 2000.

[Kad02] KADOUS, M.W. *Dynamic Time Warping* [cit. 2006-09-10]. 2002. Available from: http://www.cse.unsw.edu.au/~waleed/phd/html/node38.html.

[Kal60] KALMAN, R. E. New Approach to Linear Filtering and Prediction Problems [online]. *Transactions of the ASME – Journal of Basic Engineering.* 1960. [cit. 2006-09-10] http://www.cs.unc.edu/~welch/kalman/kalmanPaper.html.

[Ket99] KETTNAKER, V. – ZABIH, R. Bayesian Multi-camera Surveillance. *IEEE.* 1999.

[Kha05] KHAN, Z. – BALCH, T. – DELLAERT, F. *MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2005.

[Kos03] KOSCH, Harald. Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21. CRC Press, 2003. 280 s. ISBN 0-8493-1854-8.

[Kuc04] KUCERA, H. – KUCERA, G. Developing Standards for Time-varying Spatial Information, Discussion Paper. 2004. [cit. 2006-10-19] http://www.wiscorp.com/SQLStandards.html.

[Kuch04] KUCHAŘEK, T. – KRŮČEK, J. *SQL Multimedia and Application Packages* [in Czech, online]. 2004. [cit. 2006-09-22] http://kocour.ms.mff.cuni.cz/~pokorny/dj/prezentace/2_51.ppt.

[Mak04] Dimitrios Makris – Tim Ellis – James Black. Bridging the Gaps between Cameras. *Proc. of the 2004 IEEE Conf. on Computer Vision and Pattern Recognition.* 2004.

[Mar03] MARQUES, J. S. et al. Tracking Groups of Pedestrians in Video Sequences. *Proc. of the 2003 Conf. on Computer Vision and Pattern Recognition Workshop.* 2003.

[Mar04] MARIN, J.M. – MENGERSEN, K. – ROBERT, C.P. Bayesian modelling and inference on mixtures of distributions [online]. *Handbook of Statistics.* D. Dey and C.R. Rao. vol. 25. Elsevier. 2005. ISBN 0-444-51539-9. [cit 2007-01-10] http://www.ceremade.dauphine.fr/%7Exian/mixo.pdf.

[Mat05] Robust Non-Rigid Object Tracking Using Point Distribution Manifolds [online]. *British Machine Vision Conf.* 2005. [cit. 2007-01-10]. http://www.montefiore.ulg.ac.be/~mathes/.

[MDM06] *MDM/KDD2006 Seventh International Workshop on Multimedia Data Mining* [online]. 2006. [cit 2007-01-10] http://www.fortune.binghamton.edu/MDM2006/.

[Mie06] Mierswa, I. at. al. YALE: Rapid Prototyping for Complex Data Mining Tasks. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2006.

[MP2102] ISO/IEC JTC1/SC29/WG11. *MPEG-21 Overview* [online]. Bormans, J. – Hill K. Shanghai: 2002 [cit. 2006-09-14]. Available from: http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

[MP402] ISO/IEC JTC1/SC29/WG11. *MPEG-4 Overview.* [online]. Koenen, Rob. 2002 [cit. 2006-05-22]. Available from: http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm.

[MP704] ISO/IEC JTC1/SC29/WG11. *MPEG-7 Overview* [online]. Martínez, José M. Palma de Mallorca : 2004 [cit. 2005-10-22]. Available from: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.

[Neu84] NEUMANN, B. Optical Flow [online]. ACM Computer Graphics. 1984. [cit. 2007-01-16] portal.acm.org/ft_gateway.cfm?id=988528&type=pdf.

[Ngu03] NGUYEN, N. et al. Multiple camera coordination in a surveillance system, *ACTA Automatica Sinica*, Vol 29. 2003. [cit. 2007-01-16] http://www.computing.edu.au/~nguyentn/.

[Ngu06] NGUYEN, N – Bui, H. – Venkatesh, S. Recognising behaviour of multpile people with hierarchical probabilistic and statistical data association. *17th British Machine Vision Conference*. Scotland, 2006. [cit. 2007-01-16] http://www.computing.edu.au/~nguyentn/.

[Ora06] Oracle Corporation. *Oracle Database Online Documentation 10g Release2* [online]. 2005 [cit. 2006-05-16]. Dostupný z: http://www.oracle.com/pls/db102/portal.portal_db.

[Pap06] Papanikolopoulos, N. et al. *Monitoring Human Activity: A project of the Artifical Intelligence, Robotics and Vision Laboratory University of Minnesota, Department of Computer Science and Engineering* [online]. 2007. [cit. 2006-09-10] http://mha.cs.umn.edu/.

[PETS06] Procs. of the Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006). 2006.

[Pia06] PIATETSKY-CHAPIRO, G. et. al. Is There A Grand Challenge or X-Prize for Data Mining? *The Twelfth Annual SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2006.

[Pla04] PLAGEMANN, Thomas et. al. *Using Data Stream Management Systems for Traffic Analysis. A Case Study* [online]. 2004. [cit. 2006-06-19]. http://www.pam2004.org/papers/113.pdf.

[Pol03] POLAT, E. – YEASIN, M. – Sharma, R. Robust tracking of human body parts for collaborative human computer interaction. *Computer Vision and Image Understanding 89.* 2003.

[Pos06] PostgreSQL Global Development Group. *PostgreSQL* [online]. [cit. 2006-10-19] http://www.postgresql.org/.

[Rab89] RABINER, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of the IEEE.* vol. 77, no. 2. 1989.

[Rah04] RAHIMI, A. – DUNAGAN, B. – DARELL, T. Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors. Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2004.

[Ram07] Tracking People by Learning Their Appearance. Deva Ramanan, Member*, IEEE, David A. Forsyth, Senior Member, IEEE, and* Andrew Zisserman. EEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, 2007.

[RCo06] RAD COM. *Voice Over IP Reference Page* [online]. 2006. [cit. 2006-06-19]. http://www.protocols.com/pbook/VoIP.htm.

[Rea06] RealNetworks, Inc. Helix Community [online]. [cit. 2006-10-19] https://helixcommunity.org/.

[Ren05] REN, Liu et al. Learning Silhouette Features for Control of Human Motion. ACM Transactions on Graphics. vol. 24, no. 4, 2005.

[Row99] ROWEIS, S. – Ghahramani, Z. An Unifying Review of Linear Gaussian Models. *Neural Computation.* Vol 11 No 2, 1999. [cit. 2007-01-01] http://citeseer.ist.psu.edu/roweis97unifying.html.

[Sch01] Schlesinger, M. I. – Hlavac, V. Ten lectures from statistical and structural pattern recognition. Kluwer Academic Publishers. 2002.

[Sew01] Sewell, G. – Barker, J. R. Neither good, nor bad, but dangerous: Surveillance as an ethical paradox. Ethics and Information Technology 3: Kluwer. 2001.

[She03] SHEKHAR, S. *Spatial databases a tour.* 262 p. 2003. ISBN 0-13-017480-7.

[Sil02] SILBERSCHATZ, A. – KOHRT, F.H. – SUDARSHAN, S. *Database system concepts.* 4th edition. Boston, McGraw-Hill, 2002, 1064 p. ISBN 0-07-228363-7.

[Smi02] SMITH, L. I. *A tutorial on Principal Components Analysis* [online], 2002. [cit 2007-01-10] http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

[Soa01] SOATTO, S. – DORETTO, G. – WU, Y. Dynamic Textures. *Intl. Conf. on Computer Vision.* 2001.

[Soch05] SOCHMAN, J. MATAS, J. WaldBoost – Learning for Time Constrained Sequential Detection. Proc. of the 2005 IEEE Conf. *on Computer Vision and Pattern Recognition.* 2005.

[Sol06] Solomonoff, R. Machine Learning – Past and Future [online]. *The Dartmouth Artificial Intelligence Conference.* 2006. http://world.std.com/~rjs/pubs.html

[Son99] SONKA, M. – Hlavac, V. – BOYLE, R. Image processing, analysis, and machine vision. New York, PWS Publishing, 1999. 770p. ISBN 0-534-95393.

[Sor70] SORENSON, H. W. Least-squares estimation: from Gauss to Kalman, *IEEE Specrum* [online]. 1970. [cit. 2006-09-10] http://www.cs.unc.edu/~welch/kalman/media/pdf/Sorenson1970.pdf.

[Spe97] SPECHT, G. - ZIMMERMANN, S. – CLAUSNITZER, A. Introducing Parallelism in Multimedia Database Systems. *Proc. of the 2nd. Int. Symposium on Parallel Algorithms/Architectures Synthesis.* IEEE. 1997.

[Sub98] SUBRAHMANIAN, V. S. *Multimedia database systems.* Morgan Kaufmann, 1998. 442 s. ISBN 1-55860-466-9.

[Tes03] Tešic, Jelena. *Multimedia Mining Systems* [online]. 2003. [cit. 2007-01-19] http://vision.ece.ucsb.edu/~jelena/research/MiningSystems.pdf.

[Tri05] TRIVEDI, M. M. – GANDHI, T. L. – Huang, K. S. Distributed Interactive Video Arrays for Event Capture and Enhanced Situational Awareness. *IEEE Intelligent Systems.* 2005.

[VACE07] Informedia, ARDA. *VACE : Video Analysis and Content Exploitation* [online]. 2004-2007 [cit. 2007-01-16]. http://www.informedia.cs.cmu.edu/arda/index.html

[Vap63] VAPNIK, V. – LERNER, A. Pattern recognition using generalized portrait method, *Automation and Remote Control*, 24. 1963.

[VSAM00] COLLINS et al. *A System for Video Surveillance and Monitoring: VSAM Final Report.* Carnegie Mellon University, 2000. [cit. 2006-09-10] http://www.cs.cmu.edu/~vsam/research.html.

[Wan03] WANG, L. et. al. Silhouette Analysis-Based Gait Recognition for Human Identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. vol. 25, no. 12. 2003.

[WIS07] Whitemarsh Information Systems Corporation. *SQL Standards* [Online]. [cit. 2007-01-10] http://www.wiscorp.com/SQLStandards.html.

[Zha05] ZHAO, Tao et al. *Real-time Wide Area Multi-Camera Stereo Tracking. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 2005.

[Zho05] Yongluan Zhou, Y. – Ooi, B. C. – Tan, K. Dynamic Load Management for Distributed Continuous Query Systems.

[Zim06] ZIMMERMANN, K. – SVOBODA, T. – MATAS, J. Multiview 3D Tracking with an Incrementally Constructed 3D Model. *Third International Symposium on 3D Data Processing, Visualization and Transmission.* 2006.

[CLEAR] *CLEAR Evaluation Workshop* [online]. Supported by CHIL, NIST and ARDA VACE. 2007. [cit. 2007-01-19] http://www.clear-evaluation.org/.

[AXIS] Axis Communications. *AXIS IVM120 People Counter* [online]. 2007. [cit. 2007-01-19] http://www.axis.com/products/cam_ivm120/index.htm.

[COGNI] Cognitec Systems. The Face Recognition Company [online]. 2007. [cit. 2007-01-19] http://www.cognitec-systems.de/.

[Camea] Camea. Unicam [online]. 2007. [cit. 2007-01-19] http://www.unicam.cz/.

[L1ID] L-1 IDENTITY SOLUTIONS. About [online]. 2007. [cit. 2007-01-19] http://www.l1id.com/.

[CARETAKER] Content Analysis and Retrieval Technologies to Apply Knowledge Extraction to massive Recording [online]. 2007. [cit. 2007-01-19] http://www.ist-caretaker.org/.

[Cit07] Citeseer. Face recognition [online]. 2007. [cit. 2007-01-19] http://citeseer.ist.psu.edu/Applications/FaceRecognition/.

[Oli01] OLIVIER, N. – ROSARIO, B. – PENTLAND, A. A Bayesian Computer Vision System for Modeling Human Interactions. 1999.

[Sch03] Schmidhuber, J. Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements [online]. 2003. [cit. 2006-12-01] http://arxiv.org/abs/cs.LO/0309048.