Video Analysis for Ambient intelligence in Urban Environments

Andrea Prati and Rita Cucchiara

Abstract Ambient Intelligence (AmI) is an emerging field of research that comprises new paradigms, techniques and systems for intelligent processing of distributed sensing. A challenging arena for AmI framework is represented by urban environments that are characterized by high complexity, numerous sources of data, and spreading of interesting and non-trivial applications. In this context, the project LAICA (Laboratory of Ambient Intelligence for a friendly city) represents a real experiment of the usefulness of AmI for advanced services to citizens. This chapter will address solutions of video analysis that can be directly applied in urban AmI. It describes in details the uniqueness of LAICA approach, focusing in particular on the use of computer vision techniques for monitoring public parks. People surveillance and web-based video broadcasting will be taken into account.

1 Introduction

With the term "Ambient Intelligence" (or AmI) we typically refer to new paradigms, techniques and systems for acquiring data, processing information, creating and spreading knowledge in distributed environments. This new multi-disciplinary field of research has spread in the scientific community in the last decades, also thanks to the diffusion of sensors and the increase of processing power.

The typical contexts are areas where more heterogeneous sources of data coexist and can share raw and processed data. This sharing/cooperation among sensors

Andrea Prati
Dipartimento di Scienze e Metodi dell'Ingegneria, University of Modena and Reggio Emilia, Via Amendola, 2 - Pad. Morselli, 42100 Reggio Emilia, Italy, e-mail: andrea.prati@unimore.it

Rita Cucchiara
Dipartimento di Ingegneria dell'Informazione, University of Modena and Reggio Emilia, Via Vignolese, 905/b, 41100 Modena, Italy, e-mail: rita.cucchiara@unimore.it

contributes to the common scope of improving the *"intelligence"*, with the Latin meaning of the act of *"intelligere"*, i.e. to *comprehend the world*.

Ambient intelligence research can be applied in the house, to improve the processing capabilities of last generations of home automation systems [16, 15], in distributed virtual communities for data sharing with autonomous mobile agents [33, 34, 35], in complex systems, such as the interaction between remote patients and health care systems [32]. Among the possible applications, one of the most challenging is the urban context, i.e. the city as a complex entity with people and numerous sources of data. The regional project LAICA[1] (acronym for Laboratorio di Ambient Intelligence per una Città Amica - in English: Laboratory of Ambient Intelligence for a Friendly City) has been conceived in this framework and will be detailed in section 2.1.

Among the many sources of information, videos assume a central role for many reasons. Firstly, visual data can now be acquired at reasonable costs by using cheap cameras already installed in many public places (train stations, intersections, public parks, airports, etc.). Secondly, visual data are now much easier to transfer due to distributed wired and wireless networks available in most of the cities. Last but not the less important, visual data contain the highest amount of information about the environment and people that live in it.

For all these reasons, within LAICA we investigated the exploitation of visual data to extract information on status, behavior, and interaction of people and vehicles in urban contexts. Moreover, privacy and ethical issues will be taken into account and examples of applications in the LAICA projects will be described.

2 Visual Data for Urban AmI

In the past decades, visual documents have been the principal media of communication for tourist (virtual guides) and planning (remote sensing, SAR images) purposes for our cities. Nowadays, instead, the principal sources of visual information are live visual data acquired in real time from the hundreds of webcams and often cameras installed everywhere. Most of these cameras were installed only for tourist purposes (such as in the case of web pages of Time Square in New York City[2] - USA - or in Graz[3] - Austria - in which also PTZ control is made available to the user). However, also the thousands of cameras mounted as part of video surveillance systems can be potentially used.

The name of *video surveillance* is now synonymous of whichever system that uses cameras, acquires videos, possibly - but not necessarily - processes them, transfers them to remote displays in control centres and stores the data for posterity

[1] http://www.laica.re.it

[2] http://www.earthcam.com/usa/newyork/timessquare/

[3] http://www.graz.at/cms/ziel/1097909/DE/

logging. However, the existing video surveillance systems are still very prone to problem with privacy regulations.

2.1 Video Surveillance in Urban Environment

Video surveillance is motivated by three main purposes, also known as the S^3 *motivations*: security, safety, statistics.

After the tremendous acts of September 11, 2001, every city in the world became insecure and the *possibility* to add "electronic eyes" to control the urban environment became an unavoidable *requirement*. This requirement reflected on the spreading of video surveillance systems in public places (especially metro and train stations, and airports) in order to prevent crimes, vandalism, or even terrorist attacks.

In Modena (Italy) the project "Modena Secure City" consisted in installing 42 cameras in critical locations, connected to the Police control centre and equipped with PTZ (Pan Tilt Zoom) capabilities to allow active control and zooming. Stored videos are available for forensic analysis. In the city of Reggio Emilia (Italy), before the start of the LAICA project, more than 100 cameras have been installed nearby the railway station and in several public parks. Other examples can be found all over the world, like New York City which has about 5000 cameras only in Manhattan, or London which is the city with the highest number of cameras (about 150000 in 2004) in the world. In total, Italy counts about 2 millions of cameras in 2004, while UK reaches, in the same year, the considerable count of 4 millions with a citizen's picture taken on average 300 times per day.

Also European commission has expressed much interest in video surveillance in urban environments, since Fifth Framework Programme (e.g., the Urbaneye Project[4], or CAVIAR project[5] in Sixth FP), and still much interest will be devoted to this research in the task of security in the Seventh Framework Programme, where a specific strategic objective (among others) called "Intelligent urban environment observation system" is included.

Finally, many commercial systems have been developed in the last years, some of them rather sophisticated, for instance the system developed by Bosch Security System[6], or the iOmniscient[7] (Australia) company that argues to have the most intelligent video surveillance system.

Video surveillance is becoming more and more popular also for private use, in houses, offices, banks, to guarantee the personal safety of citizens and workers. New generations of video surveillance systems have been also mounted on mobile platforms. An example of this is the system developed by ELSAG (Italy) to automatically read vehicle license plates of stolen cars while a police car is moving. As an

[4] http://www.urbaneye.net/index.html

[5] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/caviar.htm

[6] http://www.boschsecurity.com

[7] http://iomniscient.com/

example of safety application, a system for smart deployment of airbags in the car has been developed at UCSD (Usa) by the group of Mohan Trivedi [8]: here, multiple cameras (both standard and omnidirectional) are used to detect driver's posture in real time in order to decide whether to deploy airbag or not.

Finally, video surveillance can be used also for collecting statistics on people, behaviors, vehicles, etc.. These statistics can be used for dissemination to citizens or public officers, or for planning purposes. This is often the purpose of vision-based traffic monitoring systems, for both urban roads and highways, employed to measure queues, quantify lane occupancy and turning rates, detect incidents, measure speed, for access control to restricted areas, and so on. An interesting application related to traffic control and part of the LAICA project is that of monitoring roundabouts for occupancy analysis and license plate recognition.

For instance, Belgian company Traficon N.V. is one of the world-leader companies in vision-based traffic control systems and has collaborated with us for more than two years for the development of a board-based system for safety in highway tunnels (called VIP-T) capable to detect automatically incidents, monitor vehicle flows, and collect statistical data. Computer vision algorithms have been developed for vehicle detection and tracking with the purpose to measure speeds, classify vehicles and detect stopped vehicles inside tunnels [11].

Statistical analysis is going to be of interest also when related to people, for example to analyze crowds of people in public places. Examples of applications are the monitoring of bus stops to plan frequencies of bus runs, crowd dynamics in department stores, etc... In all these cases privacy is a very relevant issue, being people monitored without their explicit consensus. This issue will be discussed deeply in section 4.

As a final, emerging application, it is worth of nomination the use of video analysis for posterity logging to post-process huge amount of data for supporting forensic investigation in case of crimes, vandalism, or terrorist attacks. For example, during the assassination of Prof. Marco Biagi in Italy in 2002 more than 50000 hours of video tapes have been watched and manually annotated by police officers. Having a (semi-)automatic process for pre-processing these data would definitely help in such situations.

Summarizing, possible applications of video analysis in urban environments are reported in Table 1 and examples shown in Figure 1.

Video surveillance systems have greatly improved in the last decades. As reported in [25], they can be classified mainly in three generations. The first generation (until '80) was based on analog signal, videos were only viewed in remote by human operators by means of a large set of monitors. These systems have the huge limitation to require operators' attention, resulting in a high miss rate of the events of interest. Moreover, analog signal is very noisy and requires much bandwidth to be transmitted and much space to be stored. Thanks to the rapid improvements in the camera resolution, the availability of low-cost computers, and communications improvements, in late '80, the second generation systems started to emerge. These

Application	Example of system features
Traffic control	Queue analysis, incident detection, traffic light control
Monitoring and diffusing general information	Webcam broadcasting
People detection for safety purposes	Secure road crossings, metro/bus stop control
People tracking for security and surveillance	Surveillance of public areas (stadium, museum, etc.) or, in general, crowd areas
Environmental condition analysis	Fire and smoke detection, flood control
Citizen-to-computer interactions	Video interaction and communication systems (e.g., Infopoints)
Support for investigation	Posterity analysis for forensic purposes
Security for children	In the surroundings of schools or public parks, also in connection with the soliciting of minors
Control systems for cultural heritage	Monitoring of natural and historical parks, ...
Safety for elderly and children	Remote assistance for monitoring patients in intensive care units or quarantined patients

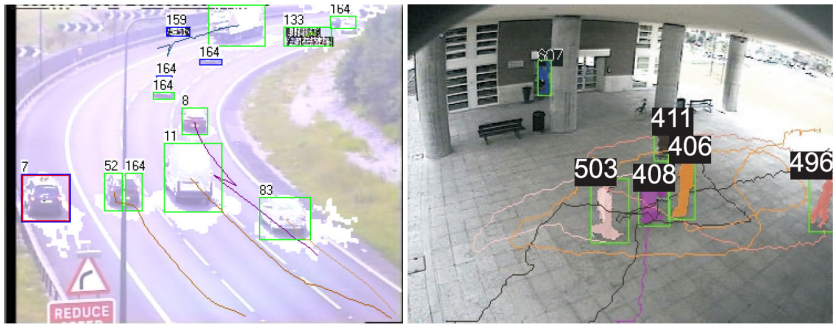
Table 1 Examples of applications of video analysis in urban environments.

systems benefited from early advances in digital video communications (e.g., digital compression, bandwidth reduction, and robust transmission) and in computer vision algorithms, and were mainly used to show the feasibility of digital, intelligent attention focusing systems on video from limited sets of cameras and for real-time analysis and segmentation of image sequences, identification and tracking of multiple objects in complex scenes, human behavior understanding, etc.. At the beginning of 21th century, the third generation of video surveillance systems takes place, providing a "full digital" solution to the design of surveillance systems: sensor and local processing layers can be physically organized together in a so-called intelligent camera; at the operator layer, an active interface is presented to the operator, assisting the operator by focusing his/her attention to a subset of interesting events.

Despite this classification, most of the existing video surveillance systems provide a limited automated processing capability, by incorporating motion detectors for automatic storage of videos or few more features. The goal of automated video surveillance is to extract meaningful objects from the observed scene, recognize them and their behavior, understand the scene by reasoning about objects and background, and infer specific conditions, alarms or interesting events.

A very promising advance to be included in the next (commercial) systems can be provided by including computer vision capabilities to the system. New types of alarms (not automatically provided in the majority of the current systems) could be:

- Low-level alarms: motion detectors, long-term change detectors, ...;
- Feature-based spatial alarms: specific-object detection in monitored areas (e.g., unattended bags in airports);
- Behavior-related alarms: anomalous trajectories, suspicious behaviors, ...;
- Complex event alarms: detection of complex scenarios related to multiple events.



(a) Traffic control (b) People tracking

(c) Web diffusion

Fig. 1 Some snapshots of possible applications of video surveillance in urban environments.

2.2 The LAICA Project

The project LAICA is an example where many of the above-mentioned advanced capabilities have been tested in a distributed environment.

LAICA is a two-years (2005-2006) project funded by Regione Emilia-Romagna for a total budget of over 2 Millions of euro and that involves universities, industries and public administrations for a total of about 320 man-months. The main objective of LAICA project is to explore the AmI capabilities in a medium-size Italian city such as Reggio Emilia. LAICA partners aim at defining innovative models and technologies for AmI in urban environments, and at studying and developing advanced services for the citizens and the public officers in order to improve personal safety and prevent crimes. The project brings together the academic expertise and the industrial knowledge into several fields, from the low-power sensor networks, to the computer vision, to the middleware and mobile agents, to the communication. Multimedia and multimodal data have been collected from different sources, such as cameras, microphones, textual data about the traffic, the security and the general situation of the city. As shown in Figure 2, the LAICA project has been structured

with a three-layers architecture, corresponding to three different level of granularity of the knowledge provided by sensors (punctual, local, and global). The processed information has been made available to both Police control centres and citizens by means of a dedicated webpage.

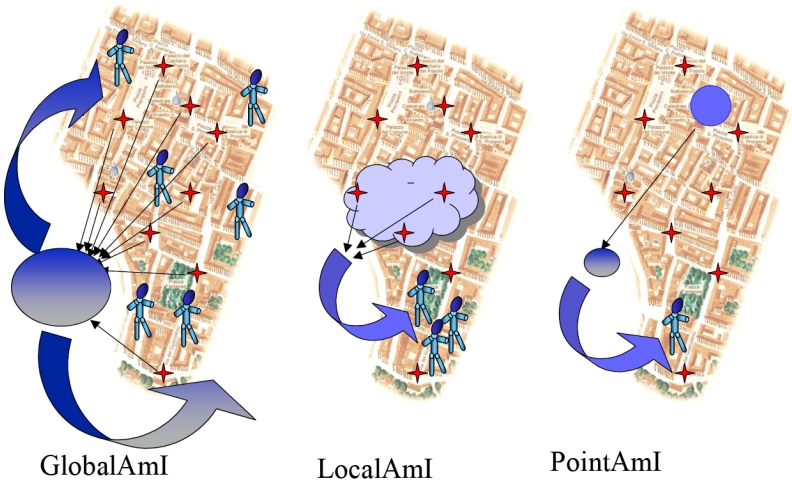


Fig. 2 The three layers of the LAICA AmI architecture.

The foreseen services should be provided by a set of prototypal systems, as for instance:

- a system for the automatic monitoring of pedestrian subways by means of mobile and low-power audio and proximity sensors [45];
- a system for the automatic monitoring of traffic scenes by cameras for data collection and web-based delivery of traffic news to citizens and police officers;
- a system that generates a feedback in pedestrian crossing systems to select the best duration of the green signal for the crossing [4];
- a platform of Urban TV to broadcast in interactive ways the data to the citizens;
- a system for the automatic monitoring of public parks with a plethora of cameras (both fixed and PTZ) [5, 41], also accounting for privacy issue [12].

In the next chapter the last prototype will be discussed in details to show how people can be detected and tracked in urban environments.

3 Automatic Video Processing for People Tracking

The motivations for tracking people in surveillance applications are numerous. The following is a list of the most important:

- Recognition of Human Motion, e.g.:
  - tracking people for statistical and security reasons, detecting moving people in dangerous zones;
  - walking, gait recognition;
  - counting, locating pedestrians;
  - abandon an object;
- Gesture Recognition, e.g.:
  - hand, arm tracking for gestures;
  - sign language recognition;
- Tracking Faces in Video, e.g.:
  - Face detection;
  - Eye tracking and gaze tracking;
  - Lip reading, lip tracking;
  - Face recognition.

Most of these applications are particularly relevant in urban environments. Moreover, the city areas under surveillance are typically large, requiring multiple cameras to cover them. Finally, PTZ cameras are often employed to either "patrol" a large scene or zoom onto a specific zone/target. With these premises, the following subchapters will briefly report on the research activity in the field of people detection and tracking by means of computer vision, starting from a single static camera, to multiple (static) cameras, to the use of PTZ cameras.

3.1 People Detection and Tracking from Single Static Camera

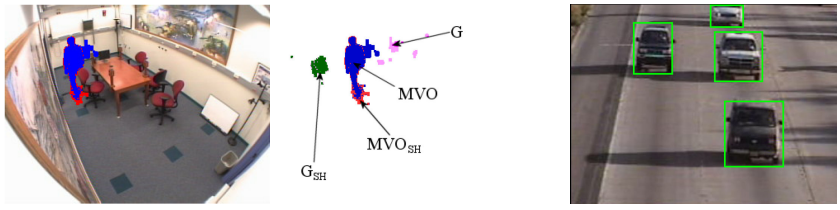
Detection of "moving objects" in video scenes is the basic step of major applications such as tracking and visual surveillance. This problem has been addressed since many years in both the scientific literature and the R&D for commercial systems and good solutions have been already proposed for static cameras. Among the many different approaches proposed, *background suppression* is the most diffused for its generality and reliability. The aim is to separate the foreground (moving visual objects, or MVOs) from the background model, i.e. a (probabilistic) model of the background as it changes in time. Thus, it is required to build and keep updated the background model, to *adapt* it to short- and long-term changes in illumination, to detect MVOs from the current frame (i.e., to suppress the background from it), and to handle difficult situations, such as shadows and the so-called "ghosts" (i.e., the false objects generated by a real still object that starts to move).

According to this, Table 2 shows a sketchy summary of these features (namely, background model construction, adaptive updating, suppression, and detection of other types of objects) and the most relevant approaches for background suppression used in the literature.

Feature	Approach used
Background model construction	Median [19, 9] Single Gaussian [43, 26] Mixture of Gaussians [38] Eigenbackground [29]
Adaptive updating	Kalman-based [23] Previous backgrounds [43, 19]
Background suppression	Intensity [23] Color [19, 26, 9] Malahanobis distance [43] Multi-valued distance [38] Eigenbackground distance [29]
Detection of other objects	Shadows [26, 9], Ghosts [38, 9]

Table 2 Summary of seminal approaches to background suppression

We also proposed an approach for background suppression from single static camera. The approach is part of the SAKBOT (Statistical And Knowledge-Based Object Tracker) described with full details in [9][21]. The background pixels are defined by two models: the first statistical model updates the pixels at each frame using the temporal median function over the previous n sampled pixels; the second model exploits the knowledge of previous background and of the corresponding moving objects. Specifically, the pixels belonging to the current moving objects are not used for updating the model in order to prevent the gradual inclusion of slowly-moving objects into the background. Instead, pixels detected as foreground at previous steps but classified as noise or shadows are included in the statistical model. This approach is critical when a stopped object starts to move, generating two "foreground" objects, one real and one apparent (the "ghost"). To avoid deadlock situations for the ghosts, a specific ghost suppression algorithm has been conceived. Moving shadows are classified using their appearance and assuming that shadows lower the brightness of the background underlying them, leaving the color components almost unchanged. Further details can be found in [31]. Figure 3 shows some examples of the output of the SAKBOT system.



(a) (b) (c)

Fig. 3 Examples of the SAKBOT output: (a) segmentation result with blue pixels indicating the MVO, while red ones indicating the shadow points; (b) complete classification of pixels for frame (a) with MVO, MVO shadow, Ghost, and Ghost shadow; (c) another example of SAKBOT's segmentation (with corresponding bounding boxes).

Tracking of MVOs is crucial for most of the applications reported above. "Tracking" means that the same label/identity is kept constant over time for the same object, allowing to compute trajectory, speed, analyze behavior, etc.. People tracking is one of the most widely explored topics in computer vision. There are many reference surveys in the field: the works of Cedras and Shah [7], of Gavrila [17], of Aggarwal and Cai [1] and Moeslund and Granum [28], or more recently, the work by Hu et al. in video surveillance [20] and the work by Wang et al. [42]. In people tracking, in order to cope with non-rigid body motion, frequent shape changes and self-occlusions, probabilistic and appearance-based tracking techniques are commonly proposed [26, 37]. In non-trivial situations, when more persons interact overlapping each other, most of the basic techniques tend to lose the previously computed tracks, detecting instead the presence of a group of people, and possibly restoring the situation after the group has split up [26]. These methods aim at keeping the track history consistent before and after the occlusion only. Consequently, during an occlusion, no information about the appearance of the single person is available, limiting the efficacy of this solution in many cases. Conversely, a more challenging solution is to try to separate the group of people into individuals also during the occlusion.

Our approach to moving object tracking is based on appearance [40]. This algorithm uses a classical predict-update approach. It takes into account not only the status vector containing position and speed, but also the memory appearance model and the probabilistic mask of the shape [10]. The former is the adaptive update of each pixel in the color space. The latter is a mask whose values, ranging between 0 and 1, can be viewed as the probability for that pixel to belong to that object. These models are used to define a MAP (Maximum A Posteriori) classifier that searches the most probable position of each person in the scene. The tracking algorithm is a suitable modification of a work, previously proposed by Senior [37], that includes a specific module for coping with large and long-lasting occlusions. Occlusions are classified into three categories: self-occlusions (or apparent occlusions), object occlusions, and people occlusions. Occlusion handling is very robust and has been tested in many applications. It can keep the shape of the tracked objects very precisely.

3.2 People Detection and Tracking from Distributed Cameras

The previous chapter briefly described the relevant issues for detecting and tracking people from a single static camera. However, as reported above, in most of the urban scenarios a single camera does not suffice to handle large areas and complex/cluttered scenes. For this reason, multiple cameras are used to both provide multiple viewpoints (useful for disambiguate difficult situations by using redundant data and for handling occlusions) and obtain the coverage of a wider area. Unfortunately, in automatic video surveillance multiple cameras are useless if uncorrelated. The exploitation of the multiple viewpoints to correlate data from multiple cameras is often called *consistent labeling*, referring to the fact that the label/identity

of moving objects is made consistent not only *over time* (as in the case of tracking from a single camera) but also *over space* (in the sense of different cameras). Consistent labeling permits to track people on wide area, increasing the potentiality of video-surveillance applications in urban scenarios.

Often cameras' fields of view are disjoint, due to installation and cost constraints. In this case, the consistent labeling should be based on appearance only, basing the matching essentially on the color of the objects (such as color histogram matching [30]).

If the fields of view are overlapped, consistent labeling can exploit geometry-based computer vision. These approaches exploit geometrical relations and constraints between different views.

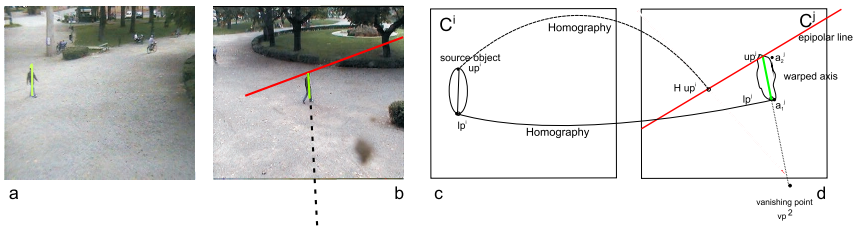
This could be done with a precise system calibration and 3D reconstruction could be used to solve any ambiguity [44]. However, this is not often feasible, in particular if the cameras are pre-installed and intrinsic and extrinsic parameters are not available. Thus, partial calibration or self-calibration methods can be adopted to extract only some of the geometrical constraints, e.g. to compute the ground-plane homography. An approach, based on the image projections of overlapped cameras' field of view lines, has been initially proposed by Khan and Shah in [22]: the lines delimiting the overlapping zones in the fields of view of the cameras are computed in a training phase with a single person moving in the scene. At run time, when one or more people have a camera handoff, the distances from the lines are used to disambiguate objects, assuring label consistency.

Another class of approaches presented in literature deals with multi-view geometry to analyze and impose continuity in the objects' trajectory across camera streams (e.g., [2, 39]).

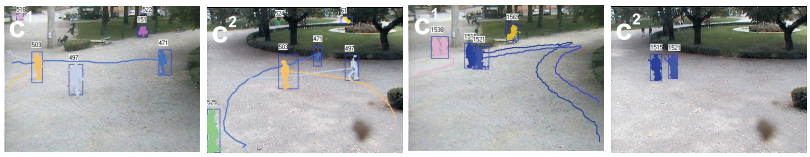
In [6], we have proposed a novel method, called *HECOL* (Homography and Epipolar-based COnsistent Labeling), to provide consistent labeling of people segmented in large areas covered by multiple overlapped cameras. The method takes into account both geometrical and shape features in a probabilistic framework. Homography and epipolar lines are computed to create relationships between cameras. The multi-camera system is modeled as a *Camera Transition Graph* (CTG) that defines the possible overlap between cameras in a given setup. When a new object is detected, the exploration of the graph selects a subset of compatible labels which may be assigned to the object in order to limit the search space. An off-line training phase allows to compute the *Entry Edges of Field of View* that define the area of overlapped FoV between cameras and permit to construct the homography. The learning phase also allows to compute the location of the epipoles of the overlapped cameras with a robust algorithm based on RANSAC optimization.

At run time, the system checks for inconsistency in label assignments among the modules of the overlapped cameras. The main novelty of the paper lies in the phase of consistent labeling that defines a probabilistic framework with forward and backward contributions: it checks the mutual correspondence of people using the axis of the objects precisely warped in the other FoV using epipolar lines. It accounts for the matching of the warped axis and the shapes of people. This makes the method

particularly robust against segmentation errors and allows to disambiguate groups of people. Figure 4 reports some images regarding the HECOL system.



(a) Process for axis warping



(b) (c)

Fig. 4 Examples of the HECOL system: (a) sketch of the process for axis warping on which the consistent labeling is based (see [6] for further details); (b) and (c) report examples from a real system working on a public park in Reggio Emilia within the LAICA project.

Multiple cameras can also be exploited directly to obtain 3D reconstruction of the moving object. In this case, the objective is not to have a consistent assignment of label among views but to correlate single points belonging to the object's shape, thus exploiting a sort of wide-baseline stereo system. An example of this use of multiple cameras can be found in [27]. The complete 3D reconstruction of the human shape has the obvious advantage of being crucial for applications such as human body modeling, gesture recognition, and similar.

3.3 People Detection and Tracking from Moving Cameras

In video-surveillance systems with multiple cameras it happens frequently that at least one of them is a PTZ camera. PTZ cameras have the main advantage to reduce the costs for covering large areas, allowing to use a single camera (even though more expensive than normal cameras), instead of a set of static cameras. PTZ camera can be programmed to patrol (automatically or manually by the operator) the scene. While patrolling, the camera should be able to extract and track moving people or even detect and track faces in order to zoom on one of them. Using a single PTZ camera solution exhibits the advantages of being basically a cheaper solution, of not requiring synchronization/communication among cameras, and of requiring, in principle, lower computational load. It has, however, also some drawbacks, namely

the need for more complex computer vision techniques and the limit of not allowing simultaneous coverage of a certain area.

Detecting and tracking people from a PTZ (i.e., moving) camera requires a rather different approach compared to what reported in section 3.1 for static cameras. One possible approach is that of creating in real time a *mosaic image* of the whole scene (by registering overlapped images provided by successive frames of the active camera) and then detect and track moving people on the mosaic image.

The segmentation of moving objects becomes more critical when the video is acquired by a moving camera with an unconstrained and a priori unknown motion. Proposals from single camera can be grouped into three classes: based on *ego-motion computation*, based on *motion segmentation*, and based on *region merging with motion*. The approaches in the first class aim at estimating the camera motion (or ego-motion) through the evaluation of the dominant motion with different techniques and models in order to obtain compensated videos and to apply algorithms developed for fixed camera (frame differencing, as in [14], or background suppression, as in [36]). In [21] Kang *et al.* define an adaptive background model that takes into account the camera motion approximated with affine transformation. Tracking of moving object is achieved by means of a joint probability data association filter (JPDAF). In methods based on motion segmentation the objects are mainly segmented by using the motion vectors computed at pixel level [24]. The vectors are then clustered to segment objects with homogeneous motion. Finally, the approaches based on region merging with motion are hybrid approaches in which the objects are obtained with a segmentation based on visual features, and next merged on motion parameters computed on a region-level [18]. It is worth noting that most of the reported approaches are computationally very expensive and cannot meet real-time constraints (and those that meet them use either special-purpose devices or a set of limiting assumptions).

In [41] we proposed a new method for fast ego-motion computation based on the so-called *direction histograms*. The method works with an uncalibrated camera that moves with an unknown path and it is based on the compensation of the camera motion (i.e., the *ego-motion*) to create the mosaic image and on the frame differencing to extract moving objects. Successive steps eliminate the noise and extract the complete shape of the moving objects in order to exploit an appearance-based probabilistic tracking algorithm. Figure 5 shows an example of the segmentation of a moving person by means of a single PTZ camera and its exploitation for automatically following the person.

4 Privacy and Ethical Issues

All the considerations reported in previous chapters are related to the usefulness of video surveillance in urban environments for increasing the (sense of) security, safety or merely for statistic collection. These are obvious and undoubted advan-

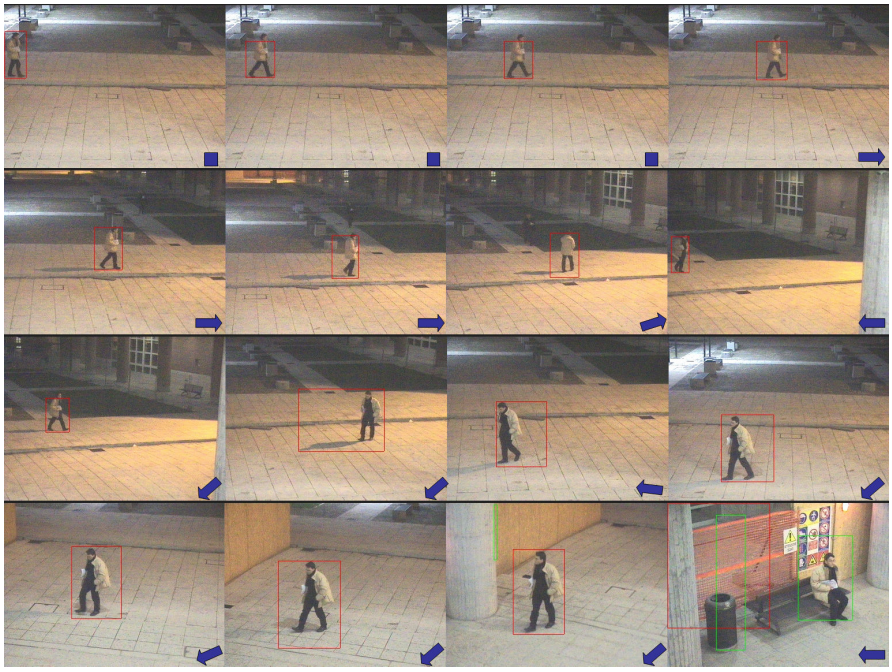


Fig. 5 Examples of the detection of a single moving person by using a single PTZ camera. Additionally, PTZ camera moves automatically to follow the moving person.

tages of (automatic) video surveillance. The use of cameras in public places brings, however, also serious problems to the citizen privacy. There is a world-spread controversy about the use of video surveillance at large, connected with risks of privacy violations. The dichotomy security vs. privacy was and is, for instance, very debated in US after September 11. K.W. Bowyer wrote a very interesting paper on pros and cons of surveillance and analyzed the risks of false claims in privacy violations [3]. Since in US the problem of security for urban vandalism, crime and terrorism is very spread, the privacy and ethical issues are often on a second order w.r.t the needs of the population. The Supreme Court of the United States declared that whichever visual data can be used for security investigation.

In Europe, this discussion has been amplified after the terrorist attacks in Madrid in March 2004 and in the London underground in July 2005, that has been ineludible even if London is the city in the world with the highest number of cameras installed. In that situation, the recorded videos have been of valuable help in terrorists' identification after the crime, but the systems were not capable to give an immediate alarm.

In Europe visual data processing is within a more general Directive (95/46/EC) than in US. This Directive covers specific features of the processing of personal information included in sound and image data and ensures "the protection of privacy and private life as well as the larger gamut of protection of personal data with regard

to fundamental rights and freedoms of natural persons". A considerable portion of the information collected by means of video surveillance concerns identified and/or identifiable persons, who have been filmed as they moved in public and/or publicly accessible premises. As a final remark, Directive states that *"in public places no automatic visual surveillance should limit the freedom of people"*.

Each European countries has then its own specific law. For instance, in Denmark surveillance of public streets, roads, squares or any similar area used for common travel is forbidden to private entities. Also in Italy, there exists a specific set of laws for video surveillance. These laws propose a basic principle called "proportionality principle", for which acquired data must be adequate, relevant and not excessive. As an example, acquiring and storing videos from a supermarket for statistical analysis is excessive, doing that for forensic analysis of crimes is not.

A good compromise between security and privacy comes from the use of computer vision. It allows to extract "biometric" information (such as faces) from the video, by still preserving semantic content to be freely distributed. This requires, as depicted in previous chapters, to detect and track people from multiple cameras, detect their faces and automatically obscuring them to prevent "identification" of the person.

In the framework of the project LAICA we have studied and developed two different solutions: the first makes use of passive sensors to develop a video-surveillance system integrated with the cameras [13], the second automatically extracts and obscures people faces from videos [12]. An example of face obscuration is reported in Figure 6.

Fig. 6 Example of face obscuration taken from [12].

References

1. Aggarwal, J. K., Cai, Q.: Human Motion Analysis: A Review. Computer Vision and Image Understanding. **73(3)**, pp. 428-440 (1999)
2. Black, J., Ellis, T.: Multi camera image tracking. Image and Vision Computing. **24(11)**, pp. 1256-1267 (2006)
3. Bowyer, K.W.: Face recognition technology and the security versus privacy tradeoff. IEEE Technology and Society. **1**, pp. 9-20 (2004)

4. Broggi, A., Fedriga, R.I., Tagliati, A., Graf, T., Meinecke, M.: Pedestrian Detection on a Moving Vehicle: an Investigation about Near Infra-Red Images. In: Proceedings of IEEE Intelligent Vehicle Symposium (IV), pp. 431-436 (2006)
5. Calderara, S., Cucchiara, R., Prati, A.: Group Detection at Camera Handoff for Collecting People Appearance in Multi-camera Systems. In: Proceedings of Conference on Advanced Video and Signal-based Surveillance (IEEE AVSS 2006), pp. 36-41 (2006)
6. Calderara, S., Prati, A., Cucchiara, R.: HECOL: Homography and Epipolar-based Consistent Labeling for Outdoor Park Surveillance. Computer Vision and Image Understanding (2007)
7. Cedras, C., Shah, M.: Motion-Based Recognition: A Survey. Image and Vision Computing. **13(2)** (1995)
8. Cheng, S.Y., Trivedi, M.M.: Human posture estimation using voxel data for "smart" airbag systems: issues and framework. In: Proceedings of IEEE Intelligent Vehicles Symposium (IV), pp. 84-89 (2004)
9. Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting Moving Objects, Ghosts and Shadows in Video Streams. IEEE Transaction on Pattern Analysis and Machine Intelligence. **25(10)**, pp. 1337-1342 (2003)
10. Cucchiara, R., Grana, C., Tardini, G., Vezzani, R.: Probabilistic People Tracking for Occlusion Handling. In: Proceedings of IAPR International Conference on Pattern Recognition (ICPR 2004), vol. 1, pp. 132-135 (2004)
11. Cucchiara, R., Melli, R., Prati, A., De Cock, L.: Predictive and Probabilistic Tracking to Detect Stopped Vehicles. In: Proceedings of Workshop on Applications of Computer Vision (WACV), pp. 388-393 (2005)
12. Cucchiara, R., Prati, A., Vezzani, R.: A System for Automatic Face Obscuration for Privacy Purposes. Pattern Recognition Letters. **27(15)**, 1809–1815 (2006)
13. Cucchiara, R., Prati, A., Vezzani, R., Benini, L., Farella, E., Zappi, P.: An Integrated Multi-Modal Sensor Network for Video Surveillance. Journal of Ubiquitous Computing and Intelligence (JUCI). **1**, pp. 1-11 (2007)
14. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. **22(8)**, pp. 781-796 (2000)
15. Friedewald, M., Da Costa, O., Punie, Y., Alahuhta, P., Heinonen, S.: Perspectives of ambient intelligence in the home environment. Telematics and Informatics. **22(3)**, 221–238 (2005)
16. Garate, A., Lucas, I., Herrasti, N., Lopez, A.: Ambient intelligence as paradigm of a full automation process at home in a real application. In: Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA, pp. 475-479. (2005)
17. Gavrila, D.M.: The Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding. **73(1)**, pp. 82-98 (1999)
18. Gelgon, M., Bouthemy, P.: A region-level motion-based graph representation and labeling for tracking a spatial image partition. Pattern Recognition. textbf33, pp. 725-740 (2000)
19. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: real-time surveillance of people and their activities. IEEE Transaction on Pattern Analysis and Machine Intelligence. **22(8)**. (2000)
20. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics - Part C. **34(3)**, pp. 334-352 (2004)
21. Kang, J., Cohen, I., Medioni, G.: Continuous tracking within and across camera streams. In: Proceedings of IEEE-CS Int'l Conf. on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. I-267 - I-272 (2003)
22. Khan, S., Shah, M.: Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. IEEE Transactions on Pattern Analysis and Machine Intelligence. **25(10)**, pp. 1355-1360 (2003)
23. Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B., Russel, S.: Towards Robust Automatic Traffic Scene Analysis in Real-Time. In: Proceedings of International Conference on Pattern Recognition (1994)
24. Lee, K.W., Ryu, S.W., Lee, S.J., Park, K.T.: Motion based object tracking with mobile camera. Electronic Letters. **34(3)**, pp. 256-258 (1998)

25. Marcenaro, L., Oberti, F., Foresti, G.L., Regazzoni, C.S.: Distributed architectures and logical-task decomposition in multimedia surveillance systems. Proceedings of the IEEE. **89(10)**, 1419–1440 (Oct. 2001)
26. McKenna, S.J., Jabri, S., Duric, Z., Rosenfeld, A., Wechsler, H.: Tracking groups of people. Computer Vision and Image Understanding. **80(1)** (2000)
27. Mikic, I., Trivedi, M.M., Hunter, E., Cosman, P.C.: Human Body Model Acquisition and Tracking Using Voxel Data. International Journal of Computer Vision. **53(3)**. pp. 199-223 (2003)
28. Moeslund, T.B., Granum, E.: A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding. **81**, pp. 231-268 (2001)
29. Oliver, N.M., Rosario, B., Pentland, A.P.: A bayesian computer vision system for modeling human interactions. IEEE Transaction on Pattern Analysis and Machine Intelligence. **22(8)** (2000)
30. Orwell, J., Remagnino, P., Jones, G.A.: Multi-camera color tracking. In: Proceedings of Second IEEE Workshop on Visual Surveillance (VS'99), pp. 14-21 (1999)
31. Prati, A., Mikic, I., Trivedi, M.M., Cucchiara, R.: Detecting Moving Shadows: Algorithms and Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence. **25(7)**, pp. 918-923 (2003)
32. Riva, G.: Ambient Intelligence in Health Care. Cyberpsychology and Behavior. **6(3)**, 295–300 (2003)
33. Riva, G., Davide, F., Ijsselsteijn, W.A.: Being There: Concepts, effects and measurements of user presence in synthetic environments. IOS Press (2003)
34. Satoh, I.: Software Agents for Ambient Intelligence. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pp.1147-1150 (2004)
35. Satoh, I.: Mobile Agents for Ambient Intelligence. In: Lecture Notes in Computer Science (LNCS), vol. 3446, Springer (2005)
36. Sawhney, H., Ayer, S.: Compact representations of videos through dominant and multiple motion estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence. **18(8)**, pp. 814-830 (1996)
37. Senior, A.: Tracking people with probabilistic appearance models. In: Proceedings of Int'l Workshop on Performance Evaluation of Tracking and Surveillance (PETS) systems, pp. 48-55 (2002)
38. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Transaction on Pattern Analysis and Machine Intelligence. **22(8)** (2000)
39. Tsutsui, H., Miura, J., Shirai, Y.: Optical Flow-based Person Tracking by Multiple Cameras. In: Proc. 2001 Int. Conf. on Multisensor Fusion and Integration in Intelligent Systems, pp. 91-96 (2001)
40. Vezzani, R.: Computer Vision for People Video Surveillance", Ph.D. Thesis, (2006) Available via Internet.
http://imagelab.ing.unimo.it/Pubblicazioni/publications_query.asp?lang=en&autore=+55+&categoria=0&tipo=5. Cited 12 Aug 2007
41. Vezzani, R., Prati, A., Cucchiara, R.: Advanced Video Surveillance with Pan Tilt Zoom Cameras, In: Proceedings of Workshop on Visual Surveillance (VS) (2006)
42. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. Pattern Recognition. **36(3)**, pp. 585-601 (2003)
43. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: real-time tracking of the human body. IEEE Transaction on Pattern Analysis and Machine Intelligence. **19(7)** (1997)
44. Yue, Z., Zhou, S.K., Chellappa, R.: Robust two-camera tracking using homography. In: Proceedings of IEEE Intl Conf. on Acoustics, Speech, and Signal Processing, vol 3, pp. 1-4 (2004)
45. Zappi, P., Farella, E., Benini, L.: A PIR based wireless sensor node prototype for surveillance applications. In: Proceedings of European Workshop on Wireless Sensor Networks (EWSN 06), pp. 26-27 (2006)