Ambient Intelligence in Urban Environments

Rita Cucchiara [1], Andrea Prati [2], Cesare Osti [3], Stefano Pavani [3]

[1] Dipartimento di Ingegneria dell'Informazione (DII), University of Modena and Reggio Emilia, Italy
[2] Dipartimento di Scienze e Metodi dell'Ingegneria, University of Modena and Reggio Emilia, Italy
[3] Regulus SpA, Bologna, Italy
{cucchiara.rita, prati.andrea}@unimore.it, {cesare.osti, stefano.pavani}@regulus.it

Abstract. This paper reports advances achieved within a project called LAICA (Laboratorio di Ambient Intelligence per una Città Amica) on Ambient Intelligence in urban environments. The overall LAICA architecture is described and the unified operative centre developed by Regulus SpA (partner of the project) to collect and correlate data from different sensors and prototypes is depicted. Moreover, the paper describes the results obtained in developing a system for video surveillance in public parks, devoted to create a mosaic image of the scene and to extract and track moving people. Moreover, the system takes the privacy issues into account, proposing a method for face detection and tracking able to obscure faces in order to protect people's identity.

1. The LAICA Project

This paper reports on a project called LAICA, acronym for Laboratorio di Ambient Intelligence per una Città Amica (in English: Laboratory of Ambient Intelligence for a Friendly City). This is a two-years (2005-2006) regional project that involves universities, industries and public administrations. The main aims of the LAICA project are to define innovative models and technologies for Ambient Intelligence in urban environments, and to study and develop advanced services for the citizens and the public officers in order to improve personal safety and prevent crimes. This multi-disciplinary 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.

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 [1];
- 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 [2];
- a system that generates a feedback in pedestrian crossing systems to select the best duration of the green signal for the crossing [3];
- a system for the automatic monitoring of public parks with a plethora of cameras (both fixed and PTZ - pan/tilt/zoom) [4];
- a platform of Urban TV to broadcast in interactive ways the data to the citizens [5].

The project represents a suitable platform for exploring different paradigms of Ambient Intelligence (AmI). Ambient Intelligence is the new multi-disciplinary field that brings together sensor fusion, computer vision, middleware, multimedia and many other research fields. In fact, these prototypes are meant both to analyze the urban environment, collecting statistics or reporting on alarming situations (in the case of traffic analysis and public park surveillance), and to control it (providing a feedback to actuators, as in the case of the traffic lights for pedestrian crossing). Additionally, they can interact with the environment, allowing the citizens to access to real-time multimedia information on the status of the city.

---

[1] Studied and developed by the Micrel lab at DEIS – University of Bologna, in collaboration with Wavelet Technology Italia SrL
[2] Developed by Wavelet Technology Italia SrL, Regulus SpA and OT Consulting SrL
[3] Studied and developed by the Artificial Vision and Intelligent Systems lab at DII – University of Parma
[4] Studied and developed by the Imagelab lab at DII – University of Modena and Reggio Emilia
[5] Studied and developed by the DSC – University of Modena and Reggio Emilia

The LAICA project has been structured with a three-layers architecture, corresponding to three different level of granularity of the knowledge provided by sensors. The test bed of the whole LAICA project is the city of Reggio Emilia, in Italy, whom municipality is active partner of the project. The project starts from the exploitation of the existing network of sensors (consisting of hundreds of cameras for surveillance purposes and intelligent sensors in road intersections and subways) and available infrastructures (hundreds of kilometres in fibre channels for transmission at large bandwidth) deployed in Reggio Emilia. However, the above mentioned sensors are of different degree of complexity and computational power, ranging from the simple sensors for environmental measurements, to matrix of proximity, temperature and positioning (GPS) sensors, to fixed and moving cameras. All these sensors are connected through a MAN or wireless networks in order to communicate useful information to an operative centre or to be queried by a human operator. SIRTI SpA and Agactel SpA have in charge the deployment of the infrastructures useful for the connectivity of the different parts of the project.

In this scenario, the foreseen architecture of LAICA is composed of three layers (see Fig. **1**):

1. *punctual layer (pointAmI)*: the AmI system is seen as a distributed collection of low-level sensors and actuators. These points of sensing and control can be singularly queried, possibly also through mobile devices, such as PDA or last-generation cell phones. For instance, it will be possible to look at a specific camera, or querying the state of the traffic at a specific intersection, and so on. The system sends the requested data with a point-to-point connection;
2. *local layer (localAmI)*: the AmI system can be seen as a cluster of sensors, dynamically reconfigurable. The expert (and authorized) users can define their own set of sensors. For example, a traffic light system can be a collection of cameras, inductive loops and proximity sensors, and of several actuators too. A citizen may want to query the state of the traffic light in its entirety, as a localized set of entities. Despite its locality, the set can require distributed processing and coordination, and must be accessible. Middleware with distributed agents will be developed for this aim;
3. *global layer (globalAmI)*: the whole AmI system must be able to interact with the environment and to collect data from distributing sensing. The global knowledge is handled by an operative centre described in the following.

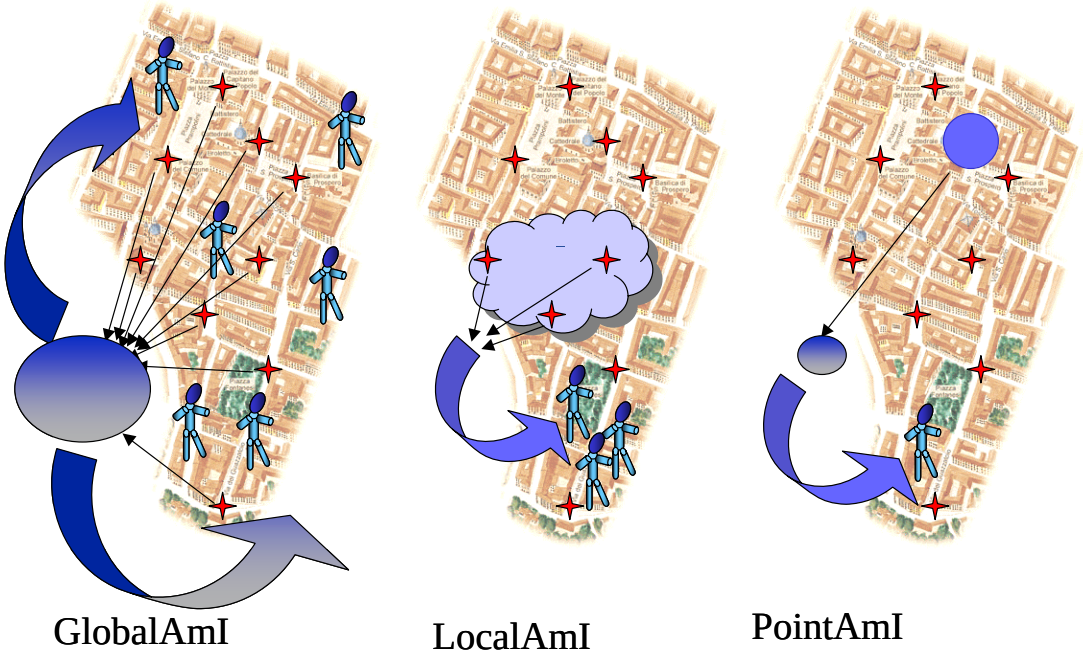


Fig. 1. The three layers of the LAICA AmI architecture.

In this scenario, each sensor or cluster of sensors must have a sufficiently high computational power to extract useful information from the scene.

The main aspect of the project is that it has been considered as a cooperative research and industrial project with strong interactions among the actors in order to implement a real AmI framework. It is not

merely a sum of services, but, conversely, the services and systems (some standards, some innovative for a public administration) will exploit their synergies, share communication supports and collecting servers, and be accessed concurrently as a single unique intelligent centre of monitoring of urban environments.

2. The LAICA operative centre

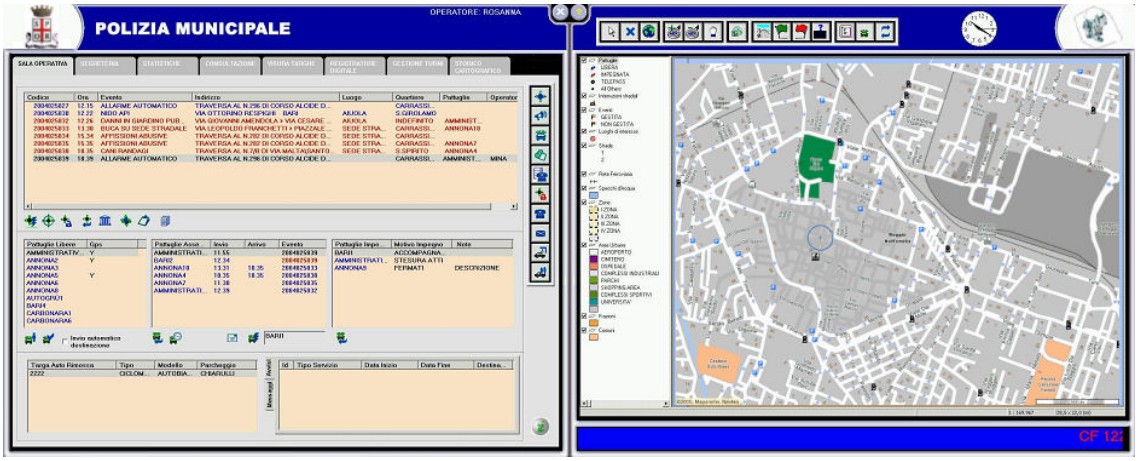


Fig. 2. A snapshot of the Regulus system for operative centre management

Road security and conditions, territory control and citizens safety, integrated road haulage, environment and pollution, are relevant and debated topics, not only at national and European level, but also at local level (city, province and region). An Ambient Intelligence software platform enables the concurrent focus on all these topics; in this context, Regulus has developed a local Operative Centre (Fig. **2**) that is the right solution to manage unexpected and planned events and activities; the Operative Centre is the final point to store and to manage information.

The Operative Centre (O.C.) has to integrate and to coordinate an intelligent network, made by sensors and devices (audio sensors, movement sensors, video cameras, etc.) distributed around the local area. These sensors and devices provided with distributed intelligent algorithms, using a wired or wireless connection, are able to signal critical events to the O.C. (a road accident, traffic jam, ice, suspicious and dangerous situations, etc.) or to signal data (counting vehicles, accessing city centre, etc.) in charge of activate all the means at its disposal.

The O.C. has a component-based architecture; each component is a solid, independent working unit, able to cooperate with the other components using standard and well-defined interfaces.

Using this approach it's possible to start with the needed functions and to increase and to integrate more components according to security and safety plans.

The base components are:

- Localization sub-system: in charge of finding the position for personnel and vehicles, using digital maps and GPS system, exchanging data and messages.
- Management sub-system: in charge of managing personnel, vehicles, activities, calls from citizenship and operators, etc.
- Communication sub-system: in charge of connecting devices, sensors, vehicles and personnel. Different infrastructures could be used, for example: Radio, GSM/GPRS, TETRA and UMTS.

Other important components are:

- Historical sub-system: in charge of retrieving all the data managed by O.C.
- Statistical sub-system: in charge of creating textual or graphical presentation about the data store in database, according to filters inserted.
- Duties sub-system: in charge of settling the work shift for the personnel, vehicles and other assets.

Moving from an Operative Centre to a Unified Operative Centre (U.O.C.) is the natural evolution to address in the right way security and safety problems. Integrating different Departments inside the same

local authority (for example local police - civil guard - road service), or connecting different Authorities (for example local police - national police) are the next steps to move towards a U.O.C.

Each department or authority will maintain its autonomy; nevertheless if data or information is common and important, immediately will be shared among all the operators.

Of course Operative Centres, both at technological and behavioural level, have to respect citizenship:

- Taking care privacy laws at local and national level.
- Returning data and information to citizens and companies to use for their:
  - Safety and security.
  - Movements and travels.

The O.C. developed for LAICA will be distributed in different strategic points of the AmI net in Reggio Emilia; it has been installed at the Police control centre and another installation will be distributed in the offices at the Reggio Emilia municipality. These O.C.s will be managed by authorized users to have (in real-time) all the available information about the city. Moreover, remote connections will be established for authorized private citizens in order to have punctual, local and global view of the AmI in Reggio Emilia.

3. Distributed Surveillance for Urban Scenarios

Among the sensors used in LAICA, cameras and computer vision applied on them are the highest level, since they can provide high-level semantics of the scene. Besides the use for traffic analysis and control, cameras are used in LAICA for the surveillance of public parks. In particular, the objectives of this prototype are mainly two: first, to provide access to videos from public parks to citizens, second, to (semi-)automatically analyze the park's scene for detecting dangerous situations (such as people's falls or suspicious behaviours) in order to help human operators in controlling the park scene for security and safety purposes.

This paper basically reports on the first objective. An interesting way of presenting to the public the scene from a wide-area park is that of *mosaic images*. A mosaic image is the result of the registration [1] of overlapped images to create a continuous description of the scene. Once the mosaic image is computed, the dynamic current scene of the park can be present in a whole, one-view image. However, in order to make the videos from public places available to citizens (for example, through an Internet web site), privacy issues must be taken into account. Most of the western countries have a set of more or less restrictive laws to assure the respect of citizens' privacy. For this reason, there is the emergent need for (semi-)automatic tools for protecting people's identity, especially in public surveillance. A possible solution is that of using more "blind" sensors, such as PIR (Passive InfraRed) or proximity sensors to detect the presence of people. However, these sensors do not provide high-level information. An alternate solution relies on the capability given by computer vision algorithms to recognize humans and, specifically, their faces. Artificially obscuring faces is an effective way to protect identities and, at the same time, save the face images for further, authorized accesses for security purposes. Unfortunately, even though face detection techniques are now mature, it can be hard to have a frontal view of the face (necessary for face detection and possible recognition) in complex environments. To help in this task, multiple cameras can be used in order to obtain different views and, potentially, at least a frontal view of the face. In addition, a multi-camera vision system enables the coverage of wider areas and the multiple viewpoints provide an effective solution to the problem of occlusions in cluttered scenes. The merge of the data provided by multiple cameras poses, however, some problems too. A main problem is that the identity of the objects moving from one camera to another must be preserved, in order to analyze their behaviours over the whole scene. This process is known in the literature as *"consistent labelling"* and becomes challenging when cameras can not be manually calibrated. Additionally, maintaining consistent labels for individuals in the scene is a key process for automated surveillance and scene analysis (see above).

To test our algorithms we created a test bed on our campus, installing four partially overlapped cameras (three fixed and one PTZ), in a zone where many people are passing through, there are some benches and the illumination conditions are typical of an outdoor environment. The cameras have been provided within the LAICA project by the Wavelet Technology Italia SrL that is one of the industrial partners of the project. In Fig. **3** some snapshots of the partial views that can be acquired by the overlapped cameras are shown. Figs. **3**.a, **3**.b, and **3**.c are acquired by three cameras with overlapped

views; Fig. **3**.d corresponds to the PTZ cameras that is partially overlapped with all the three fixed cameras.

(a) (b)

(c) (d)

Fig. 3. Snapshots acquired from the four cameras of the test bed on the Modena's campus.

3.1. Image mosaicing through homography

Image mosaicing can have two different scopes: the first is to create a dynamic background model of the environment that can be exploited by the automatic surveillance system to detect moving objects and enables consistent labelling establishment; the second scope is to create a more natural and immediate representation capable to provide a single overview of the environment. Given a set of images, image mosaicing computes a transformation (affine or projective) to map the images onto the plane of a reference image. The precise geometrical transformation yields to the image registration, that must be followed by a process of error correction to smooth the non-idealities (also known as image blending).

A large number of different approaches to image mosaicing have been proposed [1]. The methods can be classified in direct methods [2, 3] and feature-based methods [4, 5, 6]. Both of these have their pros and cons. The direct methods usually attempt to iteratively estimate the transformation parameters by minimizing an error function based on the intensity difference in the area of overlap. The advantage of the direct methods is that very accurate registration can be achieved since all the available data is used. However, direct methods are not very robust against illumination variations because of the nature of the error function to be minimized. Furthermore, the presence of moving objects in the scene can cause serious problems because all the pixel values are taken into account. Instead of using all the available data, feature-based methods try to establish feature point correspondences between the images to be registered. Most of the existing feature-based mosaicing methods use point features such as corners to find the feature points from the images. After the points have been found, they are matched by utilizing a correlation measure in the local neighbourhood. By selecting appropriate features, these methods can be done very robust against illumination variations. Also, their tolerance against image rotation and zooming is usually better compared to the direct methods. Furthermore, the scenes with moving objects can be handled robustly by detecting and removing outlier feature points with appropriate methods.

Considering that in our case we want to create a mosaic between images acquired from coupled overlapped fixed cameras, there is only the need of a one-time calibration of the pairs of overlapped

cameras in order to find a transformation onto a common plane, i.e. an homography. Homography can be obtained on the ground plane (z=0) by means of *Entry Edges of Field of View* (E^2FoV) defined in [7]. E^2FoVs are computed in an offline supervised process in which a single person moves into the scene of coupled overlapped cameras, trying to span between all the borders of the overlapped views. The correspondence of feet's position (easily obtained as the lower support point of the foreground blob) in the two views makes available pairs of points in the two ground planes and, thus, the construction of the E^2FoV lines, i.e. the lines in one camera image corresponding to the border of the field of view in the other camera. Our approach modifies the Edge of Field of View proposed by Khan and Shah [8] by taking into account only people completely entered in the scene.

For two overlapped cameras C^i and C^j, the training procedure computes the overlapping areas $Z^{i,j}$ and $Z^{j,i}$. The four corners of each overlapping area are on the same plane z = 0 and they are sufficient to compute the homography matrix $H^{i,j}$ from camera C^i to camera C^j. Obviously, the matrix $H^{j,i}$ can be easily obtained with the equation $H^{j,i} = \left(H^{i,j}\right)^{-1}$.

As introduced earlier, mosaicing onto the ground plane can be exploited for two goals: to solve the consistent labelling problem and to enlarge the view of the environment for remote users. To establish the consistent labelling (in order to keep identity of the people moving in the whole scene), each time a new object is detected in the overlapping area of a camera, its support point is projected in the overlapped camera by means of the homographic transformation. The obtained coordinates could not correspond to the support point of an actual object. For the match with the new object we select the object whose support point is at the minimum Euclidean distance in the 2D plane from these coordinates [7].

Then, the computed homographic transformation is adopted to register the two overlapped views on the ground plane. Each point of the registered image is re-mapped into the coordinates of the new mosaic image by means of the homography. An example on a public park in Reggio Emilia is shown in Fig. **4**. Figs. **4**.a and **4**.b are two different views of the same area taken from two of the hundred of cameras connected to the operative centre. The constructed mosaic image (Fig. **4**.c) gives a new enlarged view of the scene.

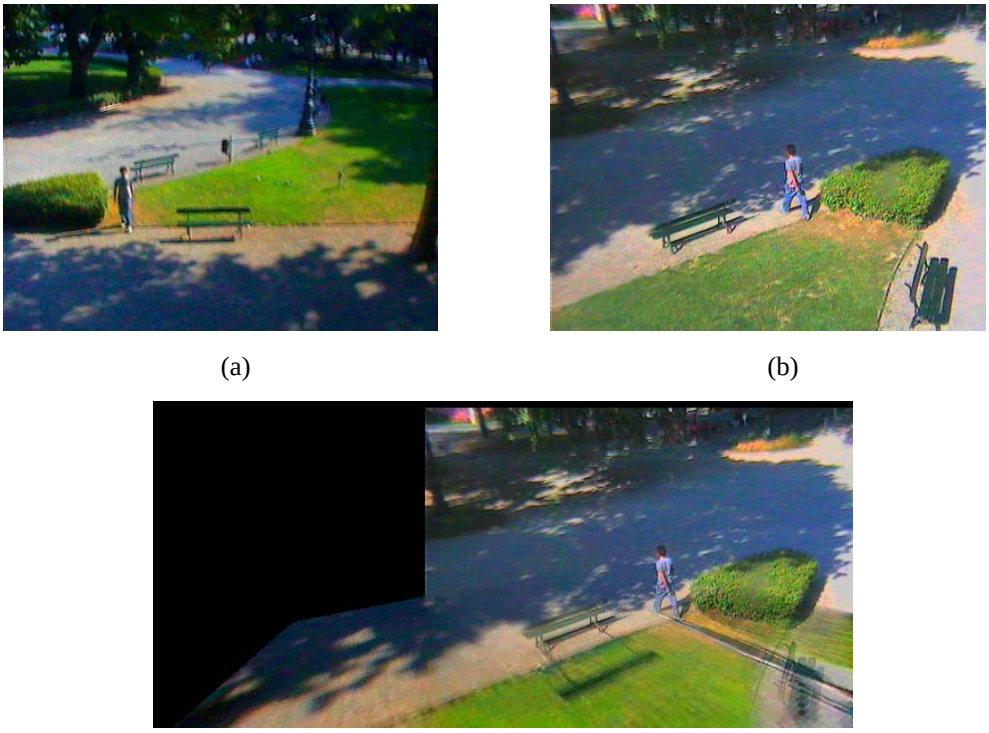
Unfortunately, this approach relies on the presence of sufficiently large and almost uncluttered ground plane overlapping between the two views. In Fig. **5** an example with occluded and limited overlapping on the ground plane and the resulting mosaic image are shown. It is evident that, in this case, the resulting mosaic image is unsatisfactory.

3.2. Face detection and tracking for identity protection

Video streams coming from single or multiple overlapped cameras are mainly used for people surveillance; people must be detected and tracked [9, 10]. An important task is the face detection to determine, whenever it is possible, the people's identity for security purposes, or, in a dual manner, to hide the identity for privacy issues.

Face detection is a widely explored research area in computer vision. Two recent surveys, Yang et al. [11] and Hjelm and Low [12], collect a large number of proposals about face detection. Most of them are based on a skin colour detection (Jones and Rehg [13]) followed by a face candidate validation achieved exploiting geometrical and topological constraints.

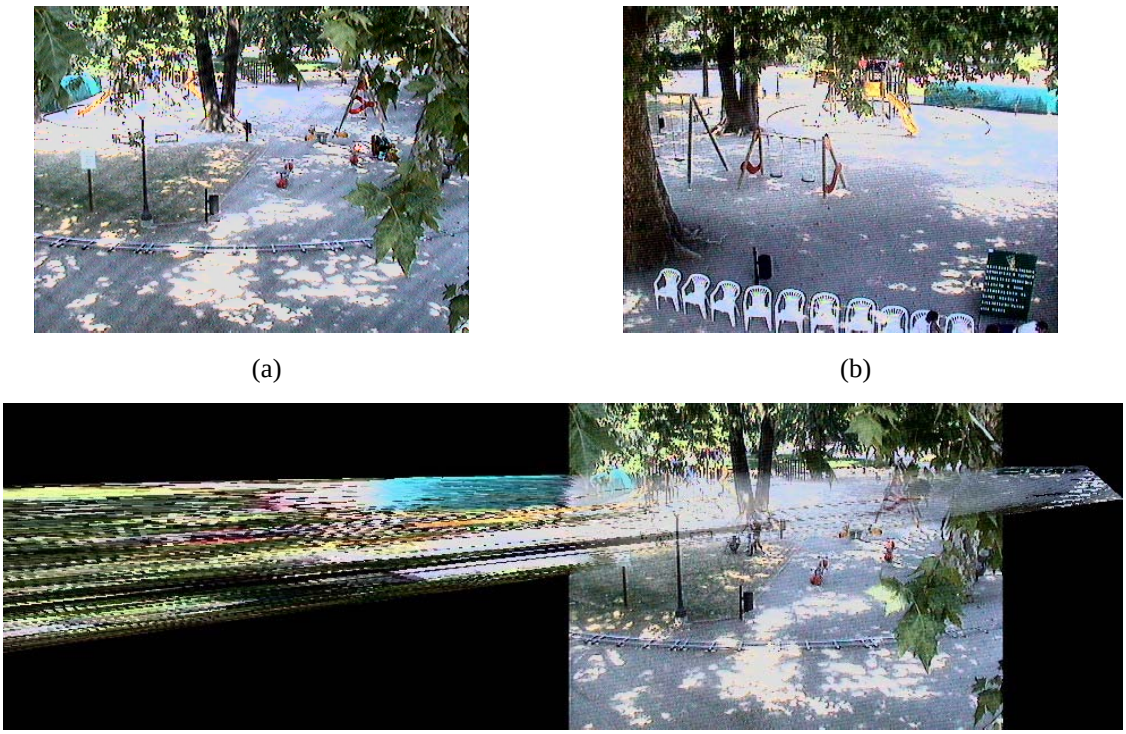
Unfortunately, most of the colour-based approaches are very expensive from the computational point of view and it is impossible to perform accurate face detection at every frame in a real time video surveillance application. To solve this problem, the face detection can be performed only when a new person enters the scene and then a face tracking can be adopted. The developed method exploits and improves the best ideas proposed in [14] and [15]. The first uses both colour and gradient information but the search of the head is limited to a neighbourhood of a predicted position. The problem of this solution is that it needs a frame rate too high to make reliable predictions. Instead, [15] adopts a solution based on the elliptical Hough transform; differently from the previous, this solution does not require any tracking nor prediction, because the processing of each frame is stand-alone. A face colour histogram must be available as a model. To this aim, a supervised learning phase is performed to compute a histogram of skin and hair colours. Thus, for each tracked object, two different Hough transforms are computed: one gradient-based and one colour-based. The points belonging to the edges of the track (obtained with Sobel edge detectors) vote for the first transform according with the gradient value. The selection of the voted



(a) (b)

(c)

Fig. 4. An example of mosaic image from a public park in Reggio Emilia: (a) and (b) report the original images, while (c) shows the mosaic obtained through homography.



(a) (b)

(c)

Fig. 5. An example of mosaic image from a public park in Reggio Emilia with a not-sufficient overlapping: (a) and (b) report the original images, while (c) shows the mosaic obtained through homography.

pixels is done by moving on the image in the same direction of the gradient with a distance obtained from the estimated head size.

Similarly, a point of the object votes for the colour-based transform if its colour has a non-zero value on the histogram of the saved head model. In this case, it votes for all the points inside an ellipse having the same size of the head and the actual pixel as the centre, and the rate is proportional to the model histogram value corresponding to the colour of the pixel. After that, the two transforms are normalized and multiplied pixel-by-pixel to obtain a single map that contains both colour and gradient information. The point with the higher value is chosen as the head centre. Once the face is detected and tracked, the head can be obscured, as shown in Fig. **6**(a).

(a) (b)

Fig. 6. Examples of face obscuration (a) and avoidable face obscuration (b).

4. Discussion and conclusions

In the LAICA project, our research attempts to exploit cameras for safety and security purposes. Citizens' safety can be improved allowing single users to connect to the network of sensors, observing the state and conditions. In particular, cameras are a plenty source of information. However, simple webcams are not sufficient, since they provide a single (thus limited) point of view and pose privacy-related problems. In this paper, we propose to adopt a plethora of cameras both to cover wider areas and to allow a more advance survey of the scene. Mosaic images are created to provide enlarged (thus enriched) view of the environment to the citizens, while automatic face obscuration is exploited to take privacy issues into account.

Citizens' security, instead, is improved adopting real-time intelligent surveillance systems. The cameras are not only connected to the operative centre, but the people detection and tracking module is capable of extracting the identity of people. Consistent labelling is necessary to follow the people in the trajectory in wide places as a public park (Fig. **7** shows a bird-eye view of three overlapped cameras in our test bed).

Acknowledgments

This work was supported by the project L.A.I.C.A. (Laboratorio di Ambient Intelligence per una Città Amica), funded by the Regione Emilia-Romagna, Italy. We wish to thank all the partners of the project for their permission to publish this paper.

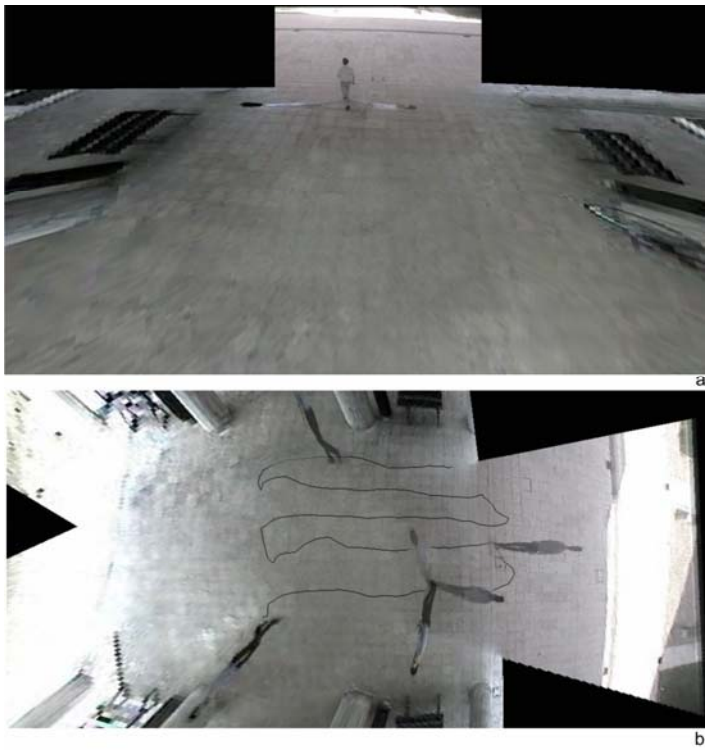


Fig. 7. Mosaic image (a) and bird-eye view (b) of three overlapped cameras in our test bed.

References

[1] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977-1000, 2003

[2] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *5th International Conference on Computer Vision*, 605–611, 1995.

[3] R. Szeliski. Image mosaicing for tele-reality applications. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 44–53, 1994.

[4] M. Brown and D. Lowe. Recognising panoramas. In *Ninth IEEE International Conference on Computer Vision*, 2:1218–1225, 2003.

[5] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 885–891, 1998.

[6] R. Marzotto, A. Fusiello, and V. Murino. High resolution video mosaicing with global alignment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:692–698, 2004.

[7] S. Calderara, R. Vezzani, A. Prati, R. Cucchiara. Entry Edge of Field of View for multi-camera tracking in distributed video surveillance, in press on *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2005.

[8] S. Khan, M. Shah. Consistent labelling of tracked objects in multiple cameras with overlapping fields of view, *IEEE Trans. on PAMI*, 25(10):1355–1360, 2003.

[9] R. Cucchiara, C. Grana, M. Piccardi, A. Prati. Detecting Moving Objects, Ghosts and Shadows in Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337-1342, 2003

[10] R. Cucchiara, C. Grana, G. Tardini. Track-based and object-based occlusion for people tracking refinement in indoor surveillance. in *Proc. of ACM 2nd International Workshop on Video Surveillance & Sensor Networks*, 81-87, 2004

[11] M. Yang, D. Kriegman, N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. on PAMI*, 24(1):34–58, 2002.

[12] E. Hjelm, B. Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83(3):236–274, 2001.

[13] M. Jones, J. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46:81–96, 2002.

[14] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. in *Proc. of IEEE Computer Vision and Pattern Recognition*, 232–237, 1998.

[15] D. Maio, D. Maltoni. Real-time face location on gray-scale static images. *Pattern Recognition*, 33 (9):1525–1539, Sept. 2000.