Group Detection at Camera Handoff for Collecting People Appearance in Multi-camera Systems

Simone Calderara, Rita Cucchiara
D.I.I. - University of Modena
and Reggio Emilia

Andrea Prati
D.I.S.M.I. - University of Modena
and Reggio Emilia

Abstract

Logging information on moving objects is crucial in video surveillance systems. Distributed multi-camera systems can provide the appearance of objects/people from different viewpoints and at different resolutions, allowing a more complete and precise logging of the information. This is achieved through consistent labeling to correlate collected information of the same person. This paper proposes a novel approach to consistent labeling also capable to fully characterize groups of people and to manage miss segmentations. The ground-plane homography and the epipolar geometry are automatically learned and exploited to warp objects' principal axes between overlapped cameras. A MAP estimator that exploits two contributions (forward and backward) is used to choose the most probable label configuration to be assigned at the handoff of a new object. Extensive experiments demonstrate the accuracy of the proposed method in detecting single and simultaneous handoffs, miss segmentations, and groups.

1. Introduction and Related Works

In the new generations of surveillance systems two requirements become crucial: to be distributed and to collect multimedia surveillance data. The first requirement answers to the need of covering wide areas. Multiple fixed cameras with partially overlapped FoVs (Field of Views) are often a good solution; in addition, they help in solving occlusions by providing different viewpoints. The second requirement, instead, is driven by emerging applications in video surveillance, such as video post-analysis for forensic purposes, retrieval of events, people, or faces from videos, and multimedia database creation. All these applications require to extract semantics from the video. While at event level the useful semantics heavily depends on the application, at the object level, most of the time, it is represented by the appearance of the moving objects. Having more cameras that look at the same scene enables to collect the appearance of the same object/person from different viewpoints and at different resolutions.

In multiple people surveillance this task poses several challenges. First, in each camera system, each moving person must be tracked, i.e. its label must be kept consistent along time. Tracking people in single camera systems is a topic well studied in the literature; in this context, appearance-based approaches [4, 10] should be adopted to extract, at each frame, the visual aspect of the objects without the background. A second challenge is to distinguish between a single person and groups of people.

Once people are detected and tracked from each single view, their identity must be kept consistent among overlapped views, by assigning the same label to different instances of the same person. The approaches to consistent labeling can be generally classified into three main categories: appearance-based (where the matching is based essentially on the color or texture of the objects) [9, 8], geometry-based (where geometrical relations and constraints between the different views are exploited) [6, 7] and mixed approaches (where information about the geometry and the visual appearance are combined by means of probabilistic information fusion [5] or Bayesian Belief Networks (BBN) [3]).

This consistency allows to correlate with the same person the visual information provided by different views. Additionally, multiple views can be also employed to solve miss segmentations and groups by exploiting redundant data provided by multiple cameras.

This paper proposes a novel approach to consistent labeling also capable to fully characterize groups of people and to manage miss segmentations.

2. System Overview

The proposed approach employs homography and epipolar geometry to solve ambiguities in the matching of people in different synchronized camera systems. The considered distributed multi-camera system is composed by a generic number n of cameras $C = \{C_1, ..., C_n\}$ so that for each C^i

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

it exists at least a camera C^j whose FoV is overlapped to that of C^i.

A first stage of offline computation of the homographic transformation (on the ground plane $z = 0$) and of the epipolar constraints must be carried out. Our approach does not require manual setups and uses, as training process, the information computed from a video containing a single person moving freely in the scene covered by the system. In this way, the homographic transformation between the projections of the ground plane in two overlapped views can be obtained exploiting the Entry Edge of Field of View (E^2oFoV) proposed in [2]. Here, we also propose to recover the epipole location during the same training phase exploiting parallax property of perspective images and RANSAC optimization technique.

Given a 3D plane Π and its projections on the camera C^i and C^j (π^i and π^j, respectively), the relation between them is pointed out by the homographic matrix H. The parallax property of projective images is exploited to compute epipole location using a single plane. Given a 3D-point M_k not laying on the plane Π, and its projection $\mathbf{m_k^i}$ on C^i, it is possible to find two true correspondences in the image plane of C^j. The former is the real projection of M_k on C^j, $\mathbf{m_k^j}$, while the latter is the point in C^j computed through the homographic transformation H given the hypothesis that $\mathbf{m_k^i}$ lies on the plane π^i. The line computed from these points must be an epipolar line since it passes through the images of the same point of image plane I^i. Given at least two lines, the epipole can be located as the intersection of these lines:

$$e^j = \min_{p^j} \left(\sum_k d^2 \left(p^j, l^j_{\mathbf{m_k^i}} \right) \right) \tag{1}$$

where $l^j_{\mathbf{m_k^i}} = \langle \mathbf{m_k^j}, H\mathbf{m_k^i} \rangle$ and $d(.)$ is a L_2 distance, e.g. the Euclidean distance. After epipole computation, the fundamental matrix is obtained as:

$$F = \left[e^j \right]_\times H \tag{2}$$

During the training phase, the knowledge of having a single person moving in the scene is exploited to collect as many points (and relative correspondences) as necessary for both E^2oFoV creation and epipolar geometry recovery. While for constructing the E^2oFoV lines, corresponding points laying on the ground plane are necessary [2], for recovering the epipolar geometry we must obtain point correspondences *not* on the ground plane. To achieve this, the lower and the upper support point (lp and up) are computed. Assuming a standing posture of the person, the lower support point is the middle point of the bottom of the bounding box. Upper support point, instead, is the highest point of the blob. Unfortunately, using the upper support point to compute the epipolar line intersection by means of LSQ can be strongly unstable. To overcome this problem we exploit RANSAC technique, to evaluate epipole location, that has been proved to be more reliable since a high number of redundant information can be accounted.

After this training phase, the system is ready to maintain system's consistency when camera handoff occurs. It is assumed that a system for people segmentation and tracking from single camera, reliable enough to be used for many people walking and interacting each other, is available. Any tracking algorithm can be used. No information about color or texture are necessary. The only requirement is the extraction of the object's appearance or, at least its silhouette, to compute the head and feet position.

3. Bayesian competitive consistency resolution

The online part of the system includes the consistency solver. To ensure system consistency we define a *maximum-a-posteriori* (MAP) estimator choosing the most probable label configuration to be assigned to new object. From this perspective, for an object appeared on camera C^l at handoff time and identified by new, a hypotheses' space Γ must be created for each overlapped camera, considering all the possible label assignments. Assuming that FoV of camera C^i is overlapped to that of camera C^l, hypotheses' space $\Gamma^{l,i}(\tau^l_{new})$ consists of all the possible combinations of candidate objects.

For each single hypothesis in hypotheses' space $\Gamma^{l,i}(\tau^l_{new})$, posterior probability is evaluated exploiting the Bayes rule where the hypothesis itself is considered as a single partition of the full hypotheses' space. Indicating with s^l_{new} the spawning event and with $\gamma^{l,i}_k$ the k^{th} considered hypothesis in $\Gamma^{l,i}(\tau^l_{new})$ space, the following equation is used to compute posterior:

$$P\left(\gamma^{l,i}_k \mid s^l_{new}\right) = \frac{P\left(s^l_{new} \mid \gamma^{l,i}_k\right) P\left(\gamma^{l,i}_k\right)}{\sum_{k=1}^n P\left(\gamma^{l,i}_k\right) P\left(s^l_{new} \mid \gamma^{l,i}_k\right)} \tag{3}$$

3.1. Prior Computation

To compute the prior probability of $\gamma^{l,i}_k$, we exploit the homographic mapping to warp the lower support point of each candidate object in the image plane of the camera C^l. A hypothesis $\gamma^{l,i}_k$ consisting of a single person should have higher prior probability if the warped lower support point is far enough from the other objects' support points. On the contrary, a hypothesis consisting of two or more objects (i.e., a possible group) will gain higher prior if the objects composing it will result close each other after the warping, and, at the same time, the whole group is far from other objects. Practically, clutterness or isolation of warped people's lower support points are considered as a discriminant element. The local search space, $\Sigma^{l,i}(\tau^l_{new})$, contains the set

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

of objects detected in C^i and located in the overlapping area between the two views, on C^i and C^l respectively. These objects are candidate to be matched with τ^l_{new}.

Priors account for a score σ assigned on the base of two contributions:

$$\sigma\left(\gamma_k^{l,i}\right) = \min_{\tau_a^i,\tau_b^i \in \Sigma^{l,i}(\tau^l_{new})\ a\neq b} Od\left(\tau_a^i,\tau_b^i\right) - \max_{\tau_c^i,\tau_d^i \in \Sigma^{l,i}(\tau^l_{new})\ c\neq d} Id\left(\tau_c^i,\tau_d^i\right) \quad (4)$$

where Od is called *outer*-hypothesis distance, Id *inner*-hypothesis distance.

The inner-hypothesis is computed as the distance between warped lower support point of two objects belonging to the same hypothesis $\gamma_k^{l,i}$:

$$Id(\tau_a^i,\tau_b^i) = \begin{cases} \left\| \left(H^{i,l}\mathbf{lp_a^i}\right) \times \left(H^{i,l}\mathbf{lp_b^i}\right) \right\| & \text{if } \{\tau_a^i,\tau_b^i\} \in \gamma_k^{l,i} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

with $H^{i,l}$ the homography matrix, while the outer-hypothesis distance is defined as the distance between a hypothesis' warped lower support point and a warped point belonging to an object inside local search space but disjoint with considered hypothesis set:

$$Od(\tau_a^i,\tau_b^i) = \begin{cases} \left\| \left(H^{i,l}\mathbf{lp_a^i}\right) \times \left(H^{i,l}\mathbf{lp_b^i}\right) \right\| & \text{if } \{\tau_a^i,\tau_b^i\} \in \Sigma^{l,i}(\tau^l_{new}) \wedge \left(\tau_a^i \in \gamma_k^{l,i} \wedge \tau_b^i \notin \gamma_k^{l,i}\right) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

3.2. Likelihood Computation

Likelihood is evaluated considering a given hypothesis on the new object's labeling and testing its fitness against current evidence. Exploiting the homography and the epipolar constraints, the principal inertial axis of the objects in each camera system can be warped on the overlapped views, supposed the objects are in the overlapping area. In absence of tilt angle, when camera's retinal plane is orthonormal to the ground plane, inertial axis of tracks appears as a straight line parallel to the lateral border of the image. This is not true for general setups, but the correct axis orientation can be obtained by computing the position of the camera's vertical vanishing point vp. The vanishing points are computed with the procedure described in [1]. In this way, the warped axis is obtained as the segment between the warped lower support point lp and the intersection with the epipolar line.

Likelihood is made up of two contributions: *forward* and *backward*. The forward probability related to hypothesis $\gamma_k^{l,i} \in \Gamma^{l,i}(\tau^l_{new})$ is computed through the warping of the principal inertial axis of new object τ^l_{new} appeared in the image plane of source camera C^l at handoff-time t:

$$P\left(\tau^l_{new} \mid \gamma_k^{l,i}\right)_{forward} = \frac{\sum\limits_{\tau_h^i \in \gamma_k^{l,i}} \frac{\sum_{(x,y)} \varrho\left(x,y,\tau^l_{new},\tau_h^i\right)}{d(as(\tau^l_{new},i),ae(\tau^l_{new},i))}}{card\left(\Sigma^{l,i}(\tau^l_{new})\right)} \quad (7)$$

where

$$\varrho\left(x,y,\tau^l_{new},\tau_h^i\right) = \begin{cases} 1 & \text{if } (x,y) \in (FG(\tau_h^i) \cap \\ & \langle as\left(\tau^l_{new},i\right), ae\left(\tau^l_{new},i\right)\rangle) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

FG is the foreground blob of the object, and as and ae are, respectively, warped axis start point and end point defined as:

$$as\left(\tau^l_{new},i\right) = H^{l,i}\mathbf{lp}^l_{new} \quad (9)$$

$$ae\left(\tau^l_{new},i\right) = \begin{cases} \langle \mathbf{e^i}, H^{l,i}\mathbf{up}^l_{new}\rangle \\ \langle \mathbf{vp^i}, H^{l,i}\mathbf{lp}^l_{new}\rangle \end{cases} \quad (10)$$

Backward probability is computed similarly to the previous contribution, except that warping source and destination camera are swapped:

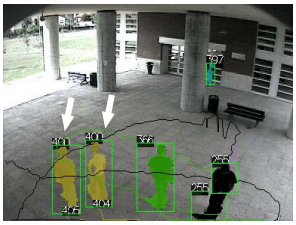
$$P\left(\tau^l_{new} \mid \gamma_k^{l,i}\right)_{backward} = \frac{\sum\limits_{\tau_h^i \in \gamma_k^{l,i}} \frac{\sum_{(x,y)} \varrho\left(x,y,\tau_h^i,\tau^l_{new}\right)}{d\left(as\left(\tau_h^i,l\right),ae\left(\tau_h^i,l\right)\right)}}{card\left(\Sigma^{l,i}(\tau^l_{new})\right)} \quad (11)$$

At the end, the maximum of the two contributions is computed and taken as likelihood value. An example of the effectiveness of the double backward/forward probability is the full characterization of groups of people. Forward contribution, in fact, is useful to detect groups on overlapped views exploiting subsequent creations of new objects, that is the case in which a group is already inside the scene and group's components appear one at a time in another view. This situation will be referred as *"group inside"* in the following. In particular, after warping, each axis of the new objects will intersect group's foreground blob on the chosen camera. This information can be exploited to characterize the new objects as belonging to the existing group. An example is reported in Fig. 1, where the group labeled with 400 in camera C^2 (Fig. 1(b)) is detected as two separate people (id #404 and #405) in camera C^1 (Fig. 1(a)).

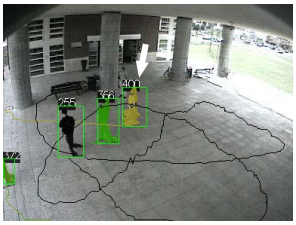
On the other hand, backward contribution is useful to solve the case of *"group enter"*, in which people appear as a new single object in a camera, being detected as separated in at least another camera with overlapped FoV, for instance a single object (actually composed of two people as in Fig 1(c)) appears at frame 432 on camera C^2 (Fig. 1(d)). The two people are detected (and tracked) as separated in camera C^1 with labels 32 and 36. The backward contribution assures that the axes of these two objects in C^1 intersect the blob in C^2 with the highest probability. As a consequence,

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

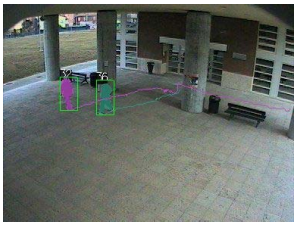
both the label 32 and 36 are assigned to the group in Fig. 1(d).



(a) C^1 at frame #459 of video GV2



(b) C^2 at frame #459 of video GV2



(c) C^1 at frame #432 of video GV1

(d) C^2 at frame #432 of video GV1

Figure 1. Examples of full groups' characterization

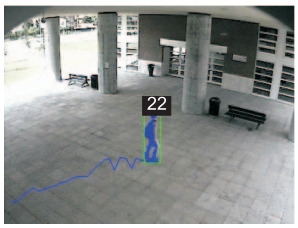
3.3 MAP label identifier

For each hypothesis $\gamma_k^{l,i}$ in hypotheses' space, the prior and the likelihood are evaluated using equations (4), (7) and (11). The marginal probability remains constant for a given hypotheses' space, thus it is discarded in maximization process. After finding, in each overlapped view, the most probable hypothesis, competition among cameras is performed to obtain the best view, in the sense of selecting the camera with the best representation of the scene in terms of likelihood. The Bayes theorem is used, where conditional probability represents the probability of choosing a given camera, under the observation of cameras' *a posteriori* best hypothesis:

$$P\left(C^i \mid s_{new}^l\right) = \frac{P\left(s_{new}^l \mid C^i\right) P\left(C^i\right)}{\sum_{k=1}^{n} P\left(C^k\right) P\left(s_{new}^l \mid C^k\right)} \qquad (12)$$

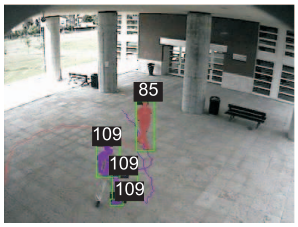
with $P\left(s_{new}^l \mid C^i\right) = \max_{\gamma_k^{l,i} \in \Gamma^{l,i}(\tau_{new}^l)} P\left(s_{new}^l \mid \gamma_k^{l,i}\right)$.

MAP approach is then exploited to find the camera having the best hypothesis representation. In maximum search, priors have all the same value considering all the overlapped camera views equally probable. Eventually, after these two maximization steps, the label is assigned to new object according to the selected hypothesis. If the new object is identified as a group, i.e. the chosen hypothesis consists of more than one object, the label set of objects composing the group is assigned as identifier.

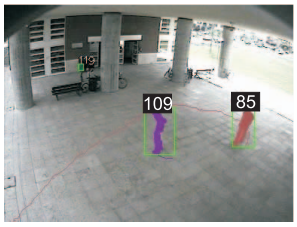


(a) C^1 at frame #724 of video MV1

(b) C^2 at frame #724 of video MV1



(c) C^1 at frame #2881 of video MSV



(d) C^2 at frame #2881 of video MSV

Figure 2. Examples of bad segmented object and consistency recovery when an object is split in several pieces or not fully detected.

4. Experimental Results

Name	Frame number	Duration
Single Handoff Video (SHV)	12600	14 min
Miss segmentation Video (MSV)	16200	18 min
Group Video 1 (GV1)	9900	11 min
Group Video 2 (GV2)	14400	16 min
Mixed Video 1 (MV1)	22500	25 min
Mixed Video 2 (MV2)	9000	10 min

Table 1. 90-minutes benchmark of videos used for the test

The presented system is currently working in a setup of four cameras (three fixed CCD cameras and a PTZ camera). Tests have been performed with videos acquired during ordinary work days in different environmental conditions. The controlled area is an open space where people can enter from each side and walk freely in each direction. In acquiring the videos, no constraints have been imposed on people's trajectories or behaviors. Segmentation with background suppression and the appearance based tracking proposed in [4] are adopted to extract people appearance and follow the person's movement. Six videos (for a total of 90 minutes), summarized in Table 1, are choosen as benchmark: video SHV)contains only camera handoff of one per-

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

son at a time, while video MSV contains several segmentation errors due to the poor illumination conditions and to the appearance of the people difficult to distinguish from the background (these segmentation errors result in people not completely detected or split in more pieces, as shown in Fig. 2 ((a)-(d))); videos GV1 and GV2 include several situations in which groups of people are formed (GV1 contains both "group enter" and "group inside" situations, while video GV2 includes only the first one, but contains also some *simultaneous handoffs*, i.e. the case in which more people enter simultaneously in the overlapping zone; eventually, videos MV1 and MV2 report a mixture of all these situations and are denoted by a high degree of complexity.

The achieved results are reported in Table 2. This table highlights the performance of the system with respect to the different situations mentioned above. Moreover, the last column reports the overall accuracy of the system. The table reports the results obtained using only the pure homography-based method described in [2], that bases the object matching process on the support points' distances measured on the inter-camera ground plane homographic mosaic, compared with the results with our Bayesian competitive approach with each single contribution (forward and backward) and with both.

In Fig. 2(b) (frame #724 of video MV1) the person on C^2 is not fully detected and a method based on ground plane homography will fail to ensure consistency. This is testified by the average accuracy of the pure homography-based method on solving miss segmentations (see Table 2) that is as low as 41.48% (56 correctly assigned labels over 135). The use of forward and backward likelihood contributions can overcome these segmentation errors (see the correct label assignment in Fig. 2(b)). Similarly, in Fig. 2(c) (frame #2881 of video MV1) the person labeled with 109 is split on C^1 in three different objects. The proposed method solves this bad segmentation issue assigning the same label to all the pieces by exploiting the information coming from the overlapped view (Fig. 2(d)). Indeed, the average accuracy increases to 46.67%, 85.93%, and 96.30% (130 correctly assigned labels over 135), using, respectively, only backward contribution, only forward contribution, or the complete MAP estimation.

Group detection can be often solved thanks to the warping of useful information between overlapped views. Results on group detection (columns 6 to 9 in Table 2) demonstrate that each of the two contributions in the MAP estimation concurs to solve different situations. In particular, the backward contribution is more suitable in assigning correct labels in the case of "group enter", doubling the average accuracy from less than 50% to a 100% of accuracy. Given the implicit duality of the problem, in the case of "group inside" (columns 8 and 9 of Table 2) the most important contribution is the forward one, able to increase from 43.96% of average accuracy (using only backward contribution) to 94.51% using only forward contribution, to 95.60% using the complete MAP estimation.

Eventually, Fig. 3 shows an example of the system working on three cameras in a complex situation and its use of consistent labeling for collecting the appearance of the moving people. For each synchronized frame, moving people are correlated among views and inserted in a database for further processing and retrieval. The figure shows that when a person is visible from more cameras' views its appearance is taken at different resolutions and viewpoints (e.g., people with id #115 and #242).

The proposed system works at a frame rate of about seven fps for each camera. The benchmarks have been carried out using three cameras connected on a desktop personal computer P4 at 3GHz with 1GB RAM.

References

[1] C. Brauer-Burchardt and K. Voss. Robust vanishing point determination in noisy images. In *Proc. of Int'l Conference on Pattern Recognition*, volume 1, pages 559–562, 2000.

[2] S. Calderara, R. Vezzani, A. Prati, and R. Cucchiara. Entry edge of field of view for multi-camera tracking in distributed video surveillance. In *Proc. of IEEE International Conference on Advanced Video and Signal-Based Surveillance*, pages 93–98, 2005.

[3] S. Chang and T.-H. Gong. Tracking multiple people with a multi-camera system. In *Proc. of IEEE Workshop on Multi-Object Tracking*, pages 19–26, 2001.

[4] R. Cucchiara, C. Grana, G. Tardini, and R. Vezzani. Probabilistic people tracking for occlusion handling. In *Proc. of International Conference on Pattern Recognition (ICPR 2004)*, volume 1, pages 132–135, Aug. 2004.

[5] J. Kang, I. Cohen, and G. Medioni. Continuous tracking within and across camera streams. In *Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–267 – I–272, 2003.

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

[7] A. Mittal and L. Davis. M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. *IJCV*, 51(3):189–203, February 2003.

[8] K. Nummiaro, E. Koller-Meier, T. Svoboda, D. Roth, and L. Van Gool. Color-based object tracking in multi-camera environments. In *DAGM03*, pages 591–599, 2003.

[9] J. Orwell, P. Remagnino, and G. Jones. Multi-camera colour tracking. In *Proc. of Second IEEE Workshop on Visual Surveillance, (VS'99)*, pages 14–21, June 1999.

[10] A. Senior. Tracking people with probabilistic appearance models. In *Proc. of Int'l Workshop on Performance Evaluation of Tracking and Surveillance (PETS) systems*, pages 48–55, 2002.

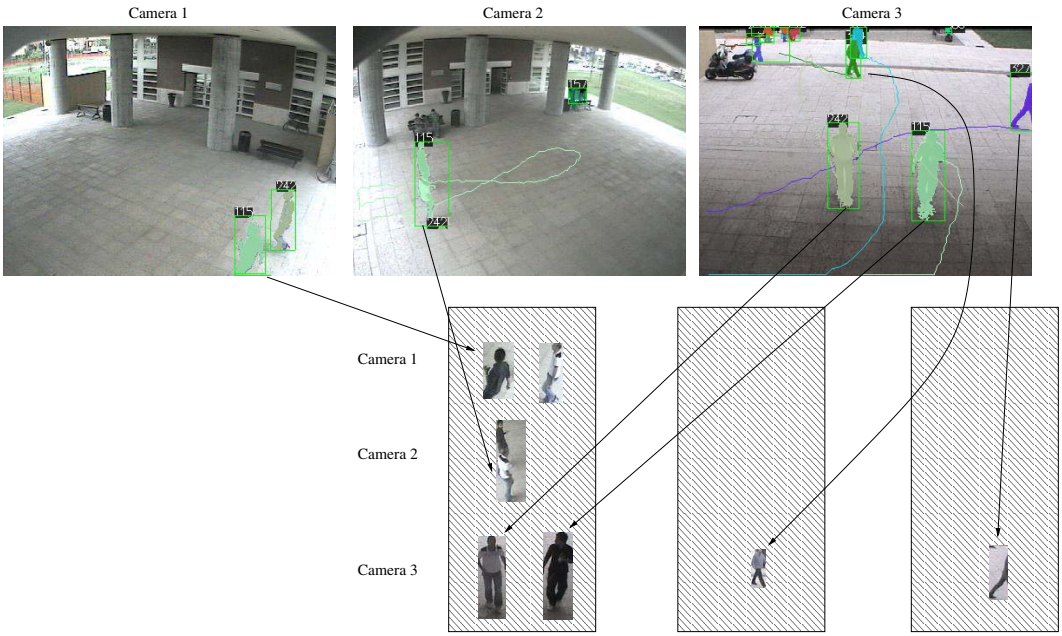


Figure 3. Example of multi-camera extraction of meaningful snapshots of moving people.

	# Single HO		# Miss sgm.		# Grp enter		# Grp insd.		# Simul. HO		Overall
	total	correct	total	correct	total	correct	total	correct	total	correct	Accuracy
Only Homography (Average overall accuracy *77.40%*)											
SHV	137	128	0	0	0	0	0	0	0	0	93.43%
MSV	28	28	76	30	0	0	7	7	0	0	58.56%
GV1	30	29	0	0	15	5	40	39	0	0	85.88%
GV2	78	73	0	0	35	16	0	0	19	18	81.06%
MV1	39	34	21	10	30	15	18	14	14	13	70.49%
MV2	52	50	38	16	30	15	26	24	18	18	75.00%
Only Forward Contribution (Average overall accuracy *80.63%*)											
SHV	137	135	0	0	0	0	0	0	0	0	98.54%
MSV	28	28	76	34	0	0	7	7	0	0	62.16%
GV1	30	30	0	0	15	6	40	40	0	0	89.41%
GV2	78	77	0	0	35	16	0	0	19	18	84.09%
MV1	39	35	21	10	30	15	18	15	14	13	72.13%
MV2	52	51	38	19	30	15	26	24	18	18	77.43%
Only Backward Contribution (Average overall accuracy *87.76%*)											
SHV	137	131	0	0	0	0	0	0	0	0	95.62%
MSV	28	28	76	65	0	0	7	3	0	0	86.49%
GV1	30	30	0	0	15	15	40	18	0	0	74.12%
GV2	78	76	0	0	35	35	0	0	19	18	97.73%
MV1	39	36	21	16	30	30	18	7	14	13	83.61%
MV2	52	52	38	35	30	30	26	12	18	17	89.02%
Proposed approach (Average overall accuracy *98.77%*)											
SHV	137	137	0	0	0	0	0	0	0	0	**100.00%**
MSV	28	28	76	75	0	0	7	7	0	0	**99.10%**
GV1	30	30	0	0	15	15	40	40	0	0	**100.00%**
GV2	78	78	0	0	35	35	0	0	19	18	**99.24%**
MV1	39	39	21	19	30	30	18	16	14	14	**96.72%**
MV2	52	52	38	36	30	30	26	24	18	18	**97.56%**

Table 2. Experimental results obtained observing a real test set of 90 minutes video evaluating five different conditions: single handoffs, miss segmentations, group enter, group inside and simoultaneous handoffs

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE