Fusion of Multiple Tracking Algorithms for Robust People Tracking

Nils T Siebel and Steve Maybank

Computational Vision Group
Department of Computer Science
The University of Reading
Reading RG6 6AY, England

{nts,sjm}@isg.cs.reading.ac.uk
http://www.cvg.cs.reading.ac.uk/

Abstract. This paper shows how the output of a number of detection and tracking algorithms can be fused to achieve robust tracking of people in an indoor environment. The new tracking system contains three co-operating parts: *i*) an Active Shape Tracker using a PCA-generated model of pedestrian outline shapes, *ii*) a Region Tracker, featuring region splitting and merging for multiple hypothesis matching, and *iii*) a Head Detector to aid in the initialisation of tracks. Data from the three parts are fused together to select the best tracking hypotheses.
The new method is validated using sequences from surveillance cameras in a underground station. It is demonstrated that robust realtime tracking of people can be achieved with the new tracking system using standard PC hardware.

Keywords. Visual Surveillance, People Tracking, Data Fusion, PCA.

1 Introduction and Related Work

Realtime automated visual surveillance has been a popular area for scientific and industrial research in the past few years. People tracking naturally plays a key role in any visual surveillance system, and a number of tracking algorithms for different applications have been presented [1–9].

While these works present a wide range of different methods, they can be classified into three main categories of increasing complexity:

1. Methods using region- or blob-based tracking, sometimes with additional classification schemes based on colour, texture or other local image properties [2–4, 6, 7].
2. Methods using 2D appearance models of human beings [1, 5].
3. Methods using full 3D modelling of human beings [8, 9].

Fig. 1. View from Surveillance Camera with Tracked Outline Shapes

The more detailed the models for people detection and tracking, the better the system can handle the particular situations for which it is trained. However, systems in category 3 with complex 3D models are usually too slow to use in realtime [9] or they require a special camera/scene setup [8] rarely available in a visual surveillance environment. This is why most of the methods used for visual surveillance fall into categories 1 or 2. The people tracker developed in this work is in category 2.

Our people tracker uses a Region Tracker, and an *Active Shape Model* [5] to model the 2D outline of a person in the image. Figure 1 shows a number of people tracked by the Active Shape Tracker. The People Tracker has been modified over time to increase tracking robustness and adapted for use in an integrated visual surveillance system. We ported the tracker from an `sgi`$^{\mathrm{TM}}$ platform to a PC running `GNU/Linux` to facilitate economical system integration. The People Tracker is now part of the `ADVISOR`[1] integrated system for automated surveillance of people in underground stations. The overall aim of the `ADVISOR` system is to track people in realtime and analyse their behaviour. It has to display video annotations and warnings in realtime to the human operator and archive them together with the digitised video. Figure 2 shows the overall system layout, with individual subsystems for tracking, detection and analysis of events, and storage and human-computer interface subsystems to meet the needs of surveillance system operators. Each of these subsystems is implemented to run in realtime on off-the-shelf PC hardware, with the ability to process input from a number of cameras simultaneously. The connections between the different subsystems are realised by Ethernet.

[1] Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval

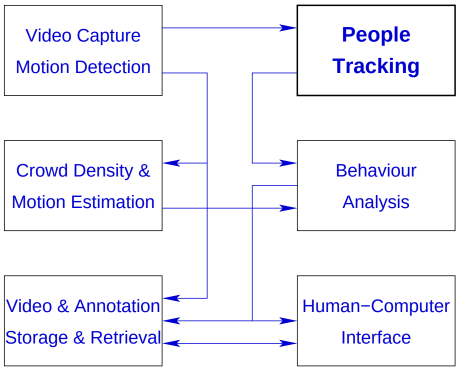


Fig. 2. The `ADVISOR` System with its 6 Subsystems Connected by Ethernet

Our research group is responsible for building the People Tracking subsystem of `ADVISOR`. This article describes the overall structure of the methods employed. Our starting point was the *Leeds People Tracker* by Baumberg and Hogg [5], but it has been modified and new region tracking and head detection modules have been added to achieve robust performance. Section 2 describes the main components of our People Tracker. Section 3 explains how these components interact and how their output is fused together to provide more robust tracking results. Validation results from experiments with the new People Tracker can be found in Section 4.

2 Overview of the People Tracking System

Our People Tracking system, as part of the `ADVISOR` system depicted in Figure 2, is itself comprised of four modules which co-operate to create the overall tracking output, aiding each other to increase tracking robustness and to overcome the limitations of individual modules. The four modules are

1. **Motion Detector** — to detect in each frame those regions containing moving objects,
2. **Region Tracker** — to track these regions over time,
3. **Head Detector** — to detect heads in the tracked regions,
4. **Active Shape Tracker** — to detect and track people over time using a model of their outline shape.

In this section we will briefly describe each of the four modules. Section 3 will describe how the modules inter-operate and how tracking output from all of them is fused together to refine tracks and create the tracking output of the system.

2.1 Module 1: The Motion Detector

In order to detect moving people in the image, the Motion Detector maintains a model of the background, that is, the appearance of the scene without people, and including all static objects. The Motion Detector obtains the background model by median-filtering the video image over time.

The Motion Detector subtracts the background image from the current video image. Thresholding of this difference image yields a binary image containing "foreground" *regions*, which are defined by bounding boxes. Each bounding box typically includes one or more moving *blobs*, i.e. connected sets of pixels where movement was detected. Those regions that match certain criteria for size and shape are classified as potential people, and together with the motion image they represent the output from the Motion Detector.

2.2 Module 2: The Region Tracker

The Region Tracker tracks all moving regions detected by the Motion Detector. It uses a frame-to-frame region matching algorithm together with information about the long-term tracking history of each region. Regions are matched according to their size and position, using a first-order motion model for their movement in the image.

In the following sections we will explain the main two features of the Region Tracker: Region Splitting/Merging and the temporal integration of static regions into the background.

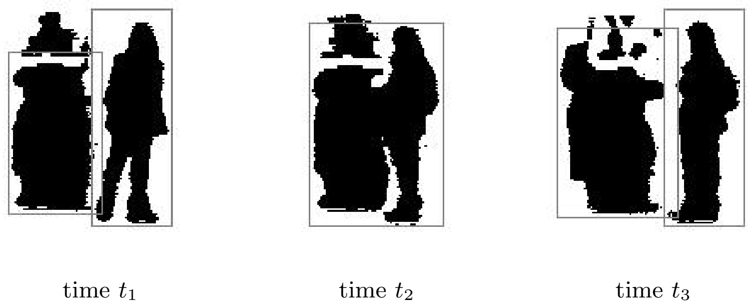


Fig. 3. Problem During Motion Detection: Two people, distinct at time t_1, come close to each other a little later, at time t_2, and their blobs merge. As a consequence, the Motion Detector now extracts the two people as one moving region (the grey box). When at time t_3 the people split again, and the Motion Detector again extracts two moving regions. These tracking and identification problems are overcome by Region Splitting in the Region Tracker.

Region Splitting and Merging

In order to match moving regions, measured in the image by the Motion Detector, to the regions predicted by the Region Tracker, the predictions are combined in a constructive approach to resemble the measurements. If a combination of predicted regions matches the measurements, these predictions are adjusted and their track accepted. This measure becomes particularly important in cases like the one demonstrated in Figure 3 when one measured region, at time t_2, corresponds to two regions predicted from a time t_1, $t_1 < t_2$.

Due to the limited capabilities of our Motion Detector, detected moving blobs are sometimes split in the image even if they are part of the same person. This is usually either due to partial occlusion or to low contrast of moving pixels with the background. Small adjacent regions detected by the Motion Detector are therefore merged by the Region Tracker, before matching them to predicted regions.

Temporal Integration of Static Objects into the Background

One problem when tracking regions in the image is the overlap of two (or more) blobs arising from different people, as in the example in Figure 3. If one of the corresponding objects is still moving and the other one has become static only a few frames previously the background model usually does not include either of the two. As a result, the Motion Detector stills report that both objects are moving. If these objects come close and overlap in the image, they are detected as one moving region which makes it difficult to maintain the correct identification of both objects. Also, the Active Shape Tracker, which uses a pixelwise difference of video image and background image, is not able to correctly detect the outline of both people in local edge search because it has an incorrect notion of the background.

Static objects can be integrated into the background, thereby enabling correct detection and identification of objects moving nearby. The background image is updated periodically using a temporal median filter with the result that all static objects are eventually incorporated into the background. However, if we incorporate detected, static objects too quickly into the background, we might not be able to identify and track them when they start moving again. Moreover, we might detect a "negative" of the object which becomes stationary and then later moves on, because the background model includes the object.

A simple procedure has been devised and implemented to incorporate static objects *temporarily* into the background, thereby

- resulting in much better detection of movement in the vicinity of objects which have become static only a few frames previously, e.g. movement of people past stationary people,
- making it possible to restore the "empty" original background at the exact time a static person starts moving again,
- enabling the identification of the object starting to move again, by keeping a copy of the necessary data (position, size, identity record).

The areas in the image where static objects have been detected and incorporated into the background are specially marked so that the background modelling algorithm does not include these objects into the background image.

2.3 Module 3: Head Detection

Inspired by the W^4 system by Haritaoglu *et al* [1] we have implemented a simple algorithm to detect head positions in moving regions detected by the Motion Detector and Region Tracker. The algorithm creates a vertical pixel histogram of the upper part of these regions, operating exclusively in the black and white motion image. It uses a simple low-pass filter to make up for noise and “holes” in the motion image. It then scans the vertical histogram for peaks, using camera calibration to analyse whether these might be heads of people. The hypotheses of detected head positions in the image are the output from this module. They are stored together with the associated region for later use by the Active Shape Tracker.

The algorithm implemented in the Head Detector is obviously much simpler than normal face detection and tracking methods which use more sophisticated means like skin colour models [10, 11]. The sole purpose of the Head Detector in our system is to aid in the initialisation and validation of tracks by the Active Shape Tracker. Wrongly detected heads (false positives) usually do not influence tracking output because no person can be detected for this head. Furthermore, the implementation of the Head Detector is very fast, adding no noticeable overhead to the overall CPU time used by the People Tracker.

2.4 Module 4: The Active Shape Tracker

Our Active Shape Tracker is based on the *Leeds People Tracker* by Baumberg and Hogg [5]. A space of suitable pedestrian outline shapes is learnt in a training stage using a set of video images containing walking pedestrians. Detected person outline shapes are represented using cubic B-splines. Each outline is specified by a point in a high-dimensional parameter space. Principal Component Analysis (PCA) is applied to the obtained set of points to generate a lower dimensional subspace S which explains the most significant modes of shape variation, and which is a suitable state space for the tracker.

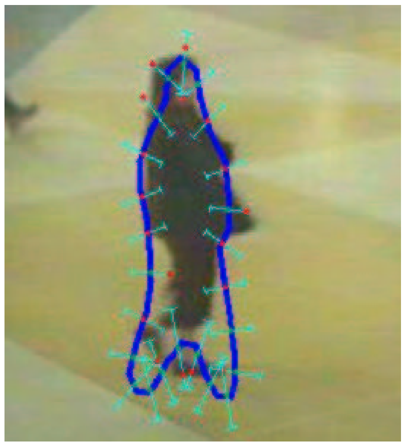


Fig. 4. Edge search for shape fitting

Moving regions detected by the Region Tracker are examined more closely by the Active Shape Tracker. The shape of the given blob is approximated by a cubic B-spline and projected into the

PCA space S of trained pedestrian outlines. The new shape obtained in this initialisation process is then used as a starting point for further shape analysis. Once people are recognised they are tracked using the trained shape model. Tracking is performed using Kalman filtering with second order models for the movement of people in 3D. The state of the tracker includes the current outline shape as a point in S. This point is updated as the observed outline changes during tracking.

In order to adjust the outline shape to each new image of a person we use an iterative optimisation method. The current estimate of the shape is projected onto the image. The shape is then fitted to the image in an optimisation loop by searching for edges of the person's outline in the neighbourhood of each spline control point around the shape. Figure 4 illustrates this process. The pale blue lines show the Mahalanobis optimal search direction [12] used in local edge search.

The Active Shape Tracker has the following deficiencies:

– tracking succeeds only if a sufficient part of the outline is visible. Although the tracker models some occlusion, it loses track in complex situations or when people are too close to each other.
– track initialisation relies on output from the Motion Detector which means that background modelling must be sufficient for the lighting conditions.
– once a track is lost, there is no means of re-establishing the identity of a person when re-gaining track.

3 Data Fusion and Hypothesis Refinement

The main goal of using more than one tracking module is to make up for deficiencies in the individual modules, thus achieving a better overall tracking performance than each single module could provide. When combining the information from different models it is important to be aware of the main sources of error for the different modules. If two modules are subject to the same type of error then there is little benefit in combining the outputs.

3.1 Interaction Between Tracking Modules

After explaining the operation of each of the four tracking modules in Section 2 we will now focus on the interaction of these modules, i.e. how each module uses the output from other modules. Figure 5 shows how the data flow between the four modules and two storage facilities: the Background Model and the Tracking Status and History database. The modules are operated in the order they are presented above:

1. Motion Detector,
2. Region Tracker,
3. Head Detector,
4. Shape Tracker.

Each of the modules has access to the output from the modules run previously, and also to the long-term tracking history which includes past and present tracks, together with the full tracking status (visibility, type of object, whether it is static etc). For each tracked object, all measurements and tracking hypotheses and predictions generated by different modules are stored in one place. As a result, a module can access all tracking data generated by other modules and associated with a given object.

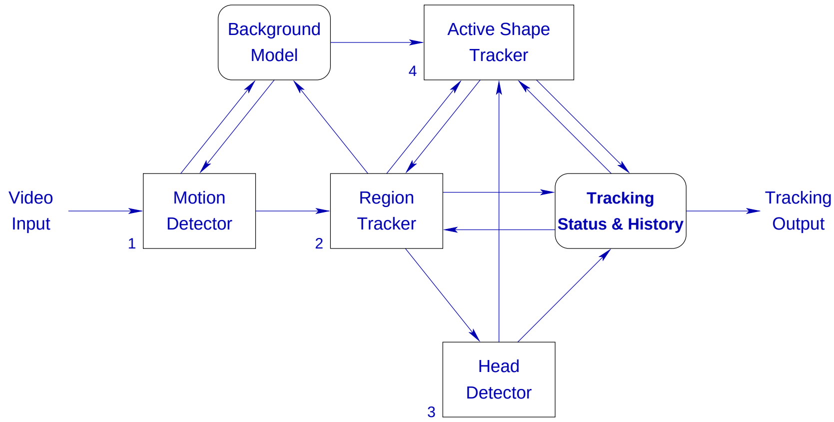


Fig. 5. Interaction Between Modules. The numbers refer to the sequence in which the modules are operated in the tracking stage.

Module 1: The Motion Detector. The Motion Detector uses the background model to detect new objects. With the help of the Region Tracker static objects are quickly incorporated into the background, thereby enabling the correct detection of other objects moving in front of the static objects. The background model is kept up to date by removing objects from it when they start moving again. This, in turn, helps the Region Tracker to track objects more reliably by improving the Motion Detector's measurements.

Module 2: The Region Tracker. The Region Tracker uses the Tracking Status and History database to track regions over time. It can create multiple hypotheses of tracks using Region Splitting and Merging, and keeps track of objects even if they are not visible in particular frames. If the Region Tracker is unsure about a track or it cannot detect an object, it consults the Active Shape Tracker's output for this object. If the associated shape model of a person is successfully tracked, the Region Tracker uses the bounding box of the tracked shape outline as a new hypothesis for the region's position and size.

Another way the Region Tracker uses the output of the Active Shape Tracker is to split up large regions if they contain more than one person. After both the region and the associated person shape(s) have been tracked, the Region Tracker checks whether a significant part of the region was not covered by the shape(s) contained in the region. This situation occurs when two or more people are close together and detected as one region, for example when they enter the scene in a group, one person occluding the other. Once the Region Tracker has established that there is more than one person in the region, the region is divided into two and each subregion tracked separately in order to establish the track of all people within the group. Camera calibration is used in this process to determine whether the remainder of the region, after splitting it, is large enough to contain more people, and how many there might be. In subsequent frames, the Region Tracker uses Region Splitting to correctly split up the output from the Motion Detector into two or more regions in order to track every person correctly. These split-up regions are then processed separately by the Head Detector, and the Active Shape Tracker tries to initialise person outline tracks in each of these.

Module 3: The Head Detector. The Head Detector uses the regions detected and tracked by the Region Tracker. It stores a list of candidate head positions together with the associated regions in the Tracking Status database. These head positions are mainly used by the Active Shape Tracker.

Module 4: The Active Shape Tracker. The Active Shape Tracker has the largest number of inputs from other modules. The most important support the Active Shape Tracker receives from other modules is in the initialisation and identification of tracks. The initialisation refers to the process of estimating the position and size of an outline shape in the image. Once a track is initialised the Active Shape Tracker uses its own predictions and tracking status, stored in the central Tracking Status and History database, to keep track of a person.

The initialisation of a track utilises detection results both from the Region Tracker and the Head Detector. In addition to the initialisation of shapes from tracked regions as described in Section 2.4 above, the heads detected in a region are used to determine possible positions of people. In this process, camera calibration is used to create hypotheses of the most probable size and position of a person in the image. This initialisation by head positions is particularly important when there is more than one person in a given region, e.g. when a group of people is detected as one region.

Additional attempts to initialise tracks are made for regions and already tracked outline shapes if the Active Shape Tracker detects that a shape is too large or too small to be a person. This situation can occur when the upper or lower part of the person is occluded. Using camera calibration, two additional hypotheses are created for the tracked object to cover the cases that either the lower or the upper part of the person's outline is visible. Hypotheses created in this way are added to the tracked object for post-processing and filtering, described in Section 3.2.

During the shape fitting process, the Active Shape Tracker also uses the current background model together with the current video image to facilitate local edge search around the current shape. In this way, the Active Shape Tracker benefits significantly from the temporal integration of static objects into the background by the Region Tracker, resulting in more accurate tracking results.

When a new track is initialised by the Active Shape Tracker, it is assigned the identity of the associated region. This is especially important in cases when a track is lost by the Active Shape Tracker, eg due to occlusion. If the Region Tracker keeps the track then the Active Shape Tracker can re-establish the identity of the tracked person when the track is re-gained at a later time.

3.2 Hypothesis Refinement

After running all four detection and tracking modules, the data and tracking hypotheses generated by them are further analysed and filtered. The trackers usually generate more than one hypothesis for each tracked person, and the information can be of different types (moving region, head position, shape model). In order to reduce the number of hypotheses, they are first pairwise compared to see whether they are multiple observations of the same object. Those which are, e.g. multiple shape models for the same person, are further compared using a track confidence measure generated by the tracker and the positional uncertainty for the predicted position. The better one of the two tracks is then accepted as valid and the other one discarded, whilst making sure that the original identity of the person or object is carried over to the next frame.

More refinement of hypotheses is done by keeping multiple hypotheses of the same track if they are considered as possibly valid. Although only the best tracks appear in the tracking output more tracks are kept in the Tracking Status database, and predictions are made of these by the associated trackers. This way, a hypothesis not matched in one frame, e.g. in the event of a partial occlusion or a missed (dropped) frame in the system, is not lost but may again be matched to measurements and the track re-acquired, at a later time.

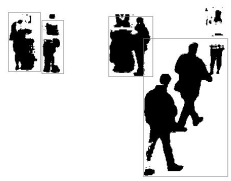
4 Experimental Results

To demonstrate the applicability of our approach, we performed several experiments with video sequences from a surveillance camera in a London Underground station. We have chosen a surveillance camera which has all the potential difficulties for people tracking:

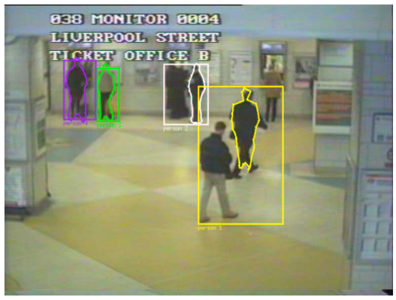
- stationary people waiting in queues at the three counters of the ticket office,
- people who occlude each other as they walk past,
- low contrast of some people against the background,
- people in groups who come close to each other and part again.

The sequence is digitised at full PAL resolution (768×576 pixels) at 5 frames per second (fps). The People Tracker runs in realtime (that is, 5 fps) with the amount of objects shown, on a 1 GHz Pentium$^{\text{TM}}$ III based PC, running under

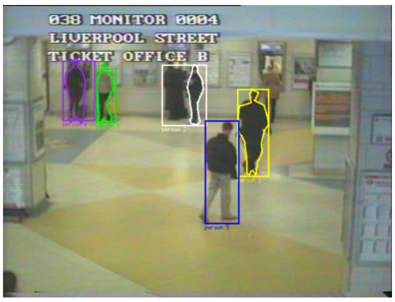
GNU/Linux. The computing time includes software JPEG decompression of the video input and annotation output in XML, but no display (this is done by the Human Computer Interface in the ADVISOR system). Screen output, when enabled during development, approximately doubles the amount of overall CPU time used by the People Tracker.



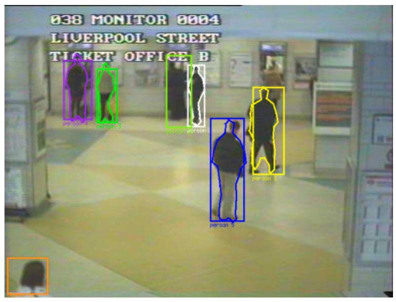
(a) Frame 2, Motion Image



(b) Frame 2, Video Image



(c) Frame 3



(d) Frame 4

Fig. 6. Frame 2 (Motion and Video Images); Frames 3 and 4 (Video Images).

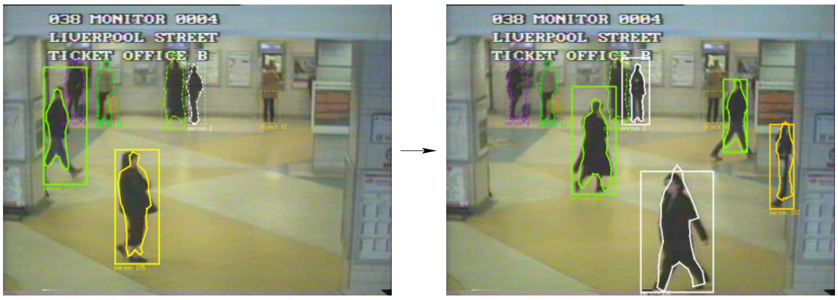
Figures 6 and 7 show the tracking output of the tracker. Regions are drawn into the video images by their defining bounding box, and associated shapes from the regions tracker are drawn in the same colour. In the following we examine how the newly implemented features work in practice.

In the motion image of **Frame 2**, Figure 6(a), we see that initially not every person is detected in a separate region because they are too close together, and no previous tracking data is available to try to correct this. As a consequence, not all shapes could be initialised correctly by the Active Shape Tracker, as can be seen in the video image, Figure 6(b). We notice, however, that by using the

Head Detector, one person's shape was initialised correctly in each of the two large regions (track 1, yellow box, and track 2, white box).

In the Hypothesis Refinement stage, the Region Tracker tries to separate the detected person within the large regions from the remainder of the region. **Frame 3**, Figure 6(c), thereby yields the new track of the person in the middle (track 5, blue box).

In **Frame 4**, Figure 6(d), the track of region and outline shape of person 2 (white) are aligned, enabling the separation of the adjacent person the left. The track of this new person is picked up by the Region Tracker (track 10, light green box) and tracked using Region Splitting. When the Active Shape Tracker also starts tracking the person a little later, their identity and track history can already be established.



(a) Frame 39

(b) Frame 55

Fig. 7. Frames 39 and 55 (Video Images).

A few frames later in the sequence, in **Frame 39**, Figure 7(a), shows that the people at the counter have been detected as static and temporarily incorporated into the background. This is indicated by a dashed outline around static objects. The person at the right counter cannot be tracked by the Active Shape Tracker because their jacket does not stand out against the background. However, the Region Tracker keeps track of their dark trousers until the full outline can be tracked by the Active Shape Tracker.

In **Frame 55**, Figure 7(b), Person 2 (white) has started to move again. Their image is removed from the background and both Region Tracker and Active Shape Tracker continue tracking with the correct id. Meanwhile, a person has walked past all five static people without their track being lost.

In the following we will focus on the three shortcomings of the Active Shape Tracker as discussed in Section 2.4 and examine in sections 4.1 to 4.3 how the new tracking system deals with these problems.

4.1 Dealing with Occlusion

The new tracking system has two advantages over the original Active Shape Tracker in handling occlusions. These are due to the combination of the Active Shape Tracker with our feature-rich Region Tracker:

- static objects are incorporated into the background. This feature of the Region Tracker, which was discussed in detail in Section 2.2, helps to avoid many types of occlusions before they happen.
- occlusion strongly affects the Active Shape Tracker when a large part of a person's outline is not detectable. To the Region Tracker, due to its nature, not observing the outline does not pose such a difficult problem. Region Splitting and Merging helps the Region Tracker to avoid problems of occlusion. By linking the Active Shape Tracker with the Region Tracker and tracking the person shapes and associated regions independently, the overall tracker does not lose track of a person so easily.

4.2 Initialisation of Tracks

(a) Initialisation of New Tracks in the Active Shape Tracker

(b) Final Tracking Output after Hypothesis Refinement

Fig. 8. Initialisation and Refined Tracks, Frame 6

Figure 8(a) shows the locations in the image where the Active Shape Tracker looks for people, in order to initialise new tracks. Initial shape estimates are projected onto the image before the shape fitting process is started. Hypotheses of already tracked people are not displayed here. Some of the shapes examined in the initialisation process are very close, and in the shape fitting process they converge to the same person. During the Hypothesis Refinement stage, only the strongest hypotheses are kept and the others are abandoned.

Potential head positions as detected by the Head Detector are marked in the image in red, showing search regions and estimated head positions. It can be seen that most of the heads are detected, and an active shape model is initialised in the vicinity of every person in the image.

The bottom left corner of Figure 8(a) shows a detected region from the Region Tracker, in this case the head of a person. The Active Shape Tracker uses camera calibration to establish that this region is too small to be a person and examines two hypotheses: either the region constitutes the head of a person or their feet.

Figure 8(b) shows the final tracking output for the same frame shown in Figure 8(a). All people but the man at the rightmost counter have been correctly detected and are tracked. The man's beige jacket has a particularly low contrast to the background and additionally, most of his body outline is occluded by the person in front of him. This is why the Active Shape Tracker does not find enough edges during the shape fitting, and consequently does not detect the person.

4.3 Establishing the identity of lost tracks

In the sequence shown, there were no instances of lost tracks. This is because the Region Tracker aids the Active Shape Tracker so the overall tracker is less likely to lose track of a person. In our system, even if a track is not detected in one frame, or the video image is unavailable for a frame or two, their data will not be removed from the Tracking Status and History database. Each tracker still makes predictions for the object in question, and if its track can be re-gained at a later stage, their original identity and tracking history can be re-established.

5 Summary and Discussion

Our People Tracking System is part of the integrated visual surveillance system `ADVISOR` developed for operation in an underground station. The People Tracker is based around an Active Shape Tracker which tracks the 2D outlines of people in video images. In order to achieve robust tracking of people, a Region Tracker and a Head Detector were added to the original tracker design. A new software design has been employed which allows for interaction between these modules.

By combining multiple modules to form one People Tracker, and fusing their output to generate the tracking output, we achieve a higher tracking reliability than any of the individual trackers can achieve on its own. The Region Tracker with its Region Splitting and Merging, as well as its Temporal Background Integration features plays an important role, helping when occlusion creates difficulties for the Active Shape Tracker, and in the identification of tracks for correct long-term tracking. Output from the Head Detector helps in the initialisation of tracks which is part of the Active Shape Tracker.

The new software structure also enables us to run the People Tracker on input from multiple cameras. This is part of ongoing work within the research and development in the `ADVISOR` project.

Acknowledgements

This research was supported by the European Union, grant `ADVISOR` (IST-1999-11287).[2] The authors wish to thank London Underground and Sergio Velastin (Kingston University) for providing video sequences, and our visitors and students for their valuable coding work.

References

1. I. Haritaoglu, D. Harwood, and L. S. Davis, "W^4: Real-time surveillance of people and their actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809–830, August 2000.
2. Q. Cai, A. Mitiche, and J. K. Aggarwal, "Tracking human motion in an indoor environment," in *Proceedings of the 2nd International Conference on Image Processing (ICIP'95)*, pp. 215–218, 1995.
3. C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," Tech. Rep. 353, MIT Media Laboratory Perceptual Computing Section, 1995.
4. F. Brémond and M. Thonnat, "Tracking multiple non-rigid objects in a cluttered scene," in *Proceedings of the 10th Scandinavian Conference on Image Analysis (SCIA '97), Lappeenranta, Finland, June 9–11, 1997*, vol. 2, pp. 643–650, 1997.
5. A. M. Baumberg, *Learning Deformable Models for Tracking Human Motion*. PhD thesis, School of Computer Studies, University of Leeds, October 1995.
6. A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," in *Proceedings of the DARPA Image Understanding Workshop (IUW'98), Monterey, CA, November 1998*, pp. 129–136, 1998.
7. S. Khan, O. Javed, Z. Rasheed, and M. Shah, "Human tracking in multiple cameras," in *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001), Vancouver, Canada. July 9–12, 2001*, pp. 331–336, July 2001.
8. D. M. Gavrila and L. S. Davis, "Tracking of humans in action: A 3-D model-based approach," in *ARPA Image Understanding Workshop*, (Palm Springs), pp. 737–746, February 1996.
9. H. Sidenbladh, M. J. Black, and D. J. Fleet, "Stochastic tracking of 3D human figures using 2D image motion," in *ECCV 2000, 6th European Conference on Computer Vision* (D. Vernon, ed.), pp. 702–718, Springer Verlag, 2000.
10. J. L. Crowley and F. Bérard, "Multi-modal tracking of faces for video communication," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 640–645, 1997.
11. N. Oliver, A. P. Pentland, and F. Bérard, "Lafter: Lips and face real time tracker," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 123–129, 1997.
12. A. Baumberg, "Hierarchical shape fitting using an iterated linear filter," in *Proceedings of the Seventh British Machine Vision Conference (BMVC96)*, pp. 313–322, BMVA Press, 1996.

[2] However, this paper does not necessarily represent the opinion of the European Community, and the European Community is not responsible for any use which may be made of its contents.