Chapter 13

Conclusions and Future Work

Omnidirectional camera systems are characterised by their ability to observe large fields-of-view. This offers several advantages to computer vision applications, in particular to surveillance and tracking applications. These benefit from having unobstructed views of their surroundings and being able to detect and track objects simultaneously and in different parts of their field-of-view without requiring any camera motion. The OmniTracking application described in this thesis makes use of such an omnidirectional camera system for the purpose of detecting and tracking objects.

13.1 Conclusions

In §1.1, we have mentioned the three main issues that are addressed by this thesis. For the first issue (robust object tracking), we have examined the background subtraction algorithm (§8) and different types of background models, and decided to use the normal distribution model. Experiments were also performed using a Gaussian mixture model but this was deemed to be computationally too costly. We also made shadow detection an integral part of the motion detection process to achieve robustness against shadow-induced problems, and used hysteresis thresholding for noise elimination.

For the tracking part, we investigated two different approaches, a blob-tracking based method (§10) and a colour-based statistical tracking method (§11). For the first method, we used a combination of match scoring using similarity measures and temporal constraints to match objects from one frame to the next. The object fragmentation issue was solved for this method by means of blob clustering.

The colour-based method (§11) uses a Gaussian mixture model to statistically represent multi-coloured objects, with object tracking consisting of the process of finding the highest probability region in an image and then using this information to define a search window for the next image frame. We used a combination of temporal constraints and the motion detection result to limit the search window area. The initial colour model is learnt automatically using the Expectation-Maximisation (EM) algorithm and we chose a variant of the EM, the Incremental EM, for updating the model. To ensure that a good representation for the object's colours is generated by the EM algorithm, we developed a technique for initialising the components of the mixture model with distinct values selected from the object's colours.

The application was tested in two different environments, using the PETS datasets mentioned in §4. We determined that the motion detection method used gives a high detection rate, locates nearly all of the moving objects, suppresses shadow correctly and eliminates small-scale noise. We also investigated the detection problems that result from background subtraction and found that the major ones were object fragmentation (caused by an object with the same colour as the background) and background 'furniture' that are moved by the subjects.

In the case of object tracking, we evaluated the results of the blob tracking method and found that this method suffers from the problems of occlusion and object merging, where the tracker is not able to differentiate between objects while they are merged together. We found out that although the tracker is able to recover the identity of most objects when they are no longer occluded or in a merged state, in some cases this leads to the tracker losing the objects completely. Therefore, we concluded that the blob tracking method is not sufficiently robust.

We also evaluated the colour-based tracking method and found out that this method is able to track objects through partial occlusion, and even through full occlusion events. This method was also found to solve the object fragmentation problem of motion detection in the majority of cases. We investigated the detection problems of this method and these fall into two categories: problems that are caused by a poor colour model as fitted by the EM algorithm and false positives due to moved background elements. The first type arises when the size of the object is very small and so there is insufficient colour information for the EM algorithm to learn a good colour model.

Concluding on the first issue, the motion detection algorithm combined with colour-based tracking achieve very good tracking results, with the majority of objects in the PETS datasets being tracked successfully from the time they appear in the scene until the moment they leave, and their identities are maintained throughout occlusion events. When compared to cameras with limited fields-of-view, including the non-omnidirectional cameras used for acquiring some of the video streams in the PETS datasets (see §4.3), we have shown the advantage that omnidirectional tracking systems offer – allowing objects to be tracked over longer periods of time which should ultimately lead to better awareness of their behaviour.

For the second issue given in §1.1 (generic tracking algorithms), we have mentioned in §4.1 the differences that arise from working in two different environments, for example, object fragmentation tends to be more prominent in indoor scenes, and we have implemented the algorithms to work in both environments without making use of any domain-specific knowledge.

For the final issue mentioned in §1.1 (non-linear image geometry), we have implemented algorithms for generating perspective and panoramic views from omnidirectional images to present the output of the omnidirectional camera and the tracking results in a human-viewable form. One of the advantages of omnidirectional cameras is their ability to see all of their environments at any one time. We have combined this with the single viewpoint constraint to implement algorithms that allow the user to create a number of simultaneous virtual cameras, centred at the camera's single viewpoint, and to control these cameras in real-time. We have used an extended pan-tilt-zoom model for controlling the virtual cameras.

We have also combined the results generated by the tracking algorithms with those of the perspective view-generation methods to allow the application to create virtual cameras that automatically track objects as they move across the field-of-view. A combination of fully automated and manually-triggered automatic tracking caters for potentially different application usage.

13.2 Future Work

We have identified three main areas that can help improve or extend the work done for this thesis. In addition to these, there are several minor enhancements that can be applied and many of these are mentioned at the end of the preceding chapters.

1. In the case of colour-based tracking, the accuracy of the results depends on how up-to-date the object's colour model is. Currently the model is regularly updated using the Incremental EM algorithm (IEM), and a full EM update is performed occasionally to compensate for any errors that might accumulate over time. These updates happen regardless of the state of the object's model.

   This can be improved during the tracking phase, by checking how much the model differs from the actual object's colour information and performing the update only when necessary. The difference can be calculated using the log-likelihood as a measure of difference. The main advantage that will be provided is that the updates will be required less often, since the current rate of update was chosen to cater for the worst case scenario where objects undergo significant changes in their appearances.

2. One of the limitations of using mixture models for multi-coloured objects is that the model does not provide any information about the spatial arrangements of the colours within the object. Recent approaches to object tracking [ELGA02], are increasingly exploiting statistical methods to model the spatial arrangements of object colours.

   Spatial information can be incorporated into the object model developed for this thesis to provide greater robustness to object tracking and this can help to alleviate the chance of mismatch where objects have regions with similar colours (like the cases identified in §11.7).

3. The third area deals with the use of virtual cameras to automatically track objects and display them on screen. Currently, the camera controller algorithm selects which objects to track on a first-come first-served basis.

   This could be enhanced by selecting objects based on some significance measure, for example, the last active person in the case of indoor scenes.