Multi-Sensor Tracking using a Scalable Condensation Filter

Simon Denman, Todd Lamb, Clinton Fookes, Sridha Sridharan, Vinod Chandran
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane 4001, Australia
{s.denman, c.fookes, s.sridharan, v.chandran}@qut.edu.au, tk.lamb@student.qut.edu.au

Abstract

Surveillance and tracking systems typically use a single colour modality for their input. These systems work well in controlled conditions but often fail with low lighting, shadowing, smoke, dust, unstable backgrounds or when the foreground object is of similar colouring to the background. With advances in technology and manufacturing techniques, sensors that allow us to see into the thermal infrared spectrum are becoming more affordable. By using modalities from both the visible and thermal infrared spectra, we are able to obtain more information from a scene and overcome the problems associated with using visible light only for surveillance and tracking. Thermal images are not affected by lighting or shadowing and are not overtly affected by smoke, dust or unstable backgrounds. We propose and evaluate three approaches for fusing visual and thermal images for person tracking. We also propose a modified condensation filter to track and aid in the fusion of the modalities. We compare the proposed fusion schemes with using the visual and thermal domains on their own, and demonstrate that significant improvements can be achieved by using multiple modalities.

1 Introduction

Surveillance and tracking systems typically use a single colour modality for their input. These systems work well in controlled conditions but often fail with low lighting, shadowing, smoke, dust, unstable backgrounds or when the foreground object is of similar colouring to the background. These conditions result in poor motion detection as well as poor tracking of an object. With advances in technology and manufacturing techniques, the cost of sensors that allow us to see into the thermal infrared spectrum has become much more affordable. Using modalities from both the visible and thermal infrared spectra, allows us to obtain more information from a scene and overcome the problems associated with using visible light only for surveillance and tracking. Thermal images are not affected by lighting or shadowing and are not overtly affected by smoke, dust or unstable backgrounds. Also, an object of interest is unlikely to be the same colour and temperature as the background. Thermal sensors on their own however are more sensitive to noise than colour sensors, and do not allow the same level of discrimination between different tracked objects (i.e. in a thermal image, two people and their clothing appear very similar due to them being close to the same temperature). Fusing the colour and thermal modalities can prove very useful in a number of applications and situations where a robust surveillance and tracking system is needed.

Previous studies have experimented with multispectral fusion for surveillance and tracking. Conaire et al [11, 3, 2] have experimented with fusion for object segmentation, background modeling and tracking using colour and thermal infrared images. Fusion for tracking is done in the appearance model by using a multi-dimensional Gaussian to represent each pixel. The scores from the visible and thermal spectra in the appearance model are fused in different ways to match the model to the incoming image. The ways of combining scores methods are compared to ascertain the best method for this form of fusion. Some of these methods for fusion in the appearance model have been implemented in this system. Blum and Liu [1] discusses different methods of early image fusion using the wavelet transform and the pyramid transform. These early fusion methods can be used to fuse the images before they are fed into a tracking system. Han and Bhanu [8] discuss techniques for the use of colour images and infrared images for use in moving human silhouette extraction as well using these silhouettes for automatic image registration between the infrared and colour images.

In this paper we propose and evaluate three approaches for fusing visual and thermal images for person tracking. We also propose a modified condensation filter to track and aid in the fusion of the modalities. We compare the proposed fusion schemes with using the visual and thermal domains on their own, and demonstrate that improvements can be achieved by using multiple modalities. Section 2 will discuss the tracking system and condensation filter used; section 3 will present the proposed fusion schemes; results are presented in section 4 and conclusions in section 5.

2 Object Tracking System

We modify the tracking system proposed in [6] to work in a multi-modal environment. The object tracking system uses a hybrid motion detector-optical flow technique[5] as a basis, and scans for appropriate regions of motion to detect people (see figure 1). A modified condensation filter (see section 2.1) is used to track the people.

The condensation filter is uses the colour image and the results of the motion detection to determine the most likely positions for any known tracked objects in the current frame. This information is used to guide the person detection routines to determine their actual locations in the image. Motion associated with these detected people is and removed from the motion image as it is now accounted for. The remaining motion must belong to new people, and so person detection is carried out on remaining areas to locate people who have morecently entered the scene.

Person detection is performed by splitting the image into sub-regions which contain concentrated areas of motion, and then locating heads and fitting ellipses within each region[9, 16]. Working within subregions overcomes problems caused by people occupying a common column of the image causing inaccurate vertical projections. Heads are detected by combining the vertical projection and pixel height of the top contour (to aid in overcoming problems caused by holes in the motion image), and finding local maxima; which are then filtered by analysing the surrounding region. Ellipses are fitted to the valid heads at an aspect dependent on the candidate head, and if there is a suitable occupancy (motion within the bounds of the ellipse) the candidate is accepted. This process is used for the detection of new tracks, and to support the condensation filter tracking. The optical flow results are used to aid both the motion based detection routines and the condensation filter.

2.1 Scalable Condensation Filter

We use a condensation filter[10] to track objects in the system. We propose a Scalable Condensation Filter (SCF), which is an extension for the Mixture Particle Filter (MPF)[15] and Boosted Particle Filter (BPF)[12]. A single filter is used for the entire system, and the particle count is scaled according to the number of objects being tracked. In addition, we allow the number of the particles for each track to vary according to the complexity of the surrounding area (see section 2.1.1). Each tracked objects particles are resampled separately to ensure that particles for a track (and thus the track itself) are not lost due to resampling. This overcomes problems associated with previous multi-target trackers where the samples for a given target could become deleted and the target lost.

Particles are four dimensional, and describe a bounding box (a centre position (x and y pixel coordinates) and the height and width, $\{x, y, h, w\}$). Each variable is free to move within the dimension limits, $\{d_{min}, d_{max}\}$, which are defined by the system (i.e. the limits of x and y are governed by the image size). The distribution of each dimension is Gaussian, with the mean at the the last observed position, and the variance equal to the maximum expected movement of a dimension from one frame to the next, e_{max}.

We use a Sequential Importance Resampling (SIR)[7, 13] procedure to update the sample set. Each new particle is adjusted according to a motion model associated with the tracked object responsible for the particle. The expected movement according to this motion model (based on a window of Q previous observations) is added to the particle as well as a noise vector.

$$S_{(i,n,t+1)} = S_{(i,n,t)} + M_i + R \qquad (1)$$

where $S_{(i,n,t+1)}$ is the nth sample for track i at the next time step; $S_{(i,n,t)}$ is the nth sample for track i at the current time step; R is the random sample, which is within the range of $-e_{max}$ to $+e_{max}$, and M_i is the expected movement for the track, i. As part of all particle updating and creation, we apply a set of limits to each particle, to ensure that it is describes a valid object (if a dimension exceeds a limit, it is set to the limit). Whilst SIR would ensure that any particles that describe invalid objects are not propagated (they would have 0 probability), performing this test on the particles at this point avoid the need to check for valid image coordinates when matching features, which allows the system to be more efficient.

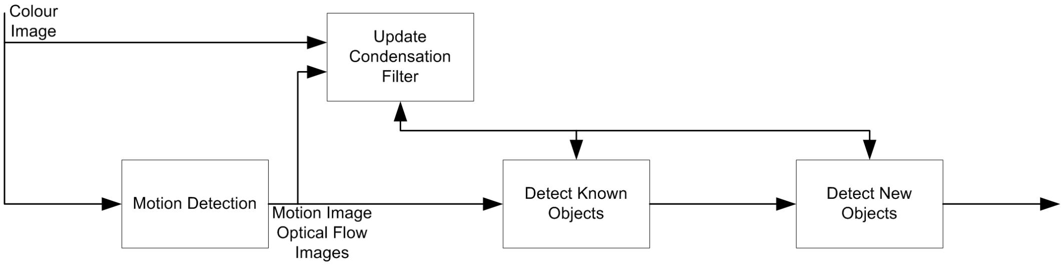


Figure 1. Tracking System Flowchart

2.1.1 Dynamic Sizing

Rather than have a fixed number of samples for the filter, we propose dynamically altering the sample count as objects enter and leave the scene, and as people move about and occlude one another. For each track, we have an arbitrary number of samples, n, that are created about the objects initial position and associated with that object.

$$s_{new} = o_{new} + 2 \times r \tag{2}$$

where s_{new} is the new sample, o_{new} is the new objects state, and r is a random value, in the range $-e_{max}$ to $+e_{max}$.

The particles initially associated with the given track remain associated with that track for duration of that tracks life, and the particle count for any individual track cannot be diminished unless it is specifically desired. This initialisation gives each tracked object a set of samples to model it immediately, rather than needing to allow a period of frames for the system to adapt to its presence. When an object leaves, n samples are removed from the system.

When tracked objects are close together, additional particles can be added and more advanced features can be used to aid in the tracking. Three levels of occlusion are defined for each track:

1. Level 0 (No Occlusion) - The tracked object is isolated within the scene, there are other objects nearby

2. Level 1 (Object Nearby) - Another tracked objects bounding box is within a distance r

3. Level 2 (Overlap) - Another tracked objects bounding box is overlapping

When a track is first created, and added to the SCF, it is at occlusion level 0 and is created with the standard number of particles. For each occlusion level increase, an additional m particles are added to the SCF for that track; and m samples are removed for each occlusion level decrease. Particle counts for tracked objects are altered during the resampling procedure by either undersampling or oversampling.

Resizing the system in this manner ensures that no unnecessary updates are done, and improves CPU utilisation.

2.1.2 Object Features

We allow each track to use multiple features. Using inheritance and polymorphism, the types of features used by tracks can be allowed vary depending on circumstances and the class of object being tracked, without any change required in the condensation filter itself. This approach allows different types of objects to use feature more suited to their individual properties.

We use two types of features, each of which has various sub-classes.

1. Histograms

2. Appearance Models

Each of these features can optionally use motion detection and optical flow as additional aids (i.e. a pixel must be in motion and must be moving in the same direction as the object being tracked), and this can be change dynamically depending on the systems status (i.e. if motion detection is unreliable for a period of time due to environmental effects, this can be omitted when matching features).

Histograms simply model colour distributions, and so while being quicker to compute, do not take geographical information into consideration (i.e. a person wearing blue pants and a red shirt will have a very similar histogram to a person wearing red pants and a blue shirt, despite having a distinct appearance). Appearance models encode position information as well as colour information, and so are more discriminative. They are however more processor intensive.

We propose varying the features used as the system complexity changes. A histogram feature is used by default, and when a tracks occlusion level increases above 0 (see section 2.1.1) an appearance model feature is used as well. When multiple features are used, the probability for the particle is the product of the probabilities for each feature.

As each tracked object has its probabilities normalised, and particles resampled separately, there is no danger of the additional matching constraints reducing a tracks probabilities to the extent that the tracks particles are removed from the system by the resampling procedure. It is feasible that multiple or different appearance models and histograms could be used for each track under appropriate circumstances.

3 Fusion Schemes

To combine the thermal infrared and visible light images, three different fusion approaches are proposed (see figure 2):

1. Fusing images during the motion detection by interlacing the images
2. Fusing the motion detection results of each image
3. Fusing when updating the tracked objects using detected object lists from each modality

The first technique can be implemented with almost no modifications to the tracking system, while the second two require more significant changes to the tracking algorithm.

3.1 Fusion in the Motion Detector

The first proposed method involves fusing the images prior to the motion detection by interlacing the luminance channel of the visible light image with the grayscale thermal infrared image. This approach is facilitated by using a motion detector which requires YCbCr 4:2:2 input [5]. The motion detector analyses images in 2 pixel (four value, two luminance, one blue chrominance and one red chrominance) blocks from which clusters containing two centroids (a luminance and chrominance cluster, $\{Y_1, Y_2; Cb, Cr\}$) are formed. The centroids of the clusters in the background model are compared to those in the incoming image to determine foreground/background.

Rather than convert the colour image to YCbCr 4:2:2 format as would be done in normal circumstances, it is converted to YCbCr 4:4:4. The thermal information is then interlaced with the colour information. By treating the thermal information as additional luminance data and doubling the luminance information, we effectively create a YCbCr 4:2:2 image (see figure 3) that can be fed directly into the tracking system without any further modifications. This results in the motion detector clusters becoming $\{Y, T; Cb, Cr\}$. This method of fusion has the advantage of consuming little processing resources on top of our existing system, is also very simple to implement. It does however require that the colour and thermal images be correctly registered, which may require additional processing, or in some situations, not be possible.

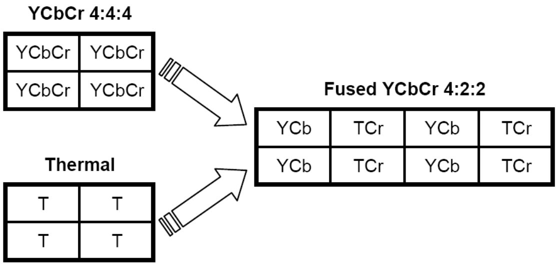


Figure 3. Fusing Visual and Thermal Information - The YCbCr 4:4:4 representation of the input images is combined with the thermal image to produce a YCbCr 4:2:2 image where every second luminance value is actually the corresponding thermal value.

3.2 Fusion After Motion Detection

The use of middle or late fusion allows for greater control over the information contained in the images that can be used by the tracking process. This information can be used to greatly improve the accuracy and robustness of the detection and tracking system. In each of these proposed systems we compute motion detection for both images. If either image shows an abnormal increase in motion, it is disregarded. In the unlikely event that both show such an abnormality, the more consistent of the two is chosen. The abnormality of the images is assessed by looking at the percentage increase of the in motion pixel count.

$$\frac{M_t}{M_{t-1}} > T \tag{3}$$

where M is the amount of motion in the image, t is the time step and T is the threshold for determining invalid motion detection results. This same check for valid motion detection results is used in out third proposed system (see section 3.3).

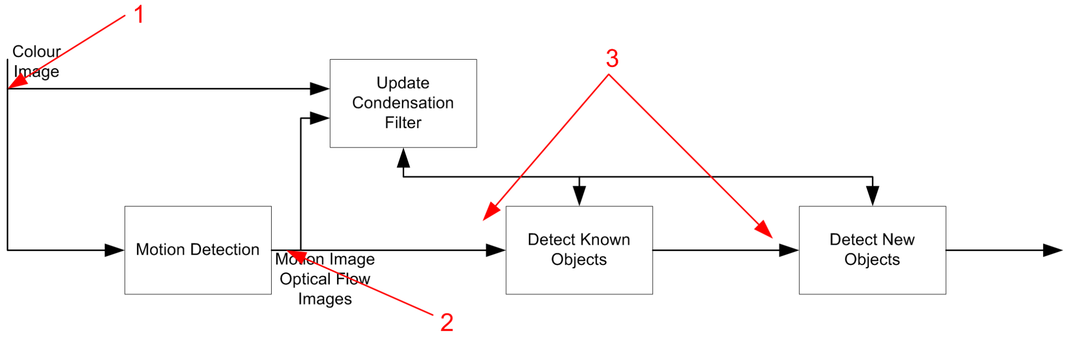


Figure 2. The points for fusion in the system

Our second proposed fusion scheme involves fusing directly after the motion detection. Once the motion detection masks are obtained for each the visible light and the thermal infrared modalities, they are combined to obtain a single mask for the scene. Rather than simply apply a logical "and" or "or" operation, we propose fusing the images the follow equations.

$$(M_{IR}(x,y,t) > T_1)\&(M_{Vis}(x,y,t) > T_1) \tag{4}$$

$$M_{IR}(x,y,t) > T_2 \tag{5}$$

$$M_{Vis}(x,y,t) > T_2 \tag{6}$$

where M_{IR} is the thermal motion image, M_{Vis} is the visual motion image, and T_1 and T_2 are thresholds to control the fusion ($T_2 > T_1$). If any other these equations are satisfied, the fused motion mask at (x,y,t) is set to indicate motion. The resultant mask is used in the remainder of the system described in section 2.

3.3 Fusion After Object Detection

A second mid-fusion scheme is proposed whereby motion detection and object detection is carried out on both modalities, and the two object lists are used to update the central list of tracked objects. The modified condensation filter (see section 2.1), an appearance model and histogram are used to aid in this process.

An efficient colour histogram [14] is used to provide a simple and quick to compute feature to evaluate object matches. The histogram is constructed such that the thermal component is separate from the visual component, making individual comparison of either domain possible. In our experiments we used a four dimensional histogram with the dimensions being: red, green, blue and thermal infrared (with the thermal channel being independent of the colour channels).

The appearance model used in this system is a Gaussian appearance model based on the work done by Conaire et al [11], which models the object of interest as a grid of pixels with each pixel being represented by a single multi-dimensional Gaussian distribution. Each pixel is also assigned a weight depending on the likelihood of it belonging to the foreground. The dimensions in the multi-dimensional Gaussian relate to features of the thermal infrared and visible light images. One of the features relates to the intensity of the input channels of the thermal or visible light images. The other feature used relates to the gradient of the input channels of the thermal infrared or visible light image. The channels used in the appearance model are the same as used in the colour histogram. The mean values of these distributions are initialised as the value of the intensity or gradient of the corresponding pixel in the initialising image, and the standard deviation of the distributions are initialised to a value k (for our experiments we used a value of $k = 30$). The means and standard deviations are updated accordingly with subsequent matched images to the appearance model, allowing changes in appearance to be learned. Fusion is performed by combining the scores obtained by matching the appearance model to the incoming image. This can be done in a number of different ways which are also outlined in [11]. For our, model we chose to use the similarity score product method to combine the scores form our appearance model. This method was used as previous studies have shown it to be the most effective method for fusion in this particular appearance model.

The two models (histogram and appearance) are both used by the condensation filter. However the appearance model is only used when there is the chance of, or an occlusion occurring. The results of the condensation filter process are used to match the candidates detected in both the visual and thermal processes to the known tracked objects. We allow objects that have been previously detected to be updated by a detection in either domain. For a new object to be added, it must be detected in both (objects in each image, at the same point, have a high overlap) or have an acceptable amount of motion in the corresponding portion of the other image.

4 Results

The data used in our experiments for development of the fusion system was obtained from an existing database of images containing both visible and infrared information of different outdoor scenes. As these images were already aligned, they were able to be used develop and test the fusion detection and tracking algorithms with out any further alignment. This data was obtained from the OTCBVS Benchmark Dataset Collection[4], which is a publicly available dataset that contains both thermal infrared and colour image sequences of two different outdoor scenes containing pedestrians. The sequences include a variety situations of interest with multiple pedestrians to test the system. We test the performance of the proposed fusion system as well as tracking with both modalities individually. We compare the performance of the system by visually comparing the tracker output, as there is no ground truth data available for the database.

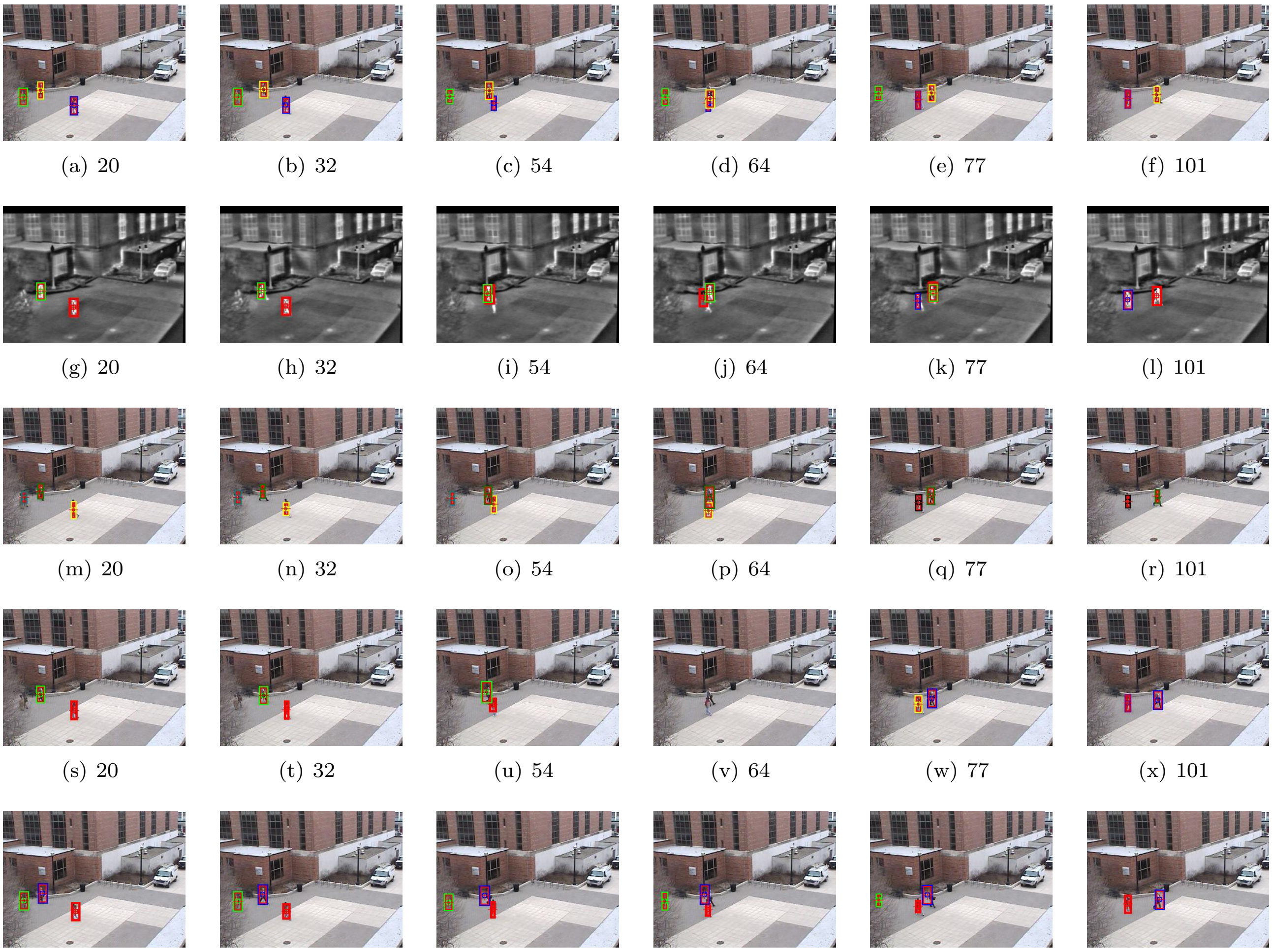
Scenario 1 (see figure 4) involves a well lit outdoor situation where two people walk past each other causing an occlusion. As this is a well lit outdoor situation, most of the tests performed well. The sequence using only infrared images performed the worst which was to be expected as it does not have any colour information available to aid with the occlusion. The sequence using only colour images handled the occlusion well in this case due to the well lit conditions, although one of the people was lost for a brief period during the occlusion. The fusion schemes also performed well in this situation due to the favorable lighting conditions and reliable colour information. Fusion after object detection performed the best in this situation, being the only method that was able to keep track of all of the objects in the scenario throughout the entire sequence.

Scenario 2 (see figure 5) is also a well lit situation but in this case the people in the scene walk toward each other and then stop for a period of time. This is also a it is a well lit situation the once again the thermal only sequence performed the worst. The other tracking sequences all worked well. In this case the fusion after object detection performed the best as it kept track of both people throughout the entire sequence, while other system lost the tracked objects at times.

Scenario 3 (see figure 6) involves heavy cloud cover causing extensive moving shadows across the scene depicting a poorly lit unstable background scenario. The colour sequence performed the worst in this scenario relying solely on colour information for detection and tracking of people, it fails as people move into and out of shadow. The high level of shadowing also results in the motion detection becoming unstable, impacting upon the detection routines. Thermal only and fusion after object detection performed the best in this situation. Fusion after object detection out performs thermal only in that it is able to detect and track some of the objects that are lost or remained undetected in the thermal only sequence. The other fusion approaches experience problems as they are still susceptible to the motion detection errors that occur in the colour image. While the effect is not as severe as it is when using colour alone, it is still sufficient to cause additional errors in the tracking. Fusion after object detection is able to ignore detection results from the colour images in favor of those from the thermal.

Scenario 4 (see figure 7) show a well lit situation with multiple people passing each other at a distance causing occlusions. As would be expected the colour only method performed reasonably well in this well lit situation. The colour only method did lose track of some of the people when they passed each other. The thermal only method performed reasonably well but failed to detect some of the people in the top left hand side of the scene. Thermal only handled the occlusion well, not losing track of the tracks people as they passed each other. The fusion in motion detection performed worst in this scenario, losing track of people multiple times throughout the sequence. Fusion after motion detection and after object detection both performed well with fusion after object detection performing the best as it maintains the tracks on the people throughout the occlusion.

Under appropriate conditions, all fusion schemes can offer some level of improvement over using either modality alone. Overall however, our third proposed fusion scheme (fusion after object detection) performs the best, out performing each camera on its own and the other fusion schemes. Fusions schemes one and two are directly reliant on the quality of the motion detection from the colour and thermal images. If either image contains excessive noise (sensor noise, or environmental effects such as shadowing) the whole system suffers as the fusion has been performed before any object detection processes, and so the object detection for the whole system is degraded. Depending on the conditions of the scene, this may still allow some improvement over either modality individually, however at other times it can result in poorer performance. This can possibly be overcome by modifying the early fusion schemes to determine fusion parameters dynamically. Fusion after the object detection



(a) 20 (b) 32 (c) 54 (d) 64 (e) 77 (f) 101

(g) 20 (h) 32 (i) 54 (j) 64 (k) 77 (l) 101

(m) 20 (n) 32 (o) 54 (p) 64 (q) 77 (r) 101

(s) 20 (t) 32 (u) 54 (v) 64 (w) 77 (x) 101

(y) 20 (z) 32 () 54 () 64 () 77 () 101

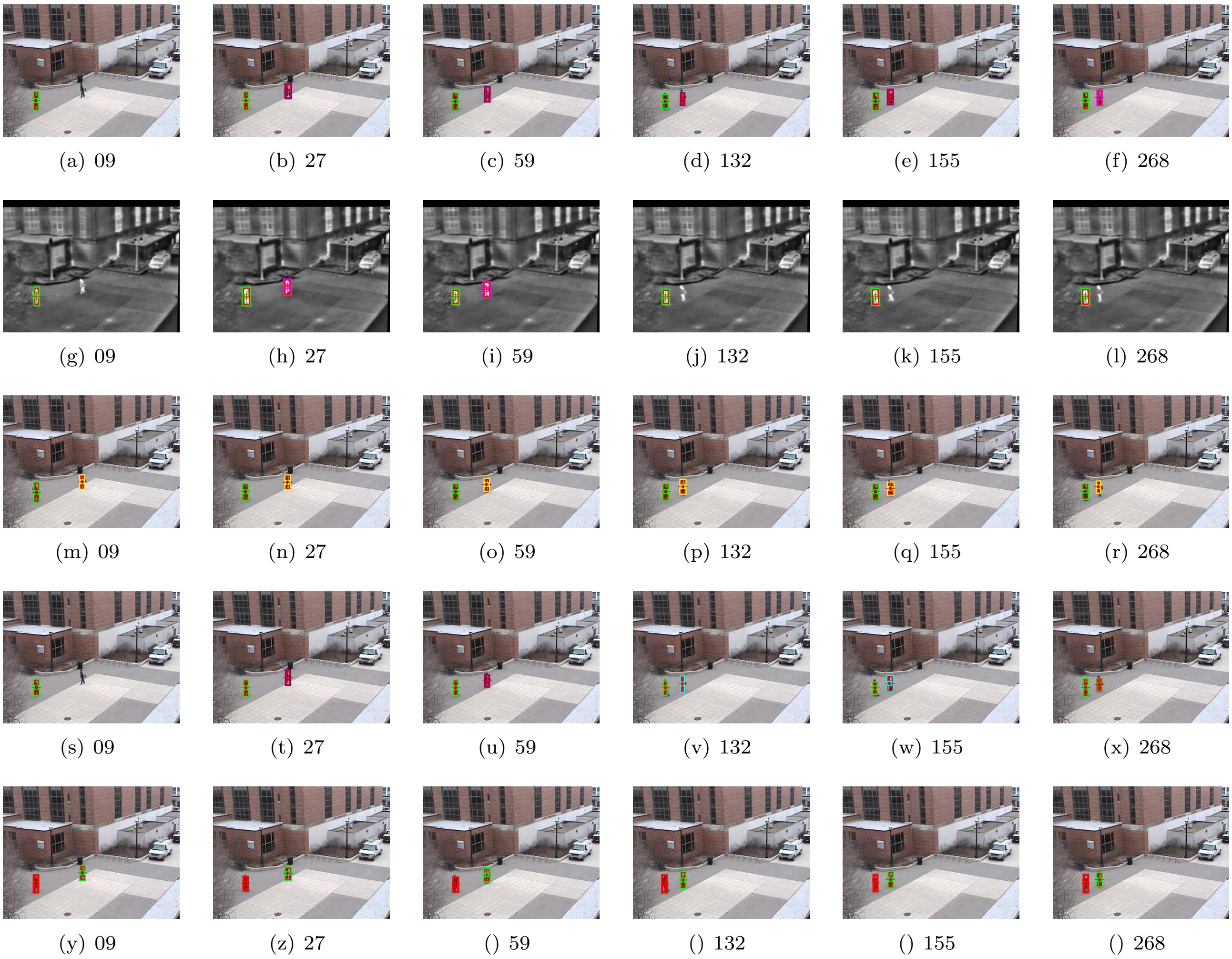
Figure 4. System Results for Scenario 1 - Top row shows the output of tracking using colour images only; second row shows the output of tracking using the thermal images only; third row shows results of tracking using fusion scheme 1; fourth row shows results of tracking using fusion scheme 4; fifth row shows results of tracking using fusion scheme 3.

overcomes this problem however, as in the event that one modality produces poor results, the system can ignore this modality entirely and fall back on the second to update the system until both modalities are producing usable results.

5 Conclusions and Future Work

In this paper, we have described a multi-sensor tracking system that combined visual and thermal data to obtain better tracking performance than can be achieved using either mode individually. We have shown that greater improvement can be achieved by leaving any fusion until the later stages of the tracking system, as fusion too early can result in errors from one modality being propagated through the system. Fusing late in the process allows more control and greater flexibility over what information we choose to use or ignore. We have also described a novel condensation filter algorithm that allows for a more flexible, computationally efficient and robust system by allowing both the number of particles used, and types of features used to change dynamically.

Future work will focus on developing features for the particle filter that better integrate the features of the motion detection and optical flow, as well as colour and shape information; and on utilising the fusion of visual and thermal data for other tasks such as abandoned ob-



(a) 09 (b) 27 (c) 59 (d) 132 (e) 155 (f) 268

(g) 09 (h) 27 (i) 59 (j) 132 (k) 155 (l) 268

(m) 09 (n) 27 (o) 59 (p) 132 (q) 155 (r) 268

(s) 09 (t) 27 (u) 59 (v) 132 (w) 155 (x) 268

(y) 09 (z) 27 () 59 () 132 () 155 () 268

Figure 5. System Results for Scenario 2 - Top row shows the output of tracking using colour images only; second row shows the output of tracking using the thermal images only; third row shows results of tracking using fusion scheme 1; fourth row shows results of tracking using fusion scheme 4; fifth row shows results of tracking using fusion scheme 3.

ject detection. Further investigation into other middle and late fusion approaches will also be carried out.

Acknowledgments

This project was supported by the Australian Government Department of the Prime Minister and Cabinet

References

[1] R. S. Blum and Z. Liu. *Multi-Sensor Image Fusion and Its Applications*. CRC Press, Boca Raton, FL, 2006.

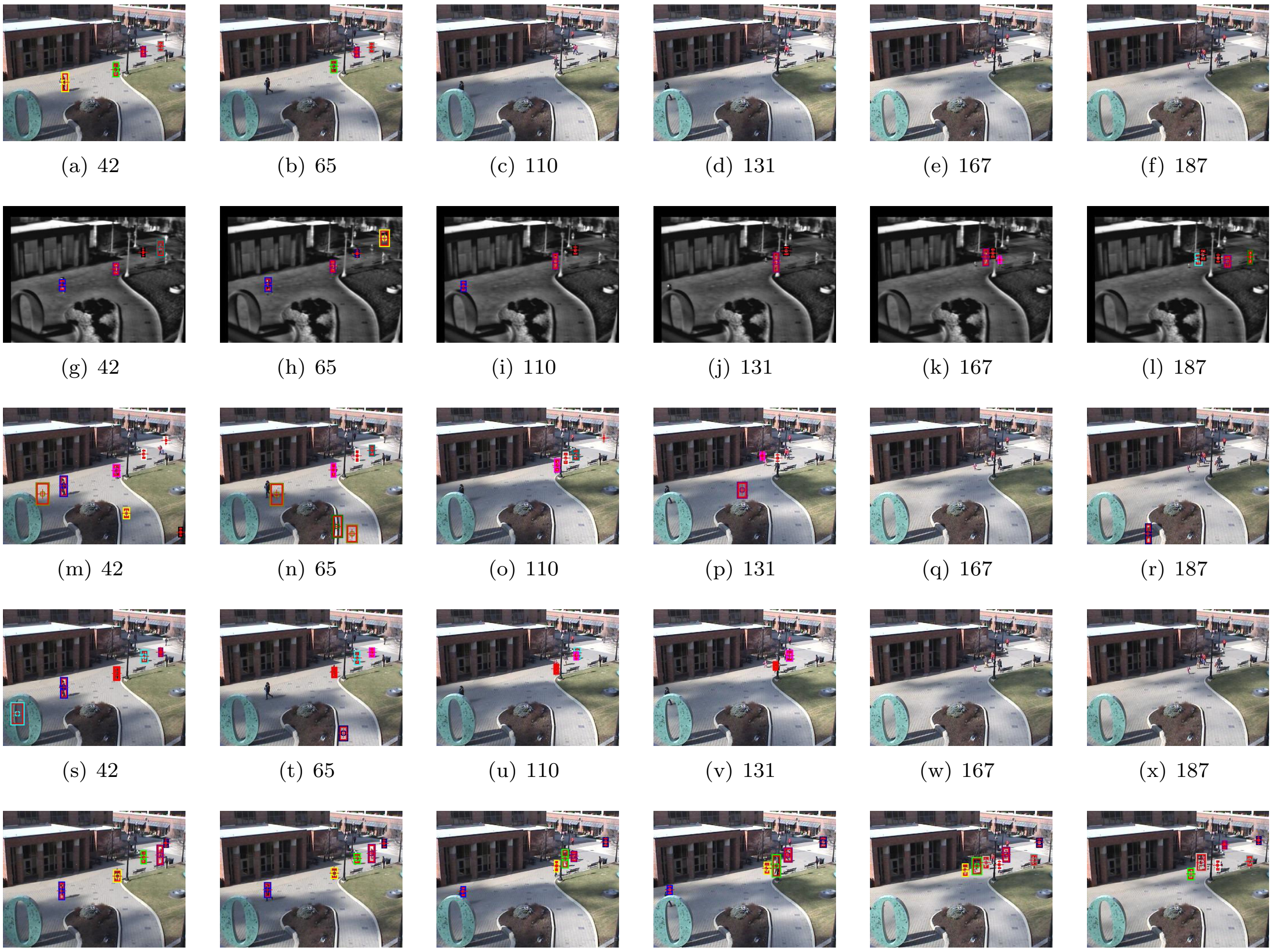
[2] C. O. Conaire, E. Cooke, N. O'Connor, N. Murphy, and A. Smearson. Background modelling in infrared and visible spectrum video for people tracking. volume 3, pages 20–20, 2005.

[3] C. O. Conaire, N. E. O'Connor, E. Cooke, and A. F. Smeaton. Multispectral object segmentation and retrieval in surveillance video. pages 2381–2384, 2006.

[4] J. Davis and V. Sharma. Ieee otcbvs ws series bench fusion-based background-subtraction using contour saliency. In *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum*, 2005.

[5] S. Denman, V. Chandran, and S. Sridharan. Adaptive optical flow for person tracking. In *Digital Image Computing: Techniques and Applications*, Cairns, Australia, 2005.

[6] S. Denman, V. Chandran, and S. Sridharan. Person tracking using motion detection and optical flow. In *The 4rd Workshop on the Internet, Telecommunications and Signal Processing*, Noosa, Australia, 2005.



(a) 42 (b) 65 (c) 110 (d) 131 (e) 167 (f) 187

(g) 42 (h) 65 (i) 110 (j) 131 (k) 167 (l) 187

(m) 42 (n) 65 (o) 110 (p) 131 (q) 167 (r) 187

(s) 42 (t) 65 (u) 110 (v) 131 (w) 167 (x) 187

(y) 42 (z) 65 () 110 () 131 () 167 () 187

Figure 6. System Results for Scenario 3 - Top row shows the output of tracking using colour images only; second row shows the output of tracking using the thermal images only; third row shows results of tracking using fusion scheme 1; fourth row shows results of tracking using fusion scheme 4; fifth row shows results of tracking using fusion scheme 3.

[7] A. Doucet. On sequential simulation-based methods for bayesian filtering. Technical report cued/f-infeng/tr 310, Department of Engineering, Cambridge University, 1998.

[8] J. Han and B. Bhanu. Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40(6):1771–1784, 2007.

[9] I. Haritaoglu, D. Harwood, and L. Davis. An appearance-based body model for multiple people tracking. In A. P. O. Sanfeliu, editor, *Vol 4; Applications, robotics systems and architectures Pattern recognition*, pages 184–187, Barcelona, Spain, 2000. IEEE Computer Society; 2000.

[10] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *IJCV*, 1998.

[11] C. O'Conaire, N. E. O'Connor, E. Cooke, and A. F. Smeaton. Comparison of fusion methods for thermo-visual surveillance tracking. pages 1–7, 2006.

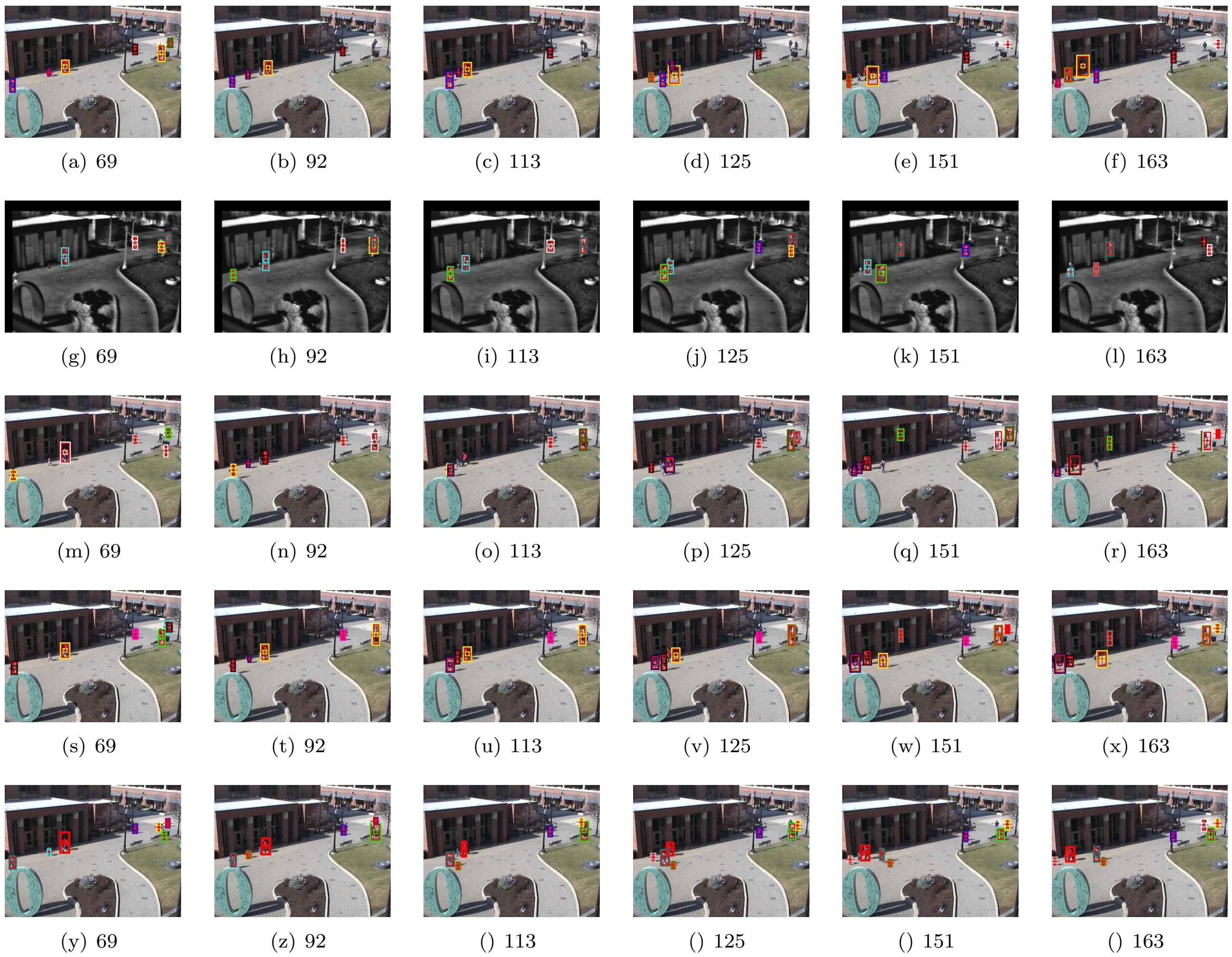
[12] K. Okuma, A. Taleghani, N. d. Freitas, J. Little, and D. Lowe. A boosted particle filter: Multitarget detection and tracking. In *ECCV*, pages 28–39, 2004.

[13] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

[14] G. Slabaugh, W. Culbertson, T. Malzbender, M. Stevens, and R. Schafer. Methods for volumetric reconstruction of visual scenes, 2003.

[15] J. Vermaak, A. Doucet, and P. Perez. Maintaining multi-modality through mixture tracking. In *Ninth IEEE International Conference on Computer Vision (ICCV'03)*, Nice, France, 2003.

[16] T. Zhao and R. Nevatia. Tracking multiple humans



(a) 69 (b) 92 (c) 113 (d) 125 (e) 151 (f) 163

(g) 69 (h) 92 (i) 113 (j) 125 (k) 151 (l) 163

(m) 69 (n) 92 (o) 113 (p) 125 (q) 151 (r) 163

(s) 69 (t) 92 (u) 113 (v) 125 (w) 151 (x) 163

(y) 69 (z) 92 () 113 () 125 () 151 () 163

Figure 7. System Results for Scenario 4 - Top row shows the output of tracking using colour images only; second row shows the output of tracking using the thermal images only; third row shows results of tracking using fusion scheme 1; fourth row shows results of tracking using fusion scheme 4; fifth row shows results of tracking using fusion scheme 3.

in complex situations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1208–1221, 2004.