Multi-Modal Object Tracking using Dynamic Performance Metrics

Simon Denman, Clinton Fookes, Sridha Sridharan, David Ryan
Image and Video Laboratory, Queensland University of Technology
GPO Box 2434, Brisbane 4001, Australia
{s.denman, c.fookes, s.sridharan, david.ryan}@qut.edu.au

Abstract

Intelligent surveillance systems typically use a single visual spectrum modality for their input. These systems work well in controlled conditions, but often fail when lighting is poor, or environmental effects such as shadows, dust or smoke are present. Thermal spectrum imagery is not as susceptible to environmental effects, however thermal imaging sensors are more sensitive to noise and they are only gray scale, making distinguishing between objects difficult. Several approaches to combining the visual and thermal modalities have been proposed, however they are limited by assuming that both modalities are perfuming equally well. When one modality fails, existing approaches are unable to detect the drop in performance and disregard the under performing modality. In this paper, a novel middle fusion approach for combining visual and thermal spectrum images for object tracking is proposed. Motion and object detection is performed on each modality and the object detection results for each modality are fused base on the current performance of each modality. Modality performance is determined by comparing the number of objects tracked by the system with the number detected by each mode, with a small allowance made for objects entering and exiting the scene. The tracking performance of the proposed fusion scheme is compared with performance of the visual and thermal modes individually, and a baseline middle fusion scheme. Improvement in tracking performance using the proposed fusion approach is demonstrated. The proposed approach is also shown to be able to detect the failure of an individual modality and disregard its results, ensuring performance is not degraded in such situations.

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, or under other environmental conditions. Thermal imaging is immune to many of these environmental effects, but offers reduced ability to discriminate between different objects and increased noise. 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 or thermal infrared only for surveillance and tracking.

Fusion for object tracking can be carried out at several points within the tracking process (see Figure 1). Early fusion involves fusing the images prior to any processing. Blum et al. [1] proposed different methods of early image fusion using the wavelet transform and the pyramid transform, which can combine the images prior to entering a tracking system. Middle fusion occurs within the tracking system, such as fusion during object detection or matching. O'Conaire et al. [9] proposed a middle fusion scheme that utilised a multi-modal appearance model to perform fusion during the tracking process. Torresan et al [11] propose a middle fusion scheme where motion segmentation and blob extraction is applied to each modality, and objects from each mode are matched to a combined object set. Leykin et al. [7] proposed the use of a multi-modal background model and particle filter to track objects in a multi-modal environment. Late fusion performs tracking across each mode independently for each frame, and then seeks to fuse the resultant tracked object lists. Denman et al. [5] evaluated four fusion schemes for use in a multi-spectral tracking system (one early fusion, two middle fusion, and one late fusion) and found that a middle fusion scheme that fused object detection results performed best.

A common limitation of these approaches is that they assume that each modality is performing equally well, and if one modality begins to perform poorly there is no mechanism to recognise the performance decrease and deemphasise (or completely disregard) the mode. In [4], Denman et al. proposes several performance metrics that can be computed in real time to evaluate the performance of an object tracking system as it runs. In this paper, we propose and evaluate a middle fusion scheme that uses a dynamic performance metric to determine the fusion weights for the modalities, allowing the system to recognise if a modality is

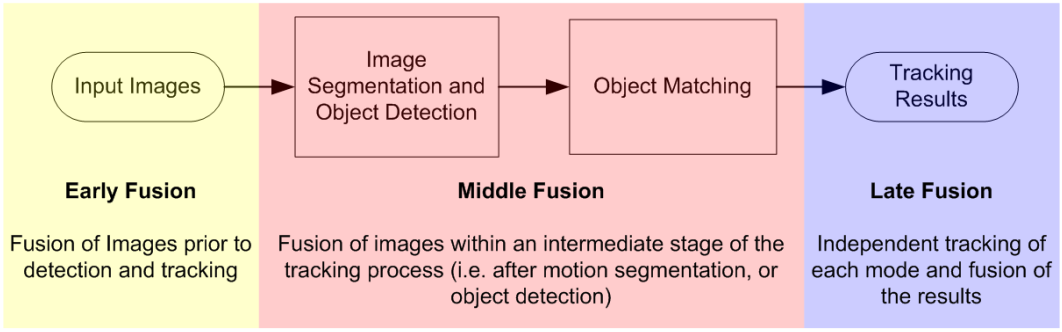


Figure 1. Early, Middle and Late Fusion for Object Tracking

failing and disregard its output. We compare the proposed fusion scheme with using the visual and thermal domains on their own and the best performing fusion scheme from [5], and demonstrate that the system is able to outperform either modality individually and the baseline approach. We also show that the proposed approach is able to recognise a poorly performing modality, and reduce its influence on the tracking process.

2. Object Tracking System

The tracking system proposed in [5] is used in this work. The object tracking system uses a hybrid motion detector-optical flow technique [3] as a basis, and scans for appropriate regions of motion to detect people (see figure 2). A scalable condensation filter [5] is used to track the people.

The condensation filter uses the input images and the results of the motion detection, and progressively updates features for each tracked object 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 which determine their precise locations in the image. Motion associated with these detected people is removed from the motion image as it is now accounted for. The remaining motion is assumed to belong to new people, and so person detection is carried out on remaining areas to locate people who have recently 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 [6, 12]. Working within sub-regions overcomes problems caused by people occupying a common column of the image causing inaccurate vertical projections. This detection 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.

When new objects are added to the system, they are considered to be *preliminary* until they have been detected for n_{prelim} consecutive frames ($n_{prelim} = 3$ in this paper). This is done to prevent erroneous objects from being added to the system as a result of a single false detection.

3. Proposed Fusion System

The study in [5] compared four simple fusion schemes for combining visual and thermal modalities. Three approaches to fusion were evaluated:

1. Early Fusion - Fusing incoming images such that only a single image was used by the object tracker.
2. Middle Fusion - Partially processing two images and fusing after an intermediate stage (i.e. motion detection or object detection).
3. Late Fusion - Track the modes independently and fuse the resulting lists of tracked objects.

It was found that the most suitable fusion scheme is a middle fusion scheme where fusion is performed after object detection has been carried out on each mode. In this approach, motion detection and object detection are carried out on both modalities, and the two object lists are used to update the central list of tracked objects. Objects that have been previously detected can be updated by a detection from either domain. For a new object to be added, the object must be detected in both, or in the modality where it is not detected, there must be a minimum amount of motion within the region where the object has been detected (attempting to ensure that a false detection in one modality, does not lead to a non-existent track being initialised). Given this, a novel middle fusion approach that dynamically determines the performance of each modality to improve fusion and reduce errors is proposed.

The single mode tracking system described in Section 2 is modified to allow multiple modalities as inputs. Motion detection and object detection are performed on both modalities, and the two object lists are used to update the central list of tracked objects and add new objects. Features used by the scalable condensation filter (SCF) are able to receive information from both modalities. The features

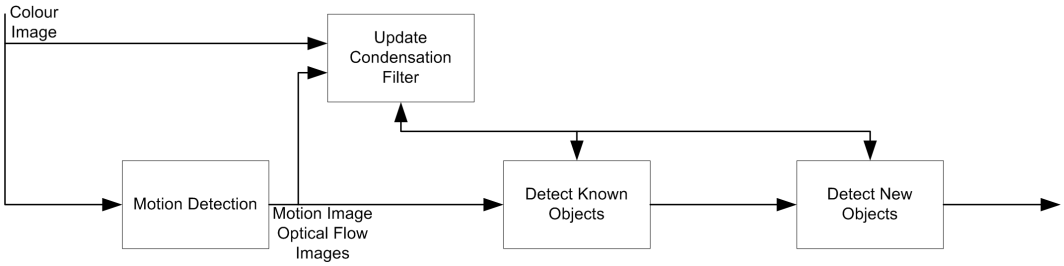


Figure 2. Tracking System Flowchart

can be used to compare and match objects in either domain individually, or to both simultaneously (an update can also be performed on only a single domain, or both). Figure 3 shows a flowchart of the proposed system.

It is important that the fusion system is able to recognise when a modality becomes unreliable or is performing poorly to ensure that poor performance of one mode does not adversely effect the whole system. To determine modality performance, the proposed system continuously monitors the performance of the object detection for each mode to ensure that when one modality is less reliable, observations from that modality are weighted less or ignored.

Object detection performance is gauged by comparing the number of objects detected to the number of objects currently being tracked. Ideally, the number of objects detected should be the same as the number of objects that are being tracked. This performance measure is simple to calculate and can be easily integrated into an existing tracking system framework. Whilst the proposed measure is not ideally suited to detecting local disturbances, it is well suited to detecting problems affecting the whole scene (i.e. lighting changes), and as such it is felt that it is a suitable measure for a preliminary study of a dynamically weighted fusion system.

Motion detection followed by object detection is performed on each modality independently resulting in two object lists, $O_{vis,t}$ and $O_{ther,t}$, of size $n_{vis,t}$ and $n_{ther,t}$ respectively. Ideally, each of these object lists should each contain the number of objects presently in the scene, n_t. The number of objects detected, when compared to the number of objects present, is used to determine the performance of each modality at a given time,

$$n_t \leq \alpha;\ q_{vis,t} = q_{ther,t} = 1, \quad (1)$$

$$q_{vis,t} = 1 - \frac{min(max(|n_t - n_{vis,t}| - \alpha, 0), n_t)}{max(max(|n_t - n_{vis,t}| - \alpha, 0), n_t)} \quad (2)$$

$$q_{ther,t} = 1 - \frac{min(max(|n_t - n_{ther,t}| - \alpha, 0), n_t)}{max(max(|n_t - n_{ther,t}| - \alpha, 0), n_t)} \quad (3)$$

where $q_{vis,t}$ and $q_{ther,t}$ are the performance measures for the visual and thermal modalities respectively for the frame at time t. A tolerance of α objects is allowed (within the proposed system α is set to 1) when comparing the number of detections. This tolerance ensures that when the system contains no objects, the appearance of an object does not result in the performance of the system dropping significantly (this is also dealt with through the use of a leaning rate to curb rapid changes in the performance metric, see Equations 4 and 5). Whilst multiple objects entering and exiting the scene will result is a drop in performance for the modality, ideally this drop should be uniform across each modality. As it is the difference in performance between the modalities that is important, this is not a problem. The metric is also designed such that it is the relative error, rather than the absolute error, that it is considered (i.e. an error of 3 detections when only 3 objects are present is considered more severe than an error of 3 detections when 6 objects are present).

The performance for a given frame ($q_{vis,t}$ and $q_{ther,t}$) is incorporated into a global performance metric ($p_{vis,t}$ and $p_{ther,t}$ for the visual and thermal modalities respectively) which is adjusted gradually,

$$p_{vis,t} = p_{vis,t-1} + \frac{q_{vis,t} - p_{vis,t-1}}{L}, \quad (4)$$

$$p_{ther,t} = p_{ther,t-1} + \frac{q_{ther,t} - p_{ther,t-1}}{L} \quad (5)$$

where L is the learning rate for the performance metric. Initial performance metrics may be specified within the system configuration, such that one modality is weighted higher than another by default (i.e. if it is known that one modality is less reliable for a given scene it can be by default set to a lower value).

These performance metrics are used to determine the weighting applied to each modality when fusing object lists, and adding objects. The relative strength of each modality for the task of object detection is calculated,

$$w_{vis,t} = \frac{p_{vis,t}}{p_{vis,t} + p_{ther,t}}, \quad (6)$$

$$w_{ther,t} = \frac{p_{ther,t}}{p_{vis,t} + p_{ther,t}} \quad (7)$$

where $w_{vis,t}$ is the performance of the visual modality relative to the thermal, and $w_{ther,t}$ is the performance of the

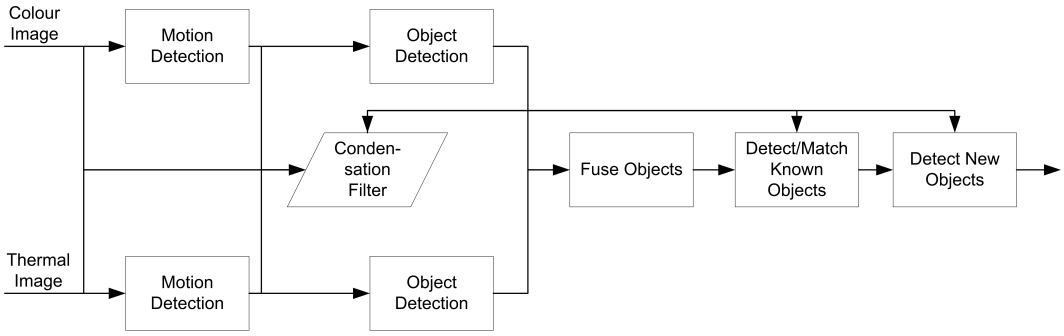


Figure 3. Flowchart for Proposed Fusion System

thermal modality relative to the visual. This process ensues that the weights of each modality sums to 1, which simplifies the process of merging objects.

The two object lists are merged, by determining the overlap between the objects. If the overlap between the two objects is greater than a threshold, T_{ov}, the objects are merged. For each object, there are several parameters such as the bounding box, centroid and velocities. Each of these values is merged according to the equation

$$O_{fused,t,i} = O_{vis,t,j} \times w_{vis,t} + O_{ther,t,k} \times w_{ther,t}. \quad (8)$$

where $O_{vis,t,j}$ is the visual object being merged, $O_{ther,t,k}$ is the thermal object being merged and $O_{fused,t,i}$ is the resultant fused object.

This yields three objects lists, $O_{fused,t}$, $O'_{vis,t}$ and $O'_{ther,t}$, representing the fused objects, the remaining visible and remaining thermal objects respectively. The updated lists of visual and thermal objects are defined as

$$O'_{vis,t} = O_{vis,t} \notin O_{fused,t}, \quad (9)$$
$$O'_{ther,t} = O_{ther,t} \notin O_{fused,t}. \quad (10)$$

The object lists are used to update the known tracks and add new tracks to the system. This update is performed in the following order:

1. Match objects in the merged list, $O_{fused,t}$, to the tracked list
2. Match objects in the individual lists ($O'_{vis,t}$ and $O'_{ther,t}$) to the tracked list, such that the best fitting object from either list is matched in turn
3. Add new objects within the merged list
4. Add new objects within the individual lists

The fourth stage involves additional checks to ensure that invalid objects are not added. A prerequisite amount of motion must be present within the other modality for such detections to be valid (set as a constant, rather than derived from the performance metrics), and the performance of the modality must be greater than a threshold, T_a (set to 0.5 within the proposed system).

Objects that are only added from a single mode are considered to be *preliminary* until they have been continuously detected for a required number of frames,

$$n_{active}(i) = \frac{n_{active}}{p_{m,t}} \quad (11)$$

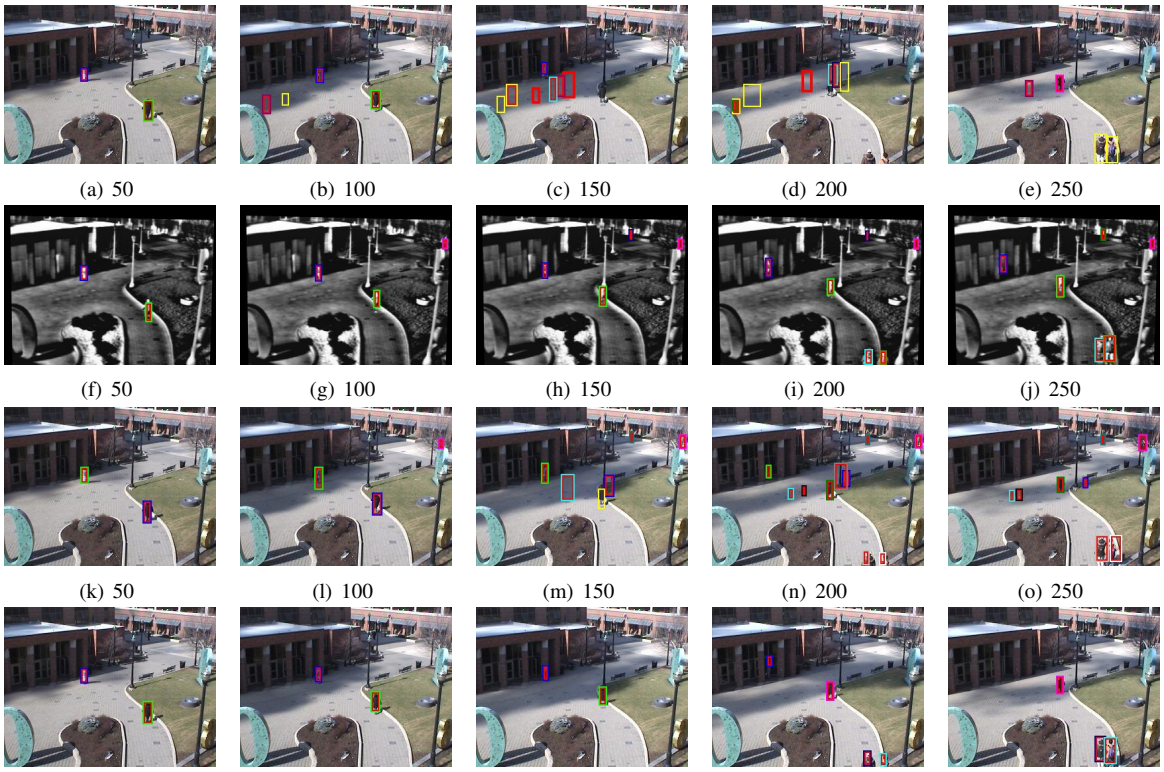
where $n_{prelim}(i)$ is the active threshold for object i (the object being added), n_{prelim} is the default threshold (see Section 2) and $p_{m,t}$ is the performance for the modality m, from which the object is being added. However, when an object that is in this state is detected in both modalities (i.e. it is updated from an object in the $O_{fused,t}$ list), the threshold is decremented by one (along with the increment in the detection count, in effect this counts as two detections). This is aimed at preventing invalid objects from being added to the system.

Even when both modalities are performing poorly, the system will still be capable of recovery and correct operation provided the value of T_a (performance modality threshold to add an object) is set appropriately. Provided objects can still be added from one or more modes, additional tracks that are detected can be added and tracked and the system can recover. With the exception of adding objects (where T_a is important), the system only considers the weights of the modes relative to one another and so existing objects can still updated and tracked, even if both modalities are performing poorly. If T_a prevents objects from being added, the system will continue to be in error until the number of objects present in the scene begins to drop at which time the system can begin to recover.

The performance weights are not used by the particle filter (i.e. the particle filter does not weight one modality above another).

4. Results

The proposed algorithm is evaluated by considering object tracking accuracy. The OTCBVS Benchmark Data



(a) 50 (b) 100 (c) 150 (d) 200 (e) 250

(f) 50 (g) 100 (h) 150 (i) 200 (j) 250

(k) 50 (l) 100 (m) 150 (n) 200 (o) 250

(p) 50 (q) 100 (r) 150 (s) 200 (t) 250

Figure 4. Example System Results for Location 1, Set 1, as cloud cover approaches - The top row is the output of the colour modality only, the second row is the thermal modality only, the third row is the baseline middle fusion system [5] and the fourth row is the proposed fusion algorithm.

set Collection [2] is used to evaluate the proposed fusion scheme. This database contains aligned thermal infrared and colour image sequences of two different outdoor scenes containing a varying number of pedestrians.

Four tracking systems are tested in the evaluation, the colour and thermal modalities individually, the middle fusion system from [5] upon which the proposed system is based, and the proposed system. The Colour and Thermal modality systems simply use the single mode tracking system outlined in Section 2, with their respective modalities as inputs.

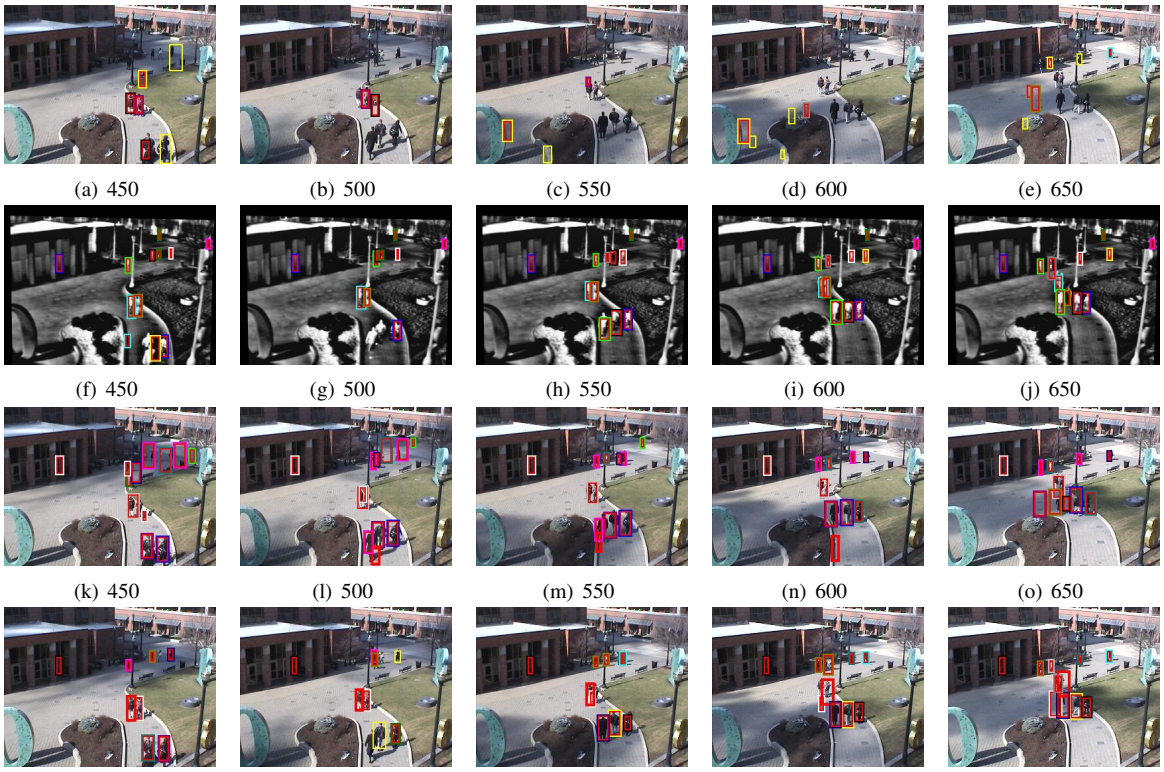
Location	Set Number	First Frame	Last Frame
1	1	1	1054
	2	1	600
2	4	1	180
	5	1	1400
	6	1	1150

Table 1. Evaluation Data Structure

Five sub-sequences from the OTCBVS database are selected to highlight various situations of interest such as stationary people, occlusions, people moving in shadowed areas, and shadowing caused by cloud cover. Two sequences from the first location, and three from the second are used. Separate results are shown for each location, as the second location contains significantly simpler scenarios than the first. Sub-sequences have been selected to ensure that there is correct synchronisation between the colour and thermal data. Table 1 shows the data used in this evaluation. It should be noted that the sequences from Location 1 are complete, whilst those from Location 2 have been cropped at a point where synchronisation between the modalities is lost. Ground truth tracking data has been computed for each of these sequences using the VIPER toolkit[1]. Ground truth is calculated on the thermal view only for all sequences.

Tracking results are evaluated using the ETISEO evaluation tool [8]. The ETISEO evaluation defined several metrics for gauging the performance of tracking systems which are split into five groups, Detection, Localisation, Tracking, Classification and Event Recognition. Each group of metrics contains several sub-metrics to evaluate specific criteria and a global metric, which is the average of all metrics within the group. Our evaluation will use the overall metrics for detection, localisation and tracking (there is

[1] ViPER-GT is a ground truth authoring tool and can be downloaded from http://vipertoolkit.sourceforge.net/docs/gt/



(a) 450 (b) 500 (c) 550 (d) 600 (e) 650

(f) 450 (g) 500 (h) 550 (i) 600 (j) 650

(k) 450 (l) 500 (m) 550 (n) 600 (o) 650

(p) 450 (q) 500 (r) 550 (s) 600 (t) 650

Figure 5. Example System Results for Location 1, Set 1, with a large portion of cloud cover - The top row is the output of the colour modality only, the second row is the thermal modality only, the third row is the baseline middle fusion system [5] and the fourth row is the proposed fusion algorithm.

only one type of object being tracked in the system - people, and there is no event recognition, so classification and event recognition performance are of no interest). All metrics yield a value in the range $[0, 1]$, with 1 being a perfect result, and 0 being complete failure. Detailed information on how the metrics are formulated can be found in [10].

Algorithm	Overall Detection	Overall Localisation	Overall Tracking
Colour	0.41	0.86	0.31
Thermal	0.67	0.94	0.54
Baseline MF	0.57	0.92	0.50
Proposed	0.73	0.94	0.56

Table 2. Evaluation Results for Location 1 - Baseline MF is the middle fusion system from [5] upon which the proposed algorithm is based.

Table 2 shows results for the Location 1 data sets. It can be seen that the proposed system outperforms all others. The scenes at Location 1 contain a large amount of moving cloud cover that results in a lot of false motion being detected within the colour modality, adversely affecting performance (see Figures 4 and 5). The thermal system cannot see the cloud cover and so is unaffected. The proposed system is able to recognise the poor performance of the colour modality and as a result the modality is largely ignored.

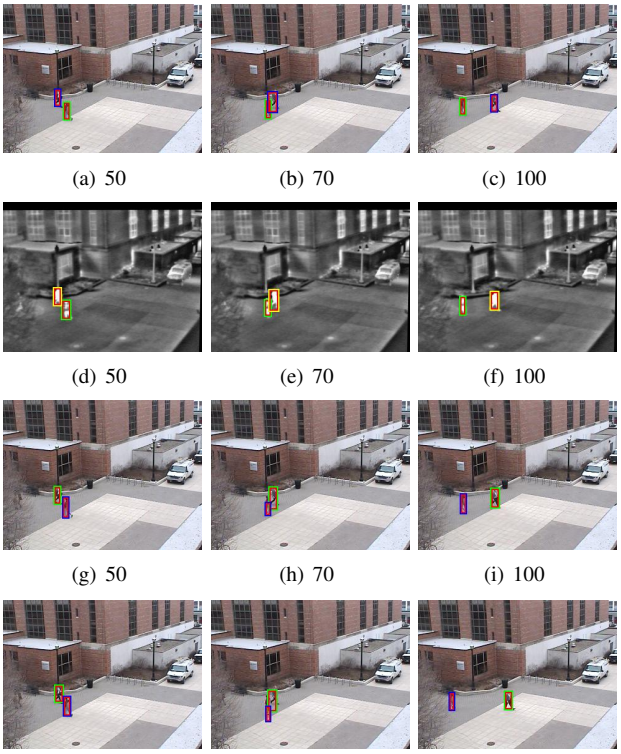
Figures 4 and 5 show output from Location 1, Set 1. In Figure 4, as cloud cover begins to cover the scene the colour system begins to struggle. The baseline middle fusion system is able to continue tracking objects, however false objects are detected and tracked for short periods of time as the system is unaware of the problems with the colour modality. The thermal system is unaffected by the changing conditions (the colour cover cannot be seen in the thermal spectrum), and the proposed system is able to adjust and continues to track correctly.

In Figure 5, the cloud cover has increased and now covers most of the scene. The colour system has failed completely by this stage. The baseline middle fusion system is still able to track the objects in the scene, however it also detects and attempts to track many false objects caused by the cloud. The nature of the thermal modality means that it is unaffected, and the proposed system is able to recognise the poor performance of the colour modality and disregard its performance.

Table 3 and Figure 6 show the evaluation results for Lo-

Algorithm	Overall Detection	Overall Localisation	Overall Tracking
Colour	0.60	0.91	0.72
Thermal	0.63	0.92	0.75
Baseline MF	0.68	0.93	0.76
Proposed	0.72	0.93	0.83

Table 3. Evaluation Results for Location 2 - Baseline MF is the middle fusion system from [5] upon which the proposed algorithm is based.



(a) 50 (b) 70 (c) 100

(d) 50 (e) 70 (f) 100

(g) 50 (h) 70 (i) 100

(j) 50 (k) 70 (l) 130

Figure 6. Example System Results for Location 2, Set 4 - The top row is the output of the colour modality only, the second row is the thermal modality only, the third row is the baseline middle fusion system [5] and the fourth row is the proposed fusion algorithm.

cation 2. As can be seen, the proposed fusion algorithm offers a noticeable improvement over both individual modalities, and the baseline fusion scheme from [5].

Figure 6 shows a simple example sequence from Location 2, Set 4 in which two people occlude one another. All systems are able to resolve the occlusion correctly. In ideal lighting conditions, with both modalities performing correctly, all systems are able to function correctly.

Figure 7 illustrates the change in the performance scores for the modalities during a scene. Figure 7 shows a scene in which a large amount of cloud cover moves over the scene (Location 1, Set 1). Performance of the colour modality drops rapidly as cloud begins to cover the scene (frame 200). Performance of both modalities remains stable until a increase in the number of people in the scene at frame 500. At this point, due to the cloud cover, the colour modality is able to detect very few of the people present and performance drops further. The thermal modality continues to perform well until frame 700, when several occlusions in the centre of the frame result in poor detection and tracking (see frame 800). As the occlusions pass, the thermal modality begins to recover. The colour modality begins recovers at frame 800 as the cloud clears, however continuing cloud cover (see frame 1000) prevents a full recovery.

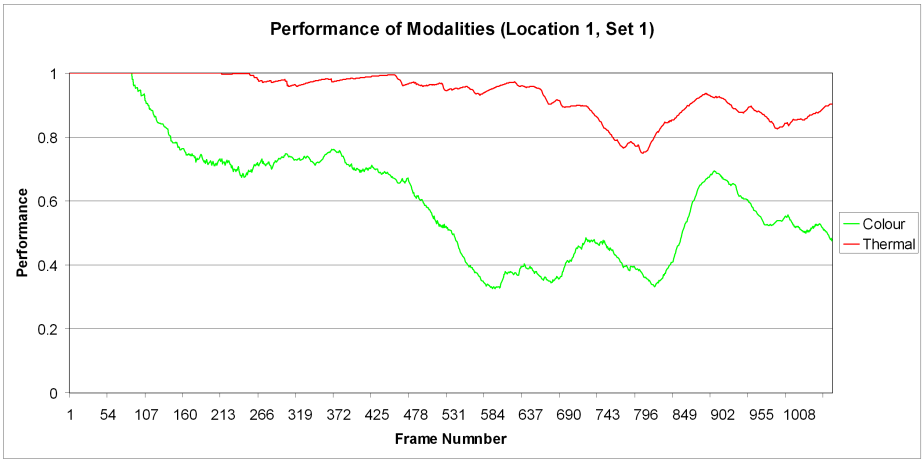
As Figure 7 shows, the proposed algorithm is able to recognise when a modality is in error and adjust its performance score (and thus weight) accordingly. However, the proposed approach is limited in that it only considers performance from a global (i.e. the whole scene) perspective. In both examples, the whole scene is not equally affected by the shadowing, with portions of the scene remaining unaffected and in full sunlight.

Ideally, the fusion system would be able to recognise that any results from the visual modality in the completely shadowed errors are to be disregarded, whilst observations from the unaffected areas are still safe to use. To overcome this, performance metrics that can dynamically determine the performance of the modalities on a localised basis (possibly in addition to more global metrics) are required.

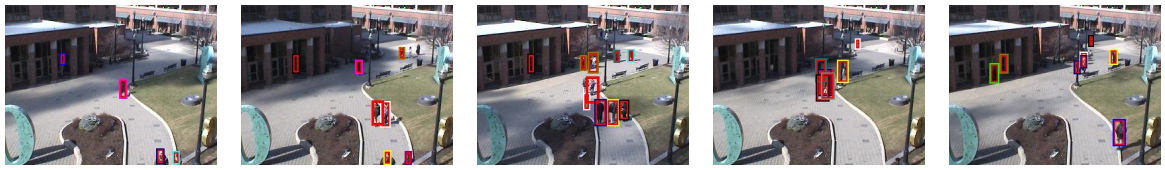
It is expected that a localised application of the performance metrics would also help when both modalities are in error. In such a situation, it is less likely that both modalities would be in error in the same location at the same time, allowing the system to favour the appropriate modality at each location. However, even if both modalities are failing at a given location, it is expected that the system should still be able to function correctly provided the value of T_a is not preventing objects from being added (see Section 3 for more details) as it is the performance of the modalities relative to one another that is important when detecting and updating locations. An evaluation of the proposed fusion scheme when both modalities are performing poorly will be undertaken when data becomes available.

5. Conclusion

This paper has proposed a novel middle fusion approach for combining colour visual spectrum and thermal spectrum images for object tracking. The proposed fusion scheme performs motion and object detection on each modality separately, and the resultant object lists are fused using dynamically determined fusion weights. Fusion weights are calculated by monitoring the performance of each modality. The number of objects detected by each modality is compared to the number of objects being tracked by the system to assess performance. The proposed fusion scheme is evaluated using a portion of the OTCBVS database and is shown to



(a) 130



(b) 200 (c) 400 (d) 600 (e) 800 (f) 1000

Figure 7. Performance of the Visual and Thermal Modality for Location 1, Set 1

outperform both the visual and thermal modalities on their own, and the baseline middle fusion system.

Future research will investigate other metrics to dynamically evaluate the performance of the modalities at both a global and local level to improve fusion. The use of multiple metrics (i.e. for motion segmentation and object detection) will also be investigated. Further testing in situations where both modalities are unreliable, and where both modalities are performing poorly will also be carried out as data becomes available.

References

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

[2] 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

[3] S. Denman, V. Chandran, and S. Sridharan. Adaptive optical flow for person tracking. In *Digital Image Computing: Techniques and Applications*, pages 8–8, Cairns, Australia, 2005. 2

[4] S. Denman, C. Fookes, S. Sridharan, and R. Lakemond. Dynamic performance measures for object tracking systems. In *6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Genoa, Italy, 2009. 1

[5] S. Denman, T. Lamb, C. Fookes, S. Sridharan, and V. Chandran. Multi-sensor tracking using a scalable condensation filter. In *International Conference on Signal Processing and Communication Systems (ICSPCS)*, volume 1, pages 429–438, Gold Coast, QLD, 2007. 1, 2, 5, 6, 7

[6] I. Haritaoglu, D. Harwood, and L. Davis. An appearance-based body model for multiple people tracking. In *15th International Conference on Pattern Recognition*, volume 4, pages 184–187, Barcelona, Spain, 2000. 2

[7] A. Leykin, Y. Ran, and R. Hammoud. Thermal-visible video fusion for moving target tracking and pedestrian classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 –8, 2007. 1

[8] A. T. Nghiem, F. Bremond, and M. T. V. Valentin. Etiseo, performance evaluation for video surveillance systems. In *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 476–481, London, UK, 2007. 5

[9] C. O'Conaire, N. E. O'Connor, E. Cooke, and A. F. Smeaton. Comparison of fusion methods for thermo-visual surveillance tracking. In *9th International Conference on Information Fusion (ICIF)*, pages 1–7, 2006. 1

[10] Silogic and Inria. Etiseo metrics definition (http://www-sop.inria.fr/orion/etiseo/download.htm). Technical report, 6th January 2006. 6

[11] H. Torresan, B. Turgeon, C. Ibarra-Castanedo, P. Hebert, and X. P. Maldague. Advanced surveillance systems: combining video and thermal imagery for pedestrian detection. In *Thermosense XXVI*, volume 5405, pages 506–515, Orlando, FL, USA, 2004. SPIE. 1

[12] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208–1221, 2004. 2