Moving Object Detection Using Particle Filter by Adaptive Fusion of Thermal and Visible Camera Data

Vrushali Mahajan, Anuja Jakhade

Dept. of Electronics and Telecommunications
Cummins College of Engineering, Pune 411052, India.

***Abstract*--This paper presents a method for tracking a moving target by fusing visual information from a thermal camera and a visible spectrum color camera.The method used for tracking objects is color based particle filtering.The algorithm selects the modality which distinguishes the foreground objects from the background. The method is evaluated by testing on a variety of challenging video sequences in which targets are camouflaged in either of imaging modalities.**

Keywords---- Bi-modal fusion, thermal imaging, Particle filter, divergence measures, Bhattacharyya distance, object tracking.

I. INTRODUCTION

Security and surveillance has received much research attention in recent years due to numerous world events. The desire to provide robust and accurate surveillance information has led to considerable research on methods to integrate information from different sensors.

In our work, we have presented a powerful algorithm by of capturing thermal infrared video along with standard visible spectrum for tracking objects in surveillance scenarios. We have implemented a method for tracking a moving target by combining visual information from a thermal camera and a conventional color camera. This helps in obtaining increased robustness against camouflaged and hidden objects.

Since objects of scene clutter and distracters are less likely to share common feature values with the target in both spectra simultaneously, tracking can be made more effective by fusingthe target features of both visible colors and heat signature.

We have implemented a color based tracking method using particle filter.The key innovation is a method for continuously relearning local background models for each frame in each imaging modality, comparing these against a model of the foreground object being tracked and then weighting the data from the two modalities in favor of whichever imaging modality is currently the most discriminating at each successive frame. In particular, we show how particle filtering with patch histograms can be used so as to make use of thermal data in addition to visible data.

II. PURPOSE OF DATA FUSION

Two main benefits of the joint use of thermal and visible sensors are:

1) The complementary nature of the two modalities that provides the thermal and color information of the scene.
2) The redundancy of information captured by the different modalities which increases the reliability and efficiency of a surveillance system.

The advantage of employing both complementary sources of datahelps in obtaining improved robustness against camouflaged and hidden targets.
For example, when the target is camouflaged against similar backgrounds in one imaging modality, the algorithm compares the two modalities and algorithm reduces the influence of the poor modality and relies more strongly on information from the more discriminating modality.

Fig.1 A thermal video frame where the moving object is undetectable due to the glass pane before the target.

III. TRACKING ALGORTIHM

A. Bhattacharyya Co-efficient

Divergence techniques for the distance between two probability distributions have been extensively researched. These measures are widely used in varied fields such as pattern recognition quantum information theory and signal detection. Distance measures try to achieve two main objectives: to process how "close" two distributions and how "easy" it is to distinguish between one pair than the other. There is plethora of distance measures available to assess the divergence of probability distributions. One of the most well-known and widely used divergence measures, the Kullback-Leibler divergence (KLD) can create problems in specific applications. Specifically, it is unbounded above and requires that the distributions be absolutely continuous with respect to each other. Various other information comparison measures have been studied keeping in view ease of computation and utility in problems of pattern recognition and signal selection. Of these measures, Bhattacharyya distance and Chernoff distance have been widely used in signal processing.

The Bhattacharyya coefficient is an approximate measurement of the amount of overlap between two successive frames. The coefficient can be used to determine the relative closeness of the two frames being considered.

The value of Bhattacharyya Coefficient ranges from '0' to '1'.A minimum value of Bhattacharyya Coefficient indicates that the foreground and the background frames do not overlap while a maximum value of Bhattacharyya Coefficient indicates that the two frames are identical.

Bhattacharyya Coefficient is calculated for thermal and visible data using patch histogram. The computed patch histograms are used to calculate the similarity between the foreground and background frames.

The learning and recognition of objects in cluttered scenes poses a crucial issue in computer vision. Here image patches are extracted at each position and stored in a histogram. The positions of the extracted patches are considered and provide a significant increase in the recognition performance. This method for plotting histograms has some immediate advantages: Changes in the geometrical relation between image parts can be modeled to be flexible or even to be ignored and the algorithm can focus on those image parts that are most important to recognize the object and can handle occlusions well. If parts of an object are occluded in an image, the remaining visible parts may still be used to recognize the object correctly and to learn about the appearance of this object from this instance.One problem with histograms is that the number of bins in a histogram grows exponentially with the number of dimensions of the data and they become difficult to handle if the dimensionality of the input data is large. For example, given 8 dimensional input data and only 4 subdivisions per dimension results in 48 = 65,536 bins. To avoid this problem, we used a patch representation of the histograms, i.e. we store only those bins whose content is not empty.

The patch histograms $\boldsymbol{h_{t1}}$ *and* $\boldsymbol{h_{t2}}$ are computed for thermal modality. Similarly, the patch histograms $\boldsymbol{h_{v1}}$ *and* $\boldsymbol{h_{v2}}$ are computed for the visible modality.

The following equation is used:

$$\boldsymbol{B_{thermal}} = \sum_{i=0}^{n} \sqrt{\boldsymbol{\Sigma h_{t1}.\Sigma h_{t2}}}$$

Where '**n**' is the number of partitions.

$$\boldsymbol{B_{visible}} = \sum_{i=0}^{n} \sqrt{\boldsymbol{\Sigma h_{v1}.\Sigma h_{v2}}}$$

Where '**n**' is the number of partitions.

The modality with greater Bhattacharyya Coefficient is selected automatically by the algorithm.

B. Particle Filter

Particle Filtering is used for object tracking. As shown by Nummiaro et al., [18], the tracking method adds an adaptive appearance model based on color distributions to particle filtering. The color-based tracker can efficiently handle non-rigid and fast moving objects under different appearance changes. Moreover, as multiple hypotheses are processed, objects can be well tracked in cases of occlusion or clutter.

The basic idea of particle filters is that any Probability Density Function can be represented as a set of samples (particles). Each particle has one set of values for the state variables.
At any time, ***t***, we represent the state **s*t*** of the tracked target by a distribution, approximated by a weighted set of *I*particles $p_0 \ldots p_i \ldots p_I$,with weights$w_0 .. w_i \ldots w_I$which are normalized so that:

$$\sum_{i=1}^{I} \boldsymbol{w_i} = \mathbf{1}$$

More powerful tracking should result from weighting particles using an appropriate combination of the above coefficients for each modality. But it is not obvious how the coefficients should be combined.

For example, if we simply took an arithmetic mean of the thermal and visible Bhattacharyya coefficients and then substituted this instead of the visible coefficient, then at times when one imaging modality is more discriminatory than the other, therefore selecting the wrong modality. Similarly, if we simply multiply the coefficients as a product of experts then we may not necessarily improve robustness - a modality that is performing poorly may spoil a meaningful coefficient from a modality that is performing well.

In general, some optimally weighted combination of the coefficients will be best. We therefore weight particles byusing:

$$\boldsymbol{B_{fused}} = \boldsymbol{\alpha}.\boldsymbol{B_{visible}} + (\mathbf{1} - \boldsymbol{\alpha}).\boldsymbol{B_{thermal}}$$

where $\boldsymbol{\alpha}$is a weighting factor (varying between 0 and 1) which is continuously relearned for every frame therefore making the data fusion process adaptive.

'**α**' is calculated as follows:

$$\alpha = \frac{B_{thermal}}{B_{thermal}+B_{visible}}$$

With appropriate normalization it is possible to assign$\boldsymbol{B_{fused}}$directly as the particle weight $\boldsymbol{w_i}$and this resultsin a reasonably effective tracker. However to make particleweights handle rather more like true probabilities, we can evaluate weights as:

$$w_i = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(1-Bfusedi)}{2\sigma^2}}$$

IV. EXPERIMENTAL RESULTS

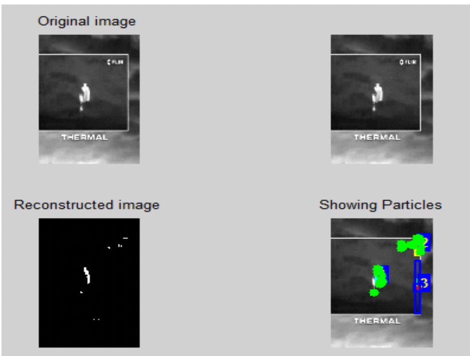
We have tested the algorithm on a variety of video sequences of many hundreds of frames each.We show here some salient examples which are designed to illustrate how our data fusion strategy has been successfully enabled.The algorithm can also successfully track multiple targets along with switching the modalities.

Fig.2.a. Visible Frameof the Video.

Fig.2.b. Thermal Frame of the Video.

Fig.2.a shows a video frame where a person wearing green clothes and green hat is camouflaged in the shrubs and Fig.2.b shows the thermal frame.

Fig.3.Output of the implementation where thermal modality is selected with simultaneous tracking with the particle filter.



The table below shows the observed parameters when thermal modality is selected. Selected Modality: Thermal**.**

Table I
EVALUATED PARAMETERS FOR CAMOUFLAGED CONDITION.

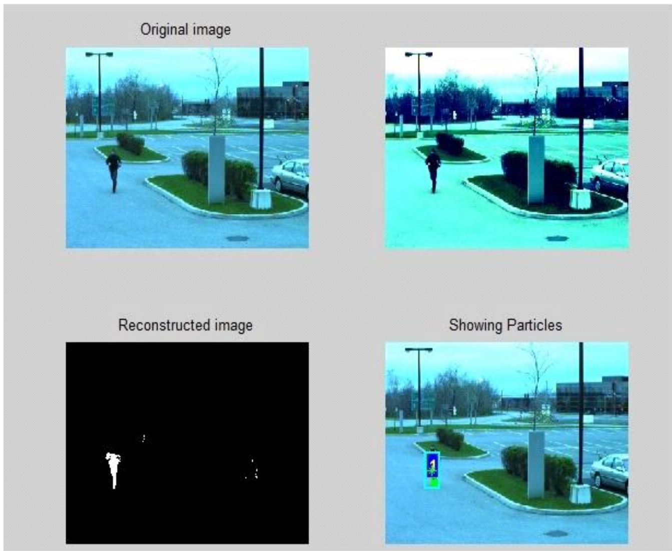
$B_{thermal}$	$B_{visible}$	α	B_{fused}
0.848087243	0.9814673	0.463548	0.46354
0.848002912	0.981009024	0.46364	0.46364
0.848062383	0.981072092	0.463641	0.46364
0.847802136	0.981445525	0.46347	0.46347
0.847605714	0.981186993	0.463478	0.46347
0.847465573	0.981312558	0.463405	0.4634

If $\boldsymbol{B_{visible}}$>$\boldsymbol{B_{thermal}}$: This indicates that the difference between the background and foreground frames for thermal modality is more hence the object can be distinctly traced. Hence thermal frames are selected.

Fig.4.a. Visible video frame

Fig.4.b.Thermalvideoframe**.**

Fig.5. Video sequence wherein the results obtained by visible camera are more discriminatory than the thermal camera, therefore selecting the former modality.



The table below shows the observed parameters when visible modality is selected.Selected Modality: Visible**.**

Table II

$B_{thermal}$	$B_{visible}$	α	B_{fused}
0.998931772	0.988218996	0.50269	0.99354
0.998977396	0.988333851	0.50267	0.99362
0.998967571	0.988281273	0.50268	0.99359
0.998998331	0.988060234	0.50275	0.99349
0.998857585	0.987741971	0.50282	0.99331

EVALUATED PARAMETERS WHEN A VIDEO IN WHICH A TARGET RUNS ACROSS THE TREE.

If $B_{thermal}$>$B_{visible}$:This indicates that the difference between the background and foreground frames for visible modality is more hence the object can be distinctly traced. Hence visible frames are selected. The algorithm can also track multiple targets and switch between the visible and thermal modality efficiently.

Fig.6.a. Visible video frame

.

Fig.6.b. Thermalvideo frame.

Fig.7 Video frame with multiple targets wherein the results obtained by visible camera are more discriminatory than the thermal camera.

V. CONCLUSION

- We have implemented an algorithm which weighs the two modalities for each video frame.
- The algorithm selects the modality which best discriminates the foreground from background.
- The process is repeated for every new frame in both the modalities thus making it an adaptive process.
- Future work will involve the investigation of a number of different research ideas, from low-level analysis of the best choice of threshold for multimodal change-detection to higher level tasks such as the detection and tracking of people in multimodal image sequences. Current work has shown some promising results in these directions.
- General principles for video data fusion will assist future work in utilizing data from other modalities such as stereo-vision depth, ultra-sound or other imagers of electromagnetic radiation.

REFERENCES

[1]. W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics*, 34(3):334–350, August 2004.

[2]. A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[3]. A. Ilie, R. Raskar, and J. Yu. Gradient domain context enhancement for fixed cameras. In *Asian Conference in Computer Vision*, January 2004.

[4]. H. Torresan, B. Turgeon, C. Ibarra, P. H´ebert, and X. Maldague. Advanced surveillance system: Combining video and thermal imagery for pedestrian detection. In *Proc. of SPIE, Thermosense XXVI*, volume 5405 of *SPIE*, pages 506–515, April 2004. 24(1), Jan/Feb 2004.

[5]. A. M. McIvor. Background subtraction techniques. In *Image and Vision Computing, Hamilton, New Zealand*, Nov 2000.

[6]. A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of Calcutta Mathematical Society, 1943.

[7]. A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. ACM Computing Surveys, 38: no 4, 2006.

[8]. Y. Yang and G. Sundaramoorthi. Modeling shape, appearance and self-occlusions for articulated object tracking. CoRR, abs/1208.4391, 2012.

[9]. V.VilaplanaandF.Marques. Region-basedmeanshifttracking: Application to face tracking. In ICIP, 2008.

[10]. K. Nummiaro, E. Koller-Meier, and L. V. Gool. A color based particle filter. First International Workshop on Generative-Model-Based Vision, pages 53–60, 2002.

[11]. A. Nakhmani and A. Tannenbaum. Particle filtering with region-based matching for tracking of partially occluded and scaled targets. SIAM Journal on Imaging Sciences, 4(1):220– 242, 2011.

[12]. T. Kailath, The Divergence and Bhattacharyya Distance Measures in Signal Selection, IEEE Transactions on Communication Technology, pp. 52-60, Vol. COM-15(1), 1967.

[13]. G. Kitagawa, Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, Journal of Computational and Graphical Statistics, pp. 1-25, Vol. 5(1), 1996. 12. D. Koller, J. Weber and J. Malik, Robust Multiple Car Tracking with Occlusion Reasoning, ECCV, pp. 189-196, 1994.