

TARGET-TRACKING IN FLIR IMAGERY USING MEAN-SHIFT AND GLOBAL MOTION COMPENSATION

Alper Yilmaz¹ Khurram Shafique¹ Niels Lobo¹ Xin Li² Teresa Olson³ Mubarak A. Shah¹

¹Unv. of Central Florida, Computer Science Dept. Orlando FL-32816

²Unv. of Central Florida, Mathematics Dept. Orlando FL-32816

³Lockheed Martin Corp. Orlando FL-32819

ABSTRACT

In this paper, we present a new approach for tracking targets in forward-looking infrared (FLIR) imagery taken from an airborne, moving platform. Our tracking approach uses the target intensity and the Gabor response distributions and computes a likelihood measure between the candidate and the model distributions by evaluating the Mean Shift Vector. When the Mean Shift Vector based tracker fails to locate the target due to large ego motion, we compensate the ego motion using a multi-resolution scheme, which employs the Gabor responses of two consecutive frames, and assumes a pseudo-perspective motion model. We present the experiments performed on an AMCOM FLIR dataset of the proposed tracking algorithm.

1. INTRODUCTION

Tracking moving or stationary targets in closing sequences of FLIR imagery is a challenging subject due to both the low contrast of the target with the background and the high ego (global) motion. In low resolution imagery the lack of the texture and the shape information of the targets make tracking even harder.

Most methods for tracking targets in FLIR imagery, use a common assumption of compensated global motion. Even under compensated global motion, the tracking results of these methods are not convincing.

To compensate global motion, Strehl and Aggarwal [1] have used a multi-resolution scheme based on the affine motion model. The affine model has its limitations and for FLIR imagery obtained from an airborne sensor, it is unable to capture the skew, pan and tilt of the planar scene. Moreover, their target tracker uses the assumption that the targets are brighter than the background and it tracks targets based on the expected mean and central moments of the targets, which are not stable measures for FLIR targets.

Shekarforoush and Chellappa [4] also use an ego motion stabilization approach. Once very hot or very cold targets are detected, the stabilization and tracking is based solely

on the goodness of the detection and the number of targets, i.e. if the number of targets is not adequate, or there is significant background texture, the system may fail to locate the target, therefore stabilization fails to correctly stabilize the image.

Braga-Neto and Goutsias [6] have presented a method based on morphological operators for target detection and tracking in FLIR imagery. Their tracker is based on the assumptions that the targets do not vary in size, they are either very hot or very cold spots, and there is small ego-motion. However, these assumptions contradict general closing sequence FLIR imagery.

Davies et al. [3] proposed a multiple target tracker system based on Kalman filters for FLIR imagery. The method assumes constant acceleration of the target, which is not valid for maneuvering targets. In addition, the method works only for sequences with no global motion.

In this paper, we present a new approach for real-time tracking of the FLIR targets in presence of high global motion. Compared to previous methods, the proposed approach does not use any constraints on the brightness of the target. Moreover, it is not required to have a target that has constant speed or acceleration. Our tracking algorithm is composed of two major modules. The first module is based on minimizing the distance between the statistical distributions of the target and the target candidate. The statistical distributions are obtained from Gabor filter responses and the intensity of the frames. We used Gabor filter response of the images since these 2-D quadrature phasor filters are conjointly optimal in providing the maximum possible resolution for information about the orientation and spatial frequency content of local image structure [8]. If the tracker module fails to locate the new target position due to high ego motion, the second module compensates the global motion. The global motion estimation module uses a multi-resolution scheme of [7] assuming a planar scene and using perspective projection for image formation. It uses Gabor filter responses of two consecutive frames to obtain the pseudo perspective parameters. The global motion compensation module is executed when the mean-shift tracker fails to track the target.

The organization of the paper is as follows. In the next section, we will give details of the target-tracking module. In section 3, the global motion estimation module is discussed. Experiments are given in section 4, and in the last section, future directions are outlined.

2. TARGET TRACKING

FLIR detector arrays generally produce spatially undersampled images, which results in poor, noisy data for further processing, such as target detection, tracking and recognition. Besides the limited capabilities of FLIR sensors, the images obtained using them often contain artifacts such as dirt on the sensor and fading brightness along with limited viewing resulting from cloudy weather. Moreover, FLIR sequences obtained via an airborne, moving platform, suffer from abrupt discontinuities in motion.

Targets in FLIR sequences can appear as either cold or hot spots due to thermal electro-magnetic spectrum. However, there is no *a priori* information about the target's temperature or shape. Due to target's low contrast with the background, our system uses both intensity images and images filtered by 2D Gabor filter kernels, which are oriented sine-wave gratings that are spatially attenuated by a Gaussian window. Two-dimensional Gabor filters, which have also been used for object detection in FLIR images [9], have the form

$$G_i(x, y) = e^{-\pi[x^2/\alpha^2 + y^2/\beta^2]} \cdot e^{-2\pi i[u_0x + v_0y]} \quad (1)$$

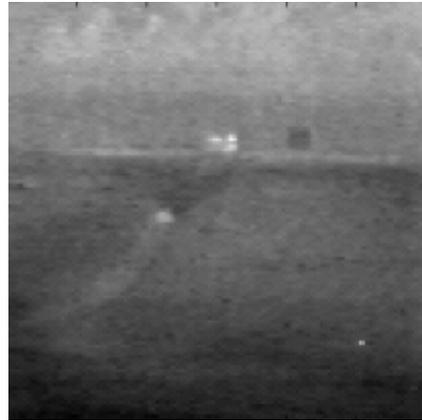
where α and β specify effective width and height, u_0 and v_0 specify modulation of the filter. We used 4 directions, 0, 45, 90, 135 and a fixed scale, 0.5 for our experiments. The response in four directions are summed for obtaining the input image to the tracker or global motion compensation module. In Figure 1 (a), a frame from one of the FLIR sequences and in (b) sum of Gabor responses are shown. In Figure 1b, the target regions are clearly emphasized.

Detection of targets in the FLIR sequences is a hard problem because of the variability of the appearance of targets due to atmospheric conditions, background, and thermodynamic state of the targets. The challenges in this regard are extremely low SNR, non-repeatability of the target signature, competing clutter, which forms similar shapes as those of the actual targets, obscured targets and lack of a priori information. In this paper, we assume that the location of target in the initial frame is given.

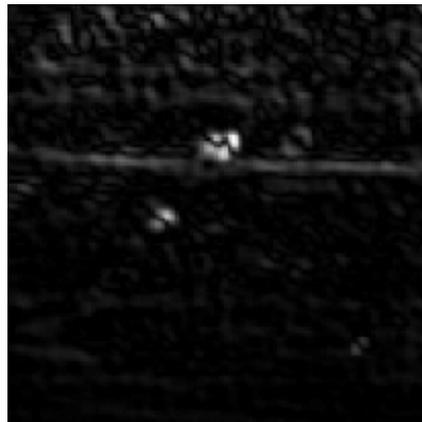
In the next section, we will give details of our model for tracking the target in two consecutive frames, which will be generalized to track the target in the whole sequence.

2.1. Tracking Model

Given the center of the target in a frame, a feature probability distribution of the target, such as weighted intensity,



(a)



(b)

Fig. 1. (a) Sample frame from one of the FLIR sequences, (b) summation of four Gabor responses of the frame in (a).

can be calculated in a circular patch using a 2-dimensional kernel centered on the target center [2]. The kernel provides the weight for a particular intensity according to its distance from the center of the kernel.

Let m be the center of a 2-dimensional kernel which is given by $K(m)$. The kernel density estimate for $K(m)$ is given by

$$f_K(m) = \frac{1}{nh^2} \sum_{i=1}^n K(x_i - \mathbf{m}) \quad (2)$$

where n is the number of points in 2-dimensional kernel $K(m)$, and h is the radius of the kernel. Possible kernels include Uniform kernel, Gaussian kernel, Triangular kernel, Bi-weight kernel and Epanechnikov kernel. Among these kernels, in the continuous case the Epanechnikov kernel yields the best minimum mean integrated square error between two kernel densities [5]. Approximating the discrete case as the continuous case, we used the Epanechnikov kernel to calculate the kernel density estimate of the data that

lies within the kernel. Two-dimensional Epanechnikov kernel is given by

$$K_E(x) = \frac{2}{\pi h^2} (h^2 - \|x\|^2) \quad (3)$$

where $\|\cdot\|$ is the magnitude operator, and h is the radius of the kernel.

Let $I : R^2 \rightarrow N$ and $G : R^2 \rightarrow N$ be the functions, which give intensity value and the Gabor filtered value at a particular spatial location respectively. Kernel density estimates for the features are given by

$$p(u|K_E) = \frac{\sum_{i=1}^n \delta(\Gamma(\mathbf{x}_i) - u) K_E(\mathbf{m} - \mathbf{x}_i)}{\sum_{i=1}^n K_E(\mathbf{m} - \mathbf{x}_i)} \quad (4)$$

where Γ is the feature function, which can be either I or G as described above and δ is Kronecker delta function.

2.2. Algorithm

Given the initial center of the target for the first frame, \mathbf{m}_0 , target model probability distributions q_I and q_G are evaluated using equation (4) for both intensity and Gabor filtered images respectively. For the next frame, center of the target is initialized at its previous location and the candidate probability distributions $p_I(\mathbf{m}_0)$ and $p_G(\mathbf{m}_0)$ are calculated calculated for intensity and Gabor filtered images respectively. The distance between model and candidate distributions is defined by

$$d(\mathbf{m}) = \sqrt{1 - \rho(\mathbf{m})} \quad (5)$$

where $\rho(\mathbf{m})$ is the Bhattacharya coefficient, which gives the likelihood of the model and the candidate distributions evaluated around target center \mathbf{m} , Bhattacharya coefficient is given by

$$\rho(\mathbf{m}) = \frac{1}{2} \sum_{u=1}^m (\sqrt{p_{I_u}(\mathbf{m})q_{I_u}} + \sqrt{p_{G_u}(\mathbf{m})q_{G_u}}) \quad (6)$$

where m is the number of bins for both the intensity and the Gabor response histograms. Minimizing the distance between two distributions is equivalent to maximizing the Bhattacharya coefficient. Displacement of the target center is calculated by the weighted mean of the kernel where the weights are obtained by the Taylor series expansion of the Bhattacharya coefficient around \mathbf{m}_0 ,

$$\rho(\mathbf{m}) = \rho(\mathbf{m}_0) + \frac{1}{4C} \sum_{i=1}^n \left(\sum_{u=1}^m \left(\delta(I(\mathbf{x}_i) - u) \sqrt{\frac{q_{I_u}}{p_{I_u}(\mathbf{m}_0)}} + \delta(G(\mathbf{x}_i) - u) \sqrt{\frac{q_{G_u}}{p_{G_u}(\mathbf{m}_0)}} \right) K_E(\mathbf{m} - \mathbf{x}_i) \right) \quad (7)$$

where C is the normalization constant. Since the translation of the center of the target is limited to the radius of the target kernel, we can discard the other terms in the Taylor series expansion. The inner sum in equation (7),

$$w_i = \sum_{u=1}^m \left(\delta(I(\mathbf{x}_i) - u) s_{qI} t \frac{q_{I_u}}{p_{I_u}(\mathbf{m}_0)} + \delta(G(\mathbf{x}_i) - u) s_{qG} t \frac{q_{G_u}}{p_{G_u}(\mathbf{m}_0)} \right) \quad (8)$$

is strictly positive and is selected as weights associated with each pixel in the kernel. Using the weights obtained from equation (8), new location of the target center is given by

$$y_1 = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (9)$$

The new location of the target is determined using the weighted mean equation, which converges in few iterations compared to correlation based approaches and due to simplicity of the calculations, it is faster than the correlation.

Once the target distribution is obtained from the first frame, we update the distribution if the function given in equation (5) is lower than a fixed threshold. We selected this threshold to be 0.35 for our experiments.

Tracking based on the weighted mean approach requires that the new target center lies within the kernel centered on the previous location of the target. In FLIR imagery, target motion in the captured frame is due to either ego (sensor) motion or object motion. However, the discontinuities in motion stems from the large ego motion, which can be compensated by a global motion compensation module. If the distance between the candidate and the model distributions, given in equation (5), does not converge, we execute the global motion compensation module, which will be discussed in detail in the next section.

3. GLOBAL MOTION COMPENSATION

We assume in our target-tracking module that the motion is small. If this condition is violated, then the output of the tracker becomes unreliable, and requires global motion compensation.

There are several approaches presented in literature for the ego motion compensation in visual images, however they may not be directly applicable to FLIR images because of the nature of imagery. One of the most significant characteristics of FLIR scenes is the lack of textural information in the scene. In addition, the scene is usually composed of patches of uniform intensities with small contrast between the patches. These characteristics of FLIR imagery makes it difficult to estimate the sensor motion from images. In order to get around the problem of limited texture, we have used Gabor filter responses of the images in the multi-resolution

framework of [10]. The proposed method uses pseudo perspective model of motion, which is able to provide better estimate of motion, where affine model fails.

The motion of the sensor between two image frames under pseudo perspective model is expressed by

$$\mathbf{U} = \mathbf{M}\mathbf{a} \quad (10)$$

where

$$\mathbf{U} = \begin{pmatrix} u & v \end{pmatrix}^T$$

$$\mathbf{M} = \begin{pmatrix} 1 & x & y & xy & x^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & y^2 & xy & 1 & x & y \end{pmatrix}$$

$$\mathbf{a} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \end{pmatrix}^T$$

where (u, v) is the optical flow. Optical flow can be calculated using the optical flow constraint equation given by

$$\mathbf{F}_\mathbf{x}^T \mathbf{U} = -f_t \quad (11)$$

where $\mathbf{F}_\mathbf{x} = \begin{pmatrix} f_x & f_y \end{pmatrix}$ is a spatial gradient vector and f_t is the temporal derivative of Gabor filter response of two consecutive frames. Combining equations (10) and (11) results in a linear system that can be solved using the least square method,

$$\left(\sum \mathbf{M}^T \mathbf{F}_\mathbf{x} \mathbf{F}_\mathbf{x}^T \mathbf{M} \right) \mathbf{a} = - \sum f_t \mathbf{M}^T \mathbf{F}_\mathbf{x} \quad (12)$$

In Figure 2, (a) reference frame (previous frame), (b) current frame, (c) registered frame and (d) the difference image between the current frame and the registered frame are shown for the Gabor filter responses.

Once projection parameters, \mathbf{a} , are calculated by solving system of equation (12), we apply the transformation to the target center of the previous and calculate the approximate new target center, by

$$\mathbf{m}_k = \mathbf{m}_{k-1} + \mathbf{U} \quad (13)$$

After compensation of the global motion by updating the target center using equation (13), we evaluate mean shift vector on this updated location. Since estimation of the global motion is done only when the tracker fails, i.e. the reference frame is always the previous frame; global motion module does not suffer from accumulation of projection errors.

4. EXPERIMENTS

We have applied the proposed method to AMCOM FLIR dataset for tracking targets. The data set was made available to us in grayscale format and was composed of 41 sequences where each frame was 128x128.

The proposed method was developed using C++ running on Windows 2000 on a Pentium III platform and current implementation of the algorithm is capable of tracking one target at a time at 15 frames per second. Currently the initializing of the target center is performed manually.

The tracking results of the algorithm were visually confirmed. Most of the time, the second module for compensation of global motion was not used. In Figure 3, we give the diagram for the distance, given in equation (5), between 2 distributions corresponding to consecutive frames for 5 selected sequences from the AMCOM dataset. The diagram is obtained without executing the global motion compensation module. Frame numbers, where the distance measure is above 0.5, show that the global motion compensation is required. In the diagram, the regions, where the distance is 1.0, show that the tracker is failing to locate the target.

Figure 4 shows a set of selected frames from a sequence where there is high but smooth global motion, i.e. the global motion compensation module is not executed. In the initial frame, the target center and the kernel, which encapsulates the target, is manually marked and the white cross in the other frames show the tracked target center. The target tracker correctly tracked the target.

Similarly, in Figure 5, the output of the target tracker for tracking a cold target is shown. The target in the sequence has very low contrast with the background and the neighboring locations hide the target due to high gradient magnitude. The sequence given in Figure 5 shows the importance of using both the intensity and the Gabor filter responses together, since the target tracker fails to locate the target if only one of the features were used.

Though the tracking results are reasonable in most of the cases, large global motion makes the tracking become unstable. In this case, the second module refines the tracking results. In figure 6, we give an example where the tracker failed to locate the target, but compensated global motion helped the tracker to locate the target correctly. In Figure 6, (a) shows the tracking results without global motion compensation and (b) shows tracking result after compensating the global motion.

To justify the increase in performance of the system, which uses both intensity and gabor filter responses as features, we performed experiments on 18 sequences of the AMCOM dataset. In the experiments, proposed mean-shift based tracker used intensity or gabor responses alone, and we compared the tracking results with the combined feature. The comparison is given in Table 1. In the table, each row represents the tracking result of 18 sequences. We displayed the results in 3 categories: robust, promising and failed, where "robust" is used for correct target tracking, "promising" stands for acceptable performance and "failed" represents the failure of the target tracker.

5. CONCLUSIONS AND FUTURE DIRECTIONS

We have proposed an algorithm for tracking targets in FLIR images. Our system tracks the target by finding the translation of the target center using the intensity and Gabor response distributions. When the distance between two distributions is not small, we compensate the global motion from the previous frame to current frame assuming pseudo-perspective motion model, and transform the target center accordingly.

Currently we are able to track one target, and we are refining the system to track multiple targets.

Acknowledgements

The authors wish to thank to Richard Sims of U.S. Army AMCOM for providing us FLIR sequences. This research is funded by Lockheed Corporation.

6. REFERENCES

- [1] Strehl, A.; Aggarwal, J.K., "Detecting moving objects in airborne forward looking infra-red sequences," *Machine Vision Applications*, Vol. 11, 2000, pp. 267-276.
- [2] Yizong Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on PAMI*, Vol. 17, 1995, pp. 790-799.
- [3] Davies, D.; Palmer, P.; Mirmehdi, "Detection and Tracking of Very Small Low Contrast Objects," *Proceedings of the 9th British Machine Vision Conference*, Sept. 1998.
- [4] Shekarforoush, H.; Chellappa, R., "A multi-fractal formalism for stabilization, object detection and tracking in flir sequences," *International Conference on Image Processing*, Vol. 3, 2000.
- [5] Comaniciu, D.; Ramesh, V.; Meer, P., "Real-time tracking of non-rigid objects using mean shift," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2000, pp. 142-149.
- [6] Braga-Neto, U.; Goutsias, J., "Automatic Target Detection and Tracking in Forward-Looking Infrared Image Sequences Using Morphological connected Operators," *33rd Conference of Information Sciences and Systems*, March 1999.
- [7] Bergen, J. R., Anandan, P., Hanna, K. J., and Hingorani, R., "Hierarchical model-based motion estimation," *Proceedings European Conference on Computer Vision*, Berlin, Germany, Vol. 588, Springer, May 1992, pp. 237-252.
- [8] Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, Vol. 2, 1985, pp. 1160-1169.
- [9] Braithwaite, R.N.; Bhanu, B., "Hierarchical Gabor filters for object detection in infrared images," *Proceedings of IEEE Conf. on Computer Vision Pattern Recognition*, 1994, pp. 628- 631.
- [10] Irani, M.; Anandan, P., "Video indexing based on mosaic representations," *Proceedings of the IEEE*, Vol. 86 Issue: 5, May 1998, pp. 905-921.

Sequence	Intensity Only	Gabor Response Only	Combined
rng14.15	promising	robust	robust
rng15.20	failed	robust	robust
rng15.NS	failed	robust	robust
rng16.04	robust	failed	robust
rng16.07	failed	failed	promising
rng16.08	robust	promising	robust
rng16.18	failed	robust	robust
rng17.01	promising	failed	robust
rng17.20	failed	failed	robust
rng18.03	failed	promising	promising
rng18.05	failed	promising	robust
rng18.07	failed	robust	robust
rng18.16	failed	failed	promising
rng18.18	failed	promising	promising
rng19.01	promising	promising	robust
rng19.06	promising	promising	robust
rng19.07	robust	promising	robust
rng19.11	promising	robust	robust

Table 1. Performance evaluation for intensity only, gabor response only and intensity and gabor combined tracking of the proposed system for 18 sequences from the AMCOM dataset.

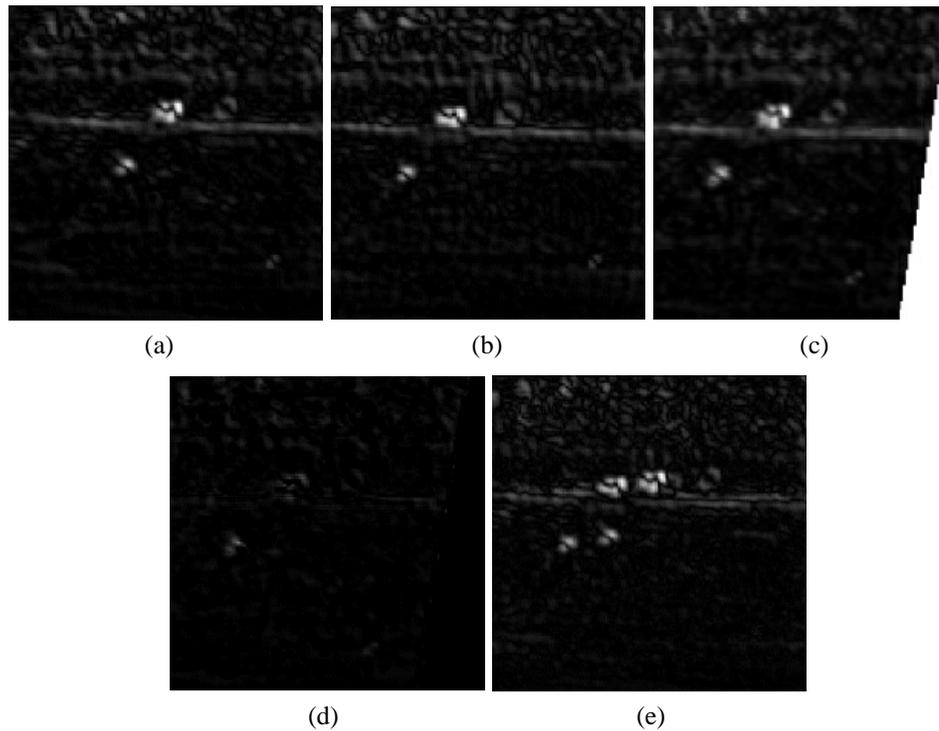


Fig. 2. (a) The reference frame, (b) current frame, (c) first frame registered onto the second frame, i.e. global motion is compensated, using the pseudo-perspective projection, (d) the difference image obtained from (b) and (c), (e) the difference image obtained from (a) and (b).

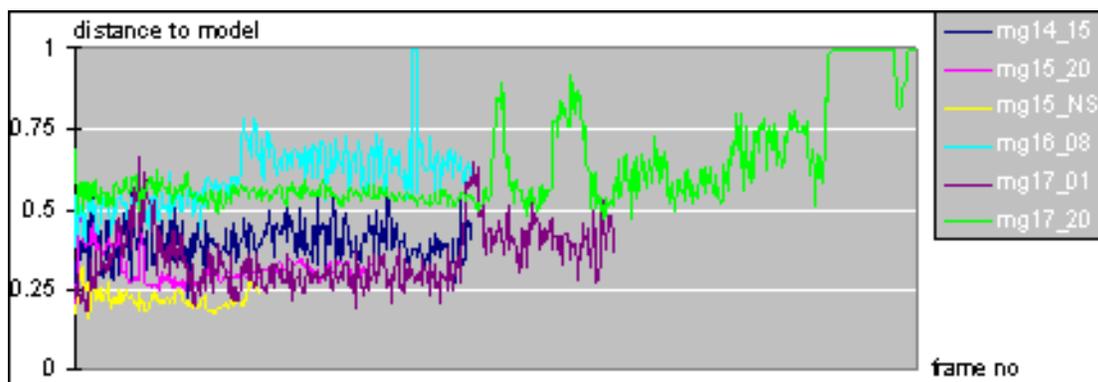


Fig. 3. Distance measure of equation (5) evaluated for consecutive frames for six different sequences from the dataset.

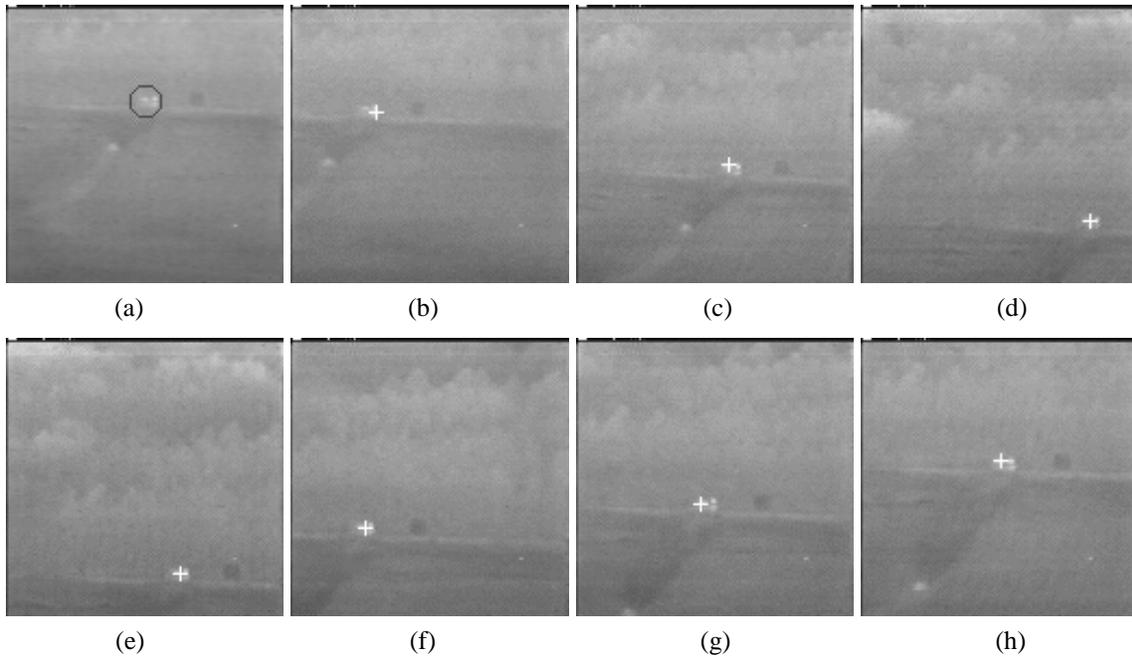


Fig. 4. Target tracking results for a sequence where there is a big global motion; (a) manually initialized target for frame 0, automatically tracked target shown by '+' in (b) frame 8, (c) frame 19, (d) frame 35, (e) frame 51, (f) frame 68, (g) frame 78 and (h) frame 89.

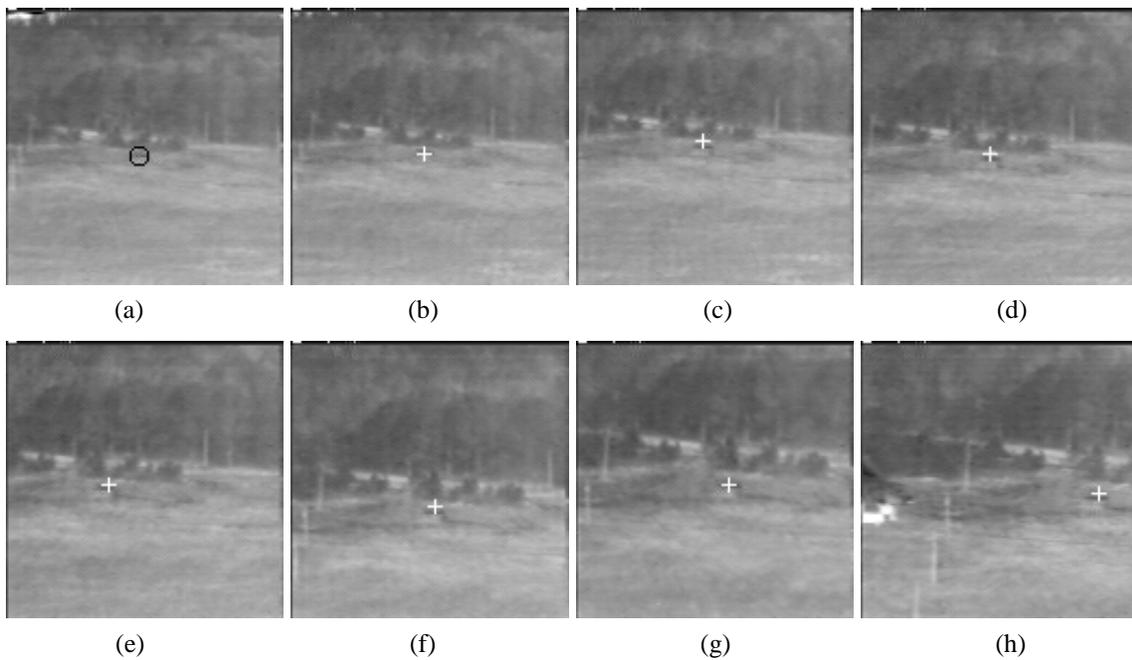
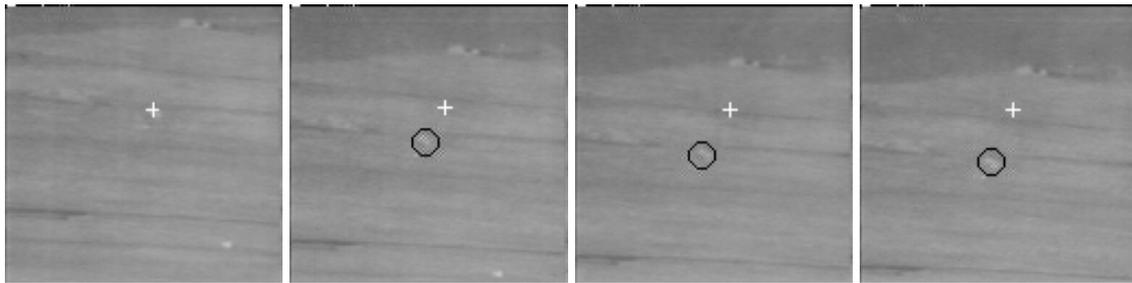
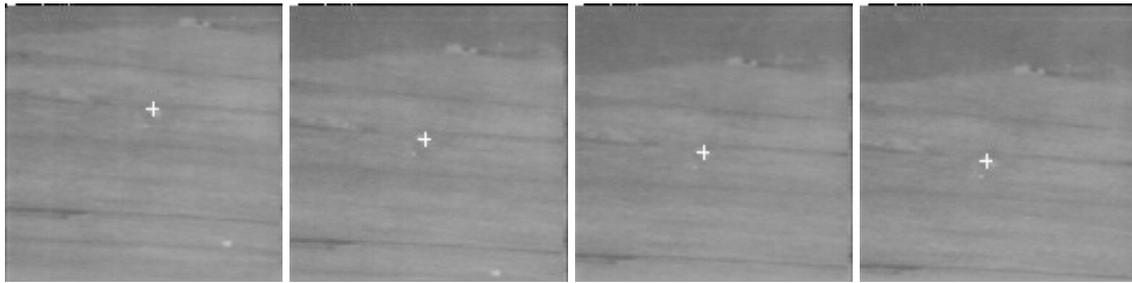


Fig. 5. Target tracking results for a cold target; (a) manually initialized target for frame 0, automatically tracked target shown by '+' in (b) frame 11, (c) frame 39, (d) frame 63, (e) frame 91, (f) frame 119, (g) frame 154 and (h) frame 170.



(a)



(b)

Fig. 6. (a) Target tracking results for a fragment of a sequence where the tracker fails due to discontinuity in motion, white cross shows the output of the tracker and the black circles show the correct location of the target. (b) Refined tracking results for the same fragment, where the global motion compensation module is executed only once for compensating the global motion for the first and second frame.