Fringe-adjusted JTC Based Target Detection and Tracking Using Subframes from a Video Sequence

M. S. Alam, M. Huq, J. Khan, H. Kettani, A. A. S. Awwal and K. M. Iftekharuddin
Department of Electrical and Computer Engineering
University of South Alabama, Mobile, AL 36688

Abstract: The joint transform correlator (JTC) technique has been found to be suitable for real-time matching and tracking operations. Among the various JTC techniques proposed in the literature, the fringe-adjusted JTC has been found to yield better correlation output for target detection. In this paper, we propose a generalized fringe-adjusted JTC based algorithm for detecting and tracking a target in a video image sequence. The proposed JTC based algorithm has been found to be suitable for near real time recognition and tracking of a static or moving target. The performance of the proposed technique has been verified with real life image sequences.

Key Words: target detection, joint transform correlation, fringe-adjusted filter, joint power spectrum, image sequence.

1. Introduction

The joint transform correlator (JTC) [1] technique has shown remarkable promise for real-time optical pattern recognition and tracking applications [2-3]. The main advantages of the JTC are that it does not require any complex filter fabrication, allows real-time update of the reference image, permits simultaneous Fourier transform of the reference image and the unknown input scene, and avoids the precise positioning otherwise required in matched filter based correlator [4].

A new class of JTCs which employs Fourier plane joint power spectrum (JPS) apodization based on the reference image has been found to yield better correlation output. Among the various apodization based techniques, the recently proposed fringe-adjusted filter (FAF) based JTC appears to be particularly attractive since it avoids the problems otherwise associated with the alternate JTC techniques [1-3, 5, 7-14]. The fringe-adjusted JTC [6] employs a real-valued filter called FAF at the Fourier plane and is found to yield significantly better correlation performance when compared to the existing JTCs. In the fringe-adjusted JTC [6], the JPS is multiplied by FAF before applying the inverse Fourier transform to yield the correlation output. The fringe-adjusted JTC yields better correlation performance than alternate JTCs for the case of noise-free single target [6, 7] and multi-target binary input scenes under normal as well as poor illumination conditions [15-17]. However, for noise corrupted input scenes, whenever the reference power spectrum contains very low values ($|R(u,v)|^2 \cong 0$), a fringe-adjusted JTC has been found to yield low correlation peak intensity. To overcome this limitation, recently a fractional fringe-adjusted JTC [18] technique has been proposed which employs a family of real-valued filters, called generalized fringe-adjusted filters (GFAF).

In this paper, we discuss the development and implementation of an algorithm, based on GFAF, capable of detecting and tracking a target in gray-level video sequence acquired by an infrared camera mounted on a moving platform. The performance of the proposed techniques has been verified with real life infrared video sequences.

Proceedings of SPIE Vol. 5201 Photonic Devices and Algorithms for Computing V, Edited by Khan M. Iftekharuddin, Abdul Ahad S. Awwal (SPIE, Bellingham, WA, 2003) · 0277-786X/03/$15.00

2. Generalized JTC Based Algorithm

In a JTC, the unknown input scene and the known reference image are displayed side-by-side in the input plane, illuminated by a coherent, collimated laser beam. The input joint image plane is located at a distance of one focal length from a lens, which Fourier transforms the input joint image at the Fourier plane. The Fourier transformed patterns constructively interfere with each other to create an interference pattern called joint power spectrum. This interference pattern is recorded in real time using a square law device such as a CCD camera. The joint power spectrum is inverse Fourier transformed by another Fourier transform lens to yield the correlation output. For a match, two correlation peaks or bright spots are produced and for a mismatch, negligible or no correlation peaks are produced.

The block diagram for the fringe-adjusted JTC algorithm is shown in Fig. 1, where the reference and the unknown input image are introduced in the input plane using a spatial light modulator (SLM) such as a liquid crystal television illuminated by a laser light source. If $r(x,y+y_0)$ represents the reference image and $t(x,y-y_0)$ represents the input scene containing n objects $t_1(x - x_1, y - y_1)$, $t_2(x - x_2, y - y_2)$, ..., $t_n(x - x_n, y - y_n)$ respectively, and if the input scene is also assumed to be corrupted by noise $n(x,y-y_0)$, the input joint image becomes

$$f(x,y) = r(x,y+y_0) + \sum_{i=1}^{n} t_i(x-x_i, y-y_i-y_0) + n(x,y-y_0) \tag{1}$$

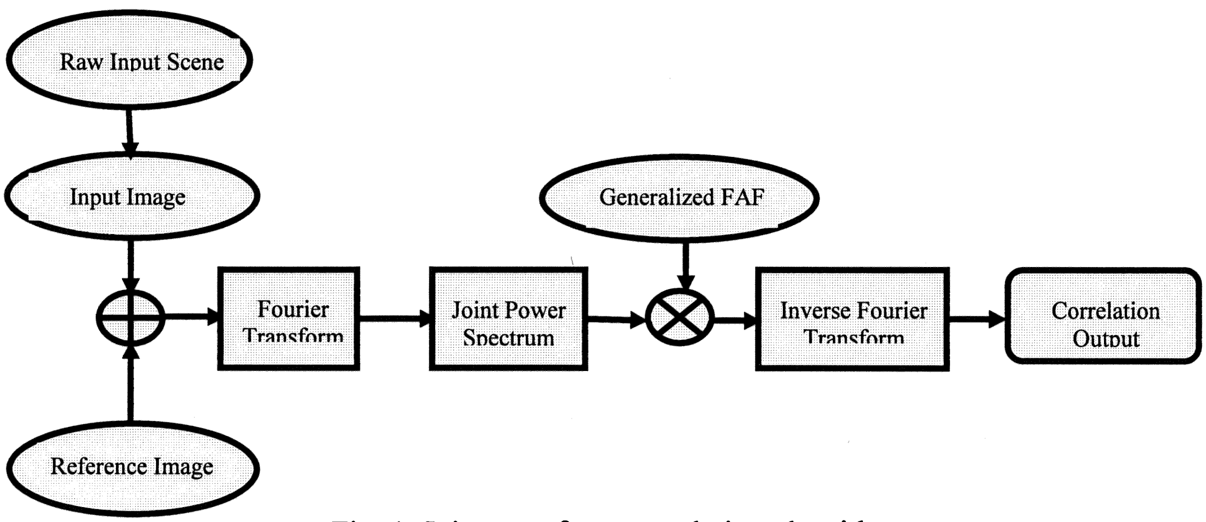


Fig. 1. Joint transform correlation algorithm.

Applying Fourier transform to the input joint image of Eq. (1) produces the JPS, given by

$$\begin{aligned}|F(u,v)|^2 &= |R(u,v)|^2 + \sum_{i=1}^{n}|T_i(u,v)|^2 + |N(u,v)|^2 + 2\sum_{i=1}^{n}|T_i(u,v)||R(u,v)|\cos[\phi_{ti}(u,v) - \phi_r(u,v) - ux_i - vy_i - 2vy_0] \\ &+ 2|R(u,v)||N(u,v)|\cos[\phi_r(u,v) - \phi_n(u,v) + 2vy_0)] + 2\sum_{i=1}^{n}|T_i(u,v)||N(u,v)|\cos[\phi_{ti}(u,v) - \phi_n(u,v) - ux_i - vy_i] \\ &+ 2\sum_{i=1}^{n-1}\sum_{k=i+1}^{n}|T_i(u,v)||T_k(u,v)|\cos[\phi_{ti}(u,v) - \phi_{tk}(u,v) - ux_i + ux_k - vy_i + vy_k]\end{aligned} \tag{2}$$

where $|R(u,v)|$, $|T_i(u,v)|$ and $|N(u,v)|$ are the amplitudes, and $\phi_r u,v)$, $\phi_{ti}(u,v)$ and $\phi_n(u,v)$ are the phases of the Fourier transforms of $r(x,y)$, $t_i(x,y)$ and $n(x,y)$ respectively; u and v are mutually independent frequency-domain variables scaled by a factor $2\pi / \lambda f$, λ is the wavelength of the collimating light and *f* is the focal length of the Fourier transforming lenses.

The fringe-adjusted JTC has been found to yield better correlation performance for the case of noise-free single/multiple target detection under varying illumination conditions of the input scene [7]. For noise-free input scenes, the fringe-adjusted JTC has been found to yield significantly better correlation output when compared to the alternate JTCs. However, for noisy input scenes, it may accentuate the noise component of the input scene which may degrade the system performance. To overcome the aforementioned problems, and at the same time, to utilize the fringe-adjusted filter as a versatile tool for various practical applications, we propose a generalized FAF, defined as

$$H(u,v) = C(u,v)\left[D(u,v) + |R(u,v)|^m\right]^{-1} \tag{3}$$

where m is a constant, and $C(u,v)$ and $D(u,v)$ are either constants or functions of u and v. When $C(u,v)$ is properly selected, one can avoid having an optical gain greater than unity. With very small values of $D(u,v)$, the pole problem otherwise associated with an inverse filter is overcome. When $m = 0$, $D(u,v) = 0$ and $C(u,v) = 1$, the GFAF-based JTC corresponds to the classical JTC; when $m = 2$, $D(u,v) << |R(u,v)|^2$ and $C(u,v) = 1$, the GFAF-based JTC corresponds to the fringe-adjusted JTC or inverse filter based JTC; and when $m = 1$, $D(u,v) << |R(u,v)|$ and $C(u,v) = 1$, the GFAF-based JTC corresponds to the phase-only JTC. The aforementioned three cases correspond to the complex matched filtering ($m = 0$), inverse filtering ($m = 2$) and phase-only filtering ($m = 1$) [13, 19]. Thus, all important types of matched filter based correlators can be implemented in real time using the proposed fractional power fringe-adjusted JTC while avoiding the limitations of matched filter based correlators. For noise-free input scenes, the fringe-adjusted JTC yields the highest correlation output. However, for noise corrupted input scenes, the fractional power phase-only JTC yields better correlation output than the fringe-adjusted JTC. If $C(u,v) = 1$ and $|R(u,v)|^m >> D(u,v)$, then Eq. (3) may be approximated as $H(u,v) \approx |R(u,v)|^{-m}$. The fringe-adjusted JPS is obtained by multiplying the JPS of Eq. (2) by the filter function $H(u,v)$ and is thus given by

$$G(u,v) = H(u,v) \times |F(u,v)|^2 \approx |R(u,v)|^{-m} \times |F(u,v)|^2 \tag{4}$$

In Eq. (2), the first three terms correspond to the zero-order peak, the fourth term corresponds to the desired crosscorrelation between the reference image and the input scene objects; the fifth and sixth terms correspond to the crosscorrelation between the reference image (input scene) and the noise term and, the last term corresponds to the crosscorrelation between the different input scene objects. If multiple identical targets are present in the input scene, then in addition to the desired autocorrelation peaks between the reference and the targets, other correlation peaks will be produced due to the autocorrelation between the targets themselves [corresponding to the last term of Eq. (2)], which yield false correlation peaks (called false alarms) in the output plane.

Most recently, we have proposed a Fourier plane image subtraction technique which can be achieved by recording the input-scene-only power spectrum ($|T(u,v)|^2$) and the reference-only power spectrum ($|R(u,v)|^2$) separately. Since the reference image is known, $|R(u,v)|^2$ may be precalculated and stored in the computer. Both the input-scene-only power spectrum and the reference-image-only power spectrum are then subtracted from the JPS of Eq. (2). Thus, the modified JPS obtained using the Fourier plane image subtraction technique may be expressed as

$$P(u,v) = |F(u,v)|^2 - |T(u,v)|^2 - |R(u,v)|^2 = 2\sum_{i=1}^{n} |T_i(u,v)||R(u,v)| \cos[\phi_{ti}(u,v) - \phi_r(u,v) - ux_i - vy_i - 2vy_0] + 2|R(u,v)||N(u,v)| \cos[\phi_r(u,v) - \phi_n(u,v) - 2vy_0] \quad (5)$$

The subtraction operation eliminates the zero order term, the false alarms generated by the similar input scene targets as well as the crosscorrelation terms between other objects that may be present in the input scene. Notice that the subtraction operation does not require any additional hardware. By exploiting the above-mentioned Fourier plane image subtraction technique [20], the fractional power fringe-adjusted JTC can significantly reduce the effect of noise and clutter which may be present in the input scene, and eliminate the crosscorrelation terms generated by the input scene targets. The modified JPS of Eq. (5) is then multiplied by the GFAF of Eq. (4) to yield the modified fringe-adjusted JPS, given by

$$O(u,v) = H(u,v) \times P(u,v) \approx |R(u,v)|^{-m} \times P(u,v) \quad (6)$$

An inverse Fourier transform of Eq. (6) yields the correlation output that will consist of pairs of delta functions corresponding to the input scene targets and negligible crosscorrelation output between the reference and non-target objects as well as the noise term.

3. Motion Estimation Using GJTC

In addition to performing raw pattern matching, a JTC can be particularly useful for tracking images in a real environment where targets move in challenging scenarios corrupted by noise, clutter and other artifacts. By correlating successive frames, for instance, the location of one or more targets can be found between two consecutive frames.

In this paper, a JTC based algorithm has been developed for distortion invariant target recognition and tracking using multimodal criterion. For example, the nonlinear JTC is capable of tracking a target by efficiently estimating image motion as shown in Fig. 2. Our algorithm consists of two modes: detection and tracking. In the detection mode, target position is estimated in the input scene and is stored as an XY coordinate vector. When the target is identified, the GJTC algorithm switches into the tracking mode.

The proposed tracking algorithm utilizes the GJTC technique for estimating object motion as shown in Fig. 2. In Fig. 2, the image at time, $t = t_1$, corresponds to a frame from a sequence of frames containing the target; and the image at time, $t = t_2$, is the next frame of that sequence. In this technique, the shift of the target between two consecutive frames of an image sequence is measured to estimate the motion of a given target. If both the reference image and the input image are same, the correlation output will contain two symmetric correlation peaks, which indicates the relative location of the target in the input image. The image motion is measured by performing GJTC based 2D correlation of the sequential images and additional post-processing of the correlation output. After post-processing, the difference between the position vector of the current frame and the previous frame is obtained to estimate the position of the target.

The shift of the correlation peaks relative to the reference axes corresponds to the shift between the input image and reference image. High redundancy of the correlation procedure permits to obtain pixel accuracy for the determination of the position of the peaks and the shift value. Thus the GJTC algorithm allows efficient real time tracking of moving targets.

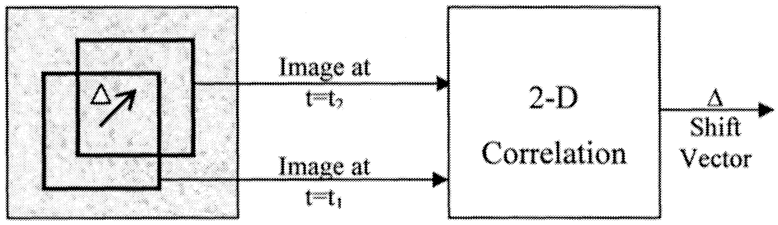


Fig. 2. Image motion determination with 2-D correlation

The proposed GJTC based algorithm for target tracking involves the following four steps:

- Image preprocessing, which involves procedures including formatting and correcting of the data to facilitate better visual interpretation.
- Image segmentation, which is useful for identifying regions of interest in a scene or annotating the data. By segmentation the search for the target is reduced to a suitable region of the image by using some *a priori* knowledge on the scene. The segmentation helps to achieve consistent tracking results for scenes with arbitrary object motion.
- Recognition stage, which tests each target by cross-correlation technique, if validated, is classified according to the database of reference images.
- Tracking stage, which keeps track of the moving target of interest in successive frames.

The proposed approach can be very helpful for the automatic moving target tracking in real world. The steps of the algorithm are discussed in the following subsections.

3.1 Preprocessing

Preprocessing operation is mainly intended for correcting the distortions of image data. Corrections may be necessary due to variations in scene illumination and viewing geometry, atmospheric conditions, and sensor noise and response. Noise in an image may be due to irregularities or errors that occur in the sensor response and/or data recording and transmission. Each of these will vary depending on the specific sensor and platform used to acquire the data and the conditions during data acquisition. All of these effects should be corrected before detection and tracking is performed. Also, it may be desirable to convert and/or calibrate the data to known (absolute) radiation or reflectance units to facilitate comparison between data. Preprocessing functions involve those operations that are normally required prior to the main data analysis and extraction of information. The images are preprocessed to simplify subsequent operations of target detection and tracking without losing relevant information. The preprocessing steps are shown in Fig. 3. After taking each frame, intensity normalization is carried out to all pixels of that frame to adjust the average intensity level. After that the mean value of that frame is subtracted from all pixel values of the image to remove the background. Thus we improve the appearance of the imagery to assist in better visual interpretation and analysis. By this processing we have intensified the target in the image relative to the pixel values of background and the useful features of the target in the image are enhanced.

3.2 Segmentation

Image and video segmentation is useful in target detection and tracking applications for identifying regions of interest in a scene or annotating the data. The objective of segmentation technique is to partition a given image into meaningful regions of components. The definition of "meaningful region" is a function of the problem being considered. For example, in our

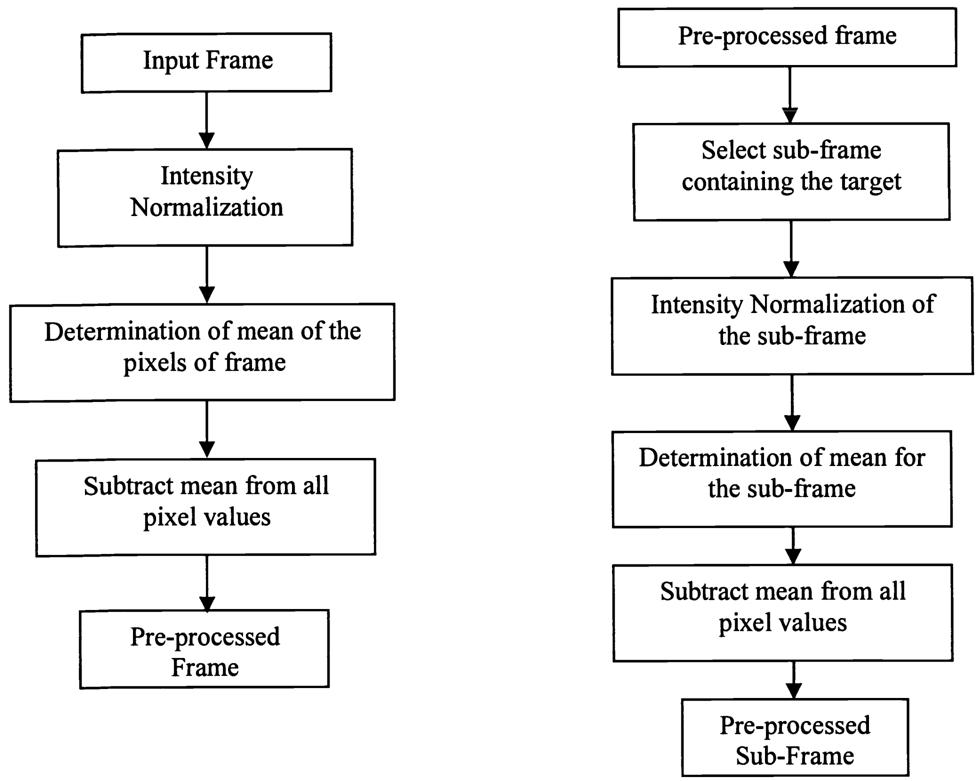
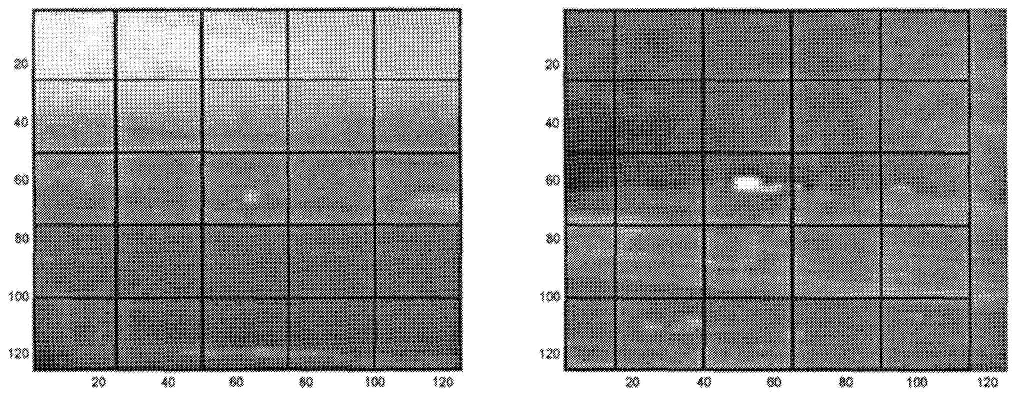


Fig. 3 (a) Preprocessing of an image frame, (b) Preprocessing of a sub-frame.

algorithm the purpose of segmentation is to identify regions corresponding to targets in the scene. In particular, in our algorithm, we used a segmentation operation to isolate the target of interest from other objects in the input scene and from the hostile background by using some *a priori* knowledge about the location of target on the scene. The main focus of this work is partitioning the image to obtain spatial segmentation of that image into sub-images and to pick one sub-image, which contains the target. The size of the window of sub-image determines the size of image regions that will be used to detect the target. In our implementation, the sub-image is a 25x25 window. The overall approach of the image-segmentation for each frame in a sequence is similar. The segmentation operation for two different sequences having different locations of target is shown in Fig. 4. Preprocessing steps of Fig. 3 are again applied to the sub-frame, before taking it as our reference image as well as input image, to further enhance relative features of target.



(a) (b)

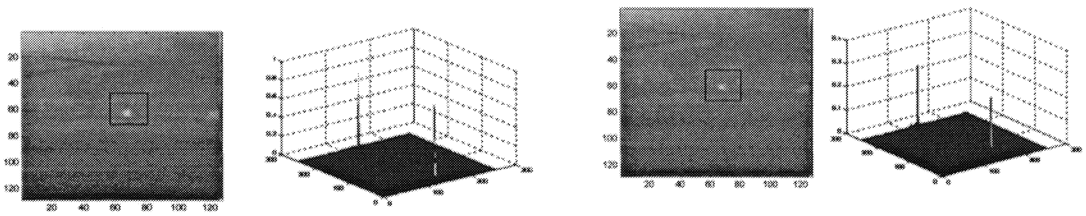
Fig. 4 Segmentation and sub-image of 25X25 from the first frame of 128X128 pixels has been selected (a) at X=51:75, Y=51:75 for image sequence 1, and (b) at X=51:75, Y=41:65 for image sequence 2.

3.3 Recognition stage

The fringe-adjusted JTC technique has been used for moving target recognition and tracking. This algorithm introduces adaptive correlation with a sequentially updated reference image and input image. For the first frame, we selected the sub-frame in such a way that the target of interest would be approximately at the center of the sub-image. This sub-frame is used as the reference image and as the input scene in the input joint image. For all subsequent frames, the sub-frame for the input scene is taken from the current frame whereas that for reference image is selected from preceding frame. In this implementation, adaptive updating of reference image and input image is used to ensure that the current frame accurately represents the new input image of the target. The Fourier plane image subtraction technique eliminates the unwanted zero-order term and cross-correlation terms produced by similar input scene objects while alleviating the detrimental effects of blur and noise that may be present in the unknown input scene. Thus the final correlation output is a pair of sharp correlation peaks as shown in Fig. 5.

3.4 Tracking stage

By correlating successive frames, for instance, the distance between two otherwise identical images can be found. From the three-dimensional correlation output matrix between reference image and input image, the X-Y coordinates of the correlation peak offers the position of the maximum correlation coefficient in the matrix, which is the location of the target in that frame. The difference between X-Y coordinates of correlation peaks for two consecutive frames indicates the measurement of the movement of target from one frame to the next frame. Thus by knowing the location of correlation peak, a moving target can be tracked using the pixel wise difference between consecutive image frames as shown in Fig. 5. The resulting system identifies targets of interest, rejects background clutter, and continually tracks a target while alleviating the effects of occlusion, appearance changes and cessation of target motion.



(a) (b)

Fig. 5. Target recognition and tracking for sequence L14_15. X-Y co-ordinate of correlation peak (a) at frame #51 is (17, 129) and (b) at frame #52 is (15, 130). The difference between these two co-ordinates provides the direction and number of pixels of target movement.

4. Experimental Results

To evaluate the performance of the developed GJTC algorithm, we used the image sequences supplied by AMCOM. The correlation output obtained from the GJTC algorithm is used for tracking the desired object. For a match, the correlation peak in the matrix gives us the approximate location of the target in the specific frame.

To investigate the performance of the proposed algorithm for tracking, we selected an arbitrary challenging sequence (sequence 1) supplied by AMCOM, which contains 398 frames. The subframe (25x25 pixels) is taken in such a way that it is centered on the target. This sub-frame is used as the reference image for Frame #1. By using the subframe as the reference image, the correlation output between the reference and the unknown input scene is obtained using the GJTC algorithm. From this correlation output the X-Y coordinates (X=17 and Y=129) of the correlation peak is determined.

For Frame #2, the entire frame is subdivided into an array of 25x25-pixel subframes. Using the reference image from Frame #1 and each of the subframes from Frame #2 as the input scene, correlation outputs are obtained using the GJTC algorithm. By examining the entire set of the correlation outputs, the location of the target is determined. For Frame #2, the X-Y coordinates of the correlation peak of the target has been found to be located at X=17 and Y=129. The difference between the X-Y coordinates of the correlation peak for Frame #1 and the X-Y coordinates of the correlation peak for Frame #2 indicates the relative displacement of the target from Frame #1 to frame #2. The difference between two subsequent X-Y coordinates are denoted by Δx and Δy, respectively. In this case, $\Delta x = 0$ and $\Delta y = 0$ indicating that the target did not move from Frame #1 to Frame #2. Now to select the sub-frame from Frame #3, the sub-frame grid is shifted by Δx and Δy along the X and Y coordinates, and the aforementioned procedure is repeated. Notice that the sub-frame corresponding to the location of the target in Frame #2 becomes the reference image for detecting the target from Frame #3. Following the aforementioned procedure, the reference image and the input image are adaptively updated to detect and track the target for subsequent frames. Satisfactory tracking results are obtained for all frames in our test sequence and some examples of tracking are shown in Fig 6.

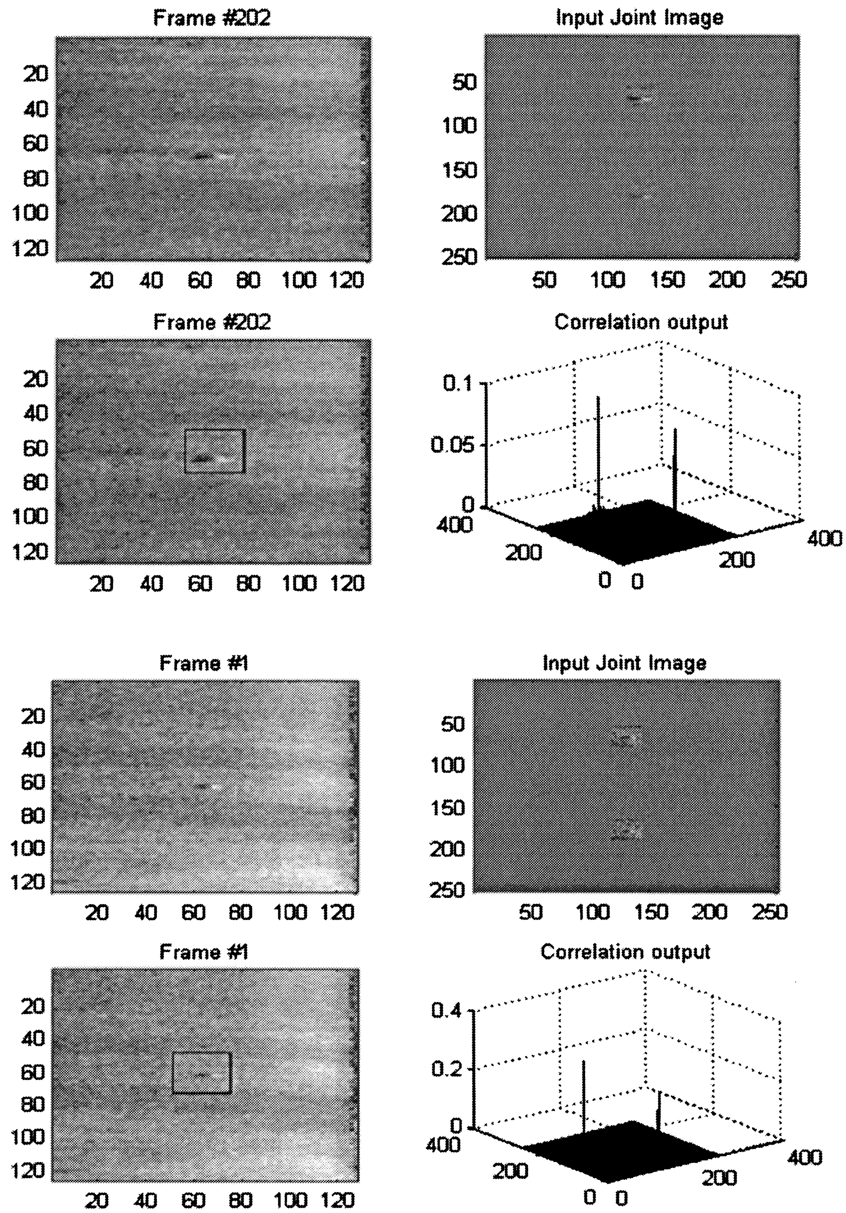
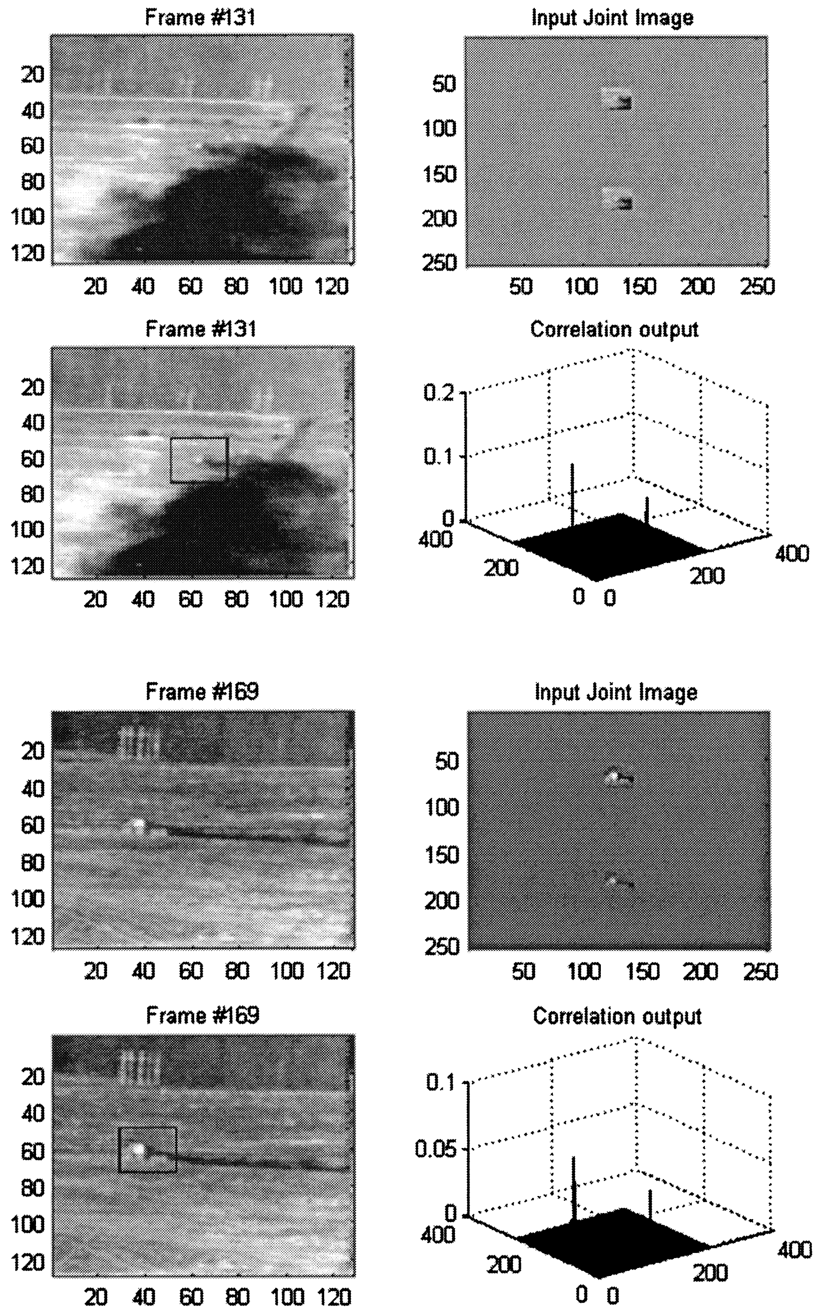


Fig. 6. Tracking results of object by using JTC

We also tested our algorithm using another arbitrary sequence (Sequence 2) supplied by AMCOM. For this sequence, the desired target appears in Frame #131 and it is a moving target. This sequence contains 195 frames and the corresponding tracking results obtained using the GJTC algorithm is shown in Fig. 7.

Fig. 7. Tracking results for another object by using JTC



5. Conclusion

In this paper, we presented a GJTC algorithm for near real-time detection and tracking of moving targets in terrestrial scenes. We consider the problem of extracting the features of moving targets in the presence of clutter and other artifacts. Experimental results show that the proposed algorithm is a powerful tool for detection and tracking of moving targets.

6. References

1. Weaver and J. W. Goodman, "A technique for optically convolving two functions," *Appl. Opt.* **5**, p. 1246-1249 (1966).
2. F. T. S. Yu and X. J. Lu, "A real-time programmable joint transform correlator," *Opt. Commun.* **52**, p. 10-16 (1984).
3. F. T. S. Yu, S. Jutamulia, T. W. Lin, and D. A. Gregory, "Adaptive real-time pattern recognition using a liquid crystal TV based joint transform correlator," *Appl. Opt.* **26**, p. 1370-1372 (1988).
4. A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory* IT-10, pp. 139-146, 1964.
5. B. Javidi and C. Kuo, "Joint transform image correlation using a binary spatial light modulator at the Fourier plane," *Appl. Opt.* **27**, p. 663-665 (1988).
6. M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlator," *Appl. Opt.* **32**, p. 4344-4350 (1993).
7. M. S. Alam and M. A. Karim, "Joint-transform correlator under varying illumination," *Appl. Opt.* **32**, p. 4344-4350 (1993).
8. M. S. Alam and M. A. Karim, "Improved correlation discrimination in a multiobject bipolar joint transform correlator," *Opt. & Laser Tech.* **24**, p. 45-50 (1992).
9. Q. Tang and B. Javidi, "Multiple-object detection with a chirp-encoded joint transform correlator," *Appl. Opt.* **32**, p. 4344-4350 (1993).
10. F. T. S. Yu, F. Cheng, T. Nagata and D. A. Gregory, "Effects of fringe binarization of multi-object joint transform correlation," *Appl. Opt.*, **28**, p. 2988-2990 (1989).
11. X. J. Lu, A. Katz and E. G. Kanterakis, N. P. Caviris, "Joint transform correlator that uses wavelet transforms," Opt. Lett. **17**, p. 1700-1702 (1992).
12. D. Feng, H. Zhao and S. Xia, "Amplitude-Modulated JTC for Improving Correlation Discrimination," Opt. Commun., **86**, p. 260-264 (1991).
13. A. A. S. Awwal, M. A. Karim and S. R. Jahan, "Improved Correlation Discrimination Using an Amplitude Modulated Phase-only Filter," *Appl. Opt.* **29**, p. 233-236 (1990).
14. J. Khoury G. Asimellies and C. L. Woods, "Incoherent-erasure nonlinear joint transform correlator," *Opt. Lett.* **20**, p. 2321-2323 (1995).
15. M. S. Alam and J. Khoury, "Fringe-adjusted Incoherent Erasure Joint Transform Correlator," *Journal of Optical Engineering*, **37** p. 75-82, (1998).
16. M. S. Alam and M. A. Karim, "Multiple target detection using a modified fringe-adjusted joint transform correlator," *Opt. Eng.* **33,** p. 1610-1617 (1994).
17. F. Cheng, P. Andreas and F. T. S. Yu, "Removal of Intraclass associations in joint transform power spectrum," *Opt. Commun.*, **99**, p. 7-12 (1993).
18. M. S. Alam, "Fractional power fringe-adjusted joint transform correlator," Op*t. Eng.* **34**, p. 3208-3216, (1995).
19. K. M. Iftekharuddin, M. A. Karim, P. W. Eloe and A. A. S. Awwal, "Discretized amplitude-modulated phase-only filter," *Opt. & Laser Tech.,* **28**(1), p. 93-100, (1996).
20. Chun-Te Li, Shizhuo Yin and Francis T. S. Yu, "Nonzero-order joint transform correlator," *Opt. Eng.* **37**, p. 58-65, (1998).
21. Alan J. Lipton, H. Fujiyoshi and Raju S. Patil, "Moving target classification and tracking from real-time video," *Proc. IEEE Workshop Application of Computer Vision,* p. 8-14, 1998.

22. J. Hwang, Y. Ooi and S. Ozawa, “Visual feedback control system for tracking and zooming a target,” *Proc. of the International Conference on Power Electronics and Motion Control*, IEEE, 1992.

Acknowledgement

This research was supported by a grant from the Army Research Office (ARO: 43004-CI).