Military Technical College
Kobry El-Kobbah,
Cairo, Egypt

ICEENG

8th International Conference
on Electrical Engineering
ICEENG 2012

Advanced image tracking techniques for target seeking

By

Mohamed Mamdouh*   Hussein Saad*   Gouda Ismail*   Tarek Ahmed*

Abstract:

The primary goal of visual tracking is to provide a computerized system resembling the capabilities of the human eye. Motion detection is used in many applications and its integration as a facility within a vision system offers many potential benefits. We introduced an investigation of target tracking techniques; two different target tracking algorithms were implemented. The first is Covariance-based target tracking technique and the second is Hu Moments-based target tracking technique. The implemented tracking techniques were investigated and analyzed to determine which algorithm is suitable. Both algorithms were enhanced by achieving segmentation process using the region growing technique and building an adaptive window size that surrounds the target to overcome the problem of scaling. The experimental results show that the enhancement was very efficient not only in target tracking but also in overcoming the occlusion problem.

Keywords:

Target tracking, Covariance matrix, Hu moments.

1. Introduction:

Tracking can be defined as the problem of estimating the trajectory of an object as it moves around a scene. Visual tracking is the vision technique that locates and follows a target within a dynamic image obtained by a visual sensor. Object tracking, in general, is a challenging problem. Difficulties in tracking objects can arise due to [1]:

---

* Egyptian Armed Forces

Abrupt object motion, changing appearance patterns of the object and the scene, complex object shapes, partial and full object occlusions, scene illumination changes, and real time processing requirements. In this paper, each tracked target in each image sequence is assigned a maximum processing time and sampling time period (temporal constraint). The maximum processing time is the maximum time spent on a particular image (frame) to locate the sampled target. The sampling time period is the maximum time interval between two consecutive samples of a particular target. However, the problem is how to process the region of interest (ROI) of each sampled target within its time constraint to achieve real time. The work aims to build a complete image tracking system that is capable of tracking a target through complex scenes. We propose 2 different algorithms for target tracking. The first algorithm is based on Covariance tracker. The second algorithm is based on the 7 Hu Invariant-Moments. Both techniques were enhanced by applying a segmentation algorithm based on region growing to build an adaptive window around the target of interest to overcome the problem of scaling. The paper is organized as follows: Section 2 presents some of the background theory and implementation techniques necessary for building a robust target tracking system. Taxonomy is introduced to group tracking approaches into three categories. Section 3 illustrates and discusses the functional block diagrams of the proposed Covariance-based and Hu moments-based target tracking techniques and their implementations with region growing segmentation method. Experimental results are given in Section 4. Conclusions are presented in Section 5.

2. Related works:

In this section, we introduce the main tracking techniques that can be classified into three categories: Template matching based object tracking, Contour based object tracking and Optical flow estimation (pixel level) tracking [2].

Template matching is a technique in digital image processing for finding small parts of an image which match a template image. It can be used as a way to detect and track moving target across frames and the region of interest in the first frame is outlined either manually or by a detector. Comaniciu et al. [3] proposed a feature histogram based target representation to represent non-rigid objects for visual tracking. Yilmaz et al. [4] extended the traditional mean shift tracking and presented an object tracking method based on the asymmetric kernel mean shift, in which the scale and orientation of the kernel adaptively change depending on the observations at each iteration. Lim et al. [5] proposed to learn the eigenbasis online during the object tracking process, and constantly update this representation as the appearance changes due to pose, view angle, and illumination variation. Tao et al. [6] introduced a general solution for estimating motion layers in extended image sequences.

<u>Contour tracking</u> goal is to find the object's silhouette across time given a sequence of images in which a known object of interest is in motion. Boundary tracking with deformable planar contours, known as snakes, was originally introduced by Terzopoulos and Szeliski [7]. MacCormick and Blake [8] modeled the objects as a set of parametric image curves. Chen et al. [9] used parametric shape model to represent object contours. The hidden states are defined to be the contour point located along each normal line. Cohen [10] introduced a balloons model, which starts from the external force balance equation directly, inflates snakes with an additional force and stops the evolution at strong edges.

<u>Optical flow estimation</u> task consists of recovering a 2D displacement vector establishing correspondences between the consecutive projections of a 3D patch in the image. The central assumption in such a context is the visual preservation constraint [11] that assumes pixels corresponding to the same 3D projection will have the same intensities. Optical flow algorithms can be classified into three categories: variational methods [12], Markov random field based approaches [13] and statistical techniques [14].

3. Proposed target tracking techniques:

<u>3.1 Covariance-based target tracking technique</u>

This technique operates in three phases, namely target selection, region of interest determination in the next frame, and finding the best match. The first phase is the target selection phase. In this phase, the target is selected manually by the center of the desired target and the window size that includes the desired target in the first frame is defined. Then, compute the covariance matrix of the features (f_k) as a model of the desired target for the selected window. These features may be associated directly to the pixel coordinates

$$f_k = [\,x \quad y \quad I(x,y) \quad I_x(x,y) \quad I_y(x,y) \quad I_{xx}(x,y) \quad I_{yy}(x,y)\,] \tag{1}$$

Where I is the observed image and I_x, I_{xx}, I_y, I_{yy} are the image gradients in the x and y directions respectively.

The second phase is the region of interest determination in the next frame. In this phase, simply, the ROI is evaluated by summing and subtracting the ROI factor ***k*** with upper-left corner of the desired target. Since the movement of the tracked target between two successive frames is unknown, then ROI should be determined such that the tracked target is covered in all directions. According to the tracked target speed (# pixel/msec), the ROI size is evaluated, so that the ROI increases as the speed of the tracked target increases and vice versa. For example, if the desired target moves with 0.1 pixel/msec in a video sequence of 25 fps, then the target will move approximately 4 pixels between

two successive frames, so the ROI factor **k** should be at least 4 pixels. Thus, Covariance matrix of a sliding window with the same target size will be evaluated in the ROI from its upper-left to lower-right corners for each increment step **d**. Consequently as **d** decreases, the no. of overlapped windows inside ROI increases and vice versa. We represent an M×N rectangular region R with a d×d covariance matrix c_R of the feature points as [15]:

$$c_R = \frac{1}{MN} \sum_{k=1}^{MN} (f_k - \mu_R)(f_k - \mu_R)^T \tag{2}$$

Where μ_R is the vector of the means of the corresponding features for the points within the region R.

The third phase is finding the best match phase. In this phase, for a given ROI in the current frame, we compute the covariance matrix of the features as a model of the object for the different windows. Then, compute the dissimilarity measure between the covariance measure of the first frame window and the covariance measure of the ROI windows using the Generalized Eigen values method as shown in equation (3):

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^{d} ln^2 \lambda_k(C_i, C_j)} \tag{3}$$

Where $\{\lambda_k(C_i, C_j)\}$ are the generalized eigenvalues of C_i and C_j, computed from

$$\lambda_k C_i x_k - C_j x_k = 0 \qquad k = 1 \ldots d \tag{4}$$

And x_k are the generalized eigenvectors. The distance measure ρ satisfies the metric axioms, positivity, symmetry, triangle inequality, for positive definite symmetric matrices. After that, we find the window that has the minimum covariance distance from the model and assign it as the estimated location.

There are several advantages of using covariance matrices as region descriptors. The covariance matrix proposes a natural way of fusing multiple features without normalizing features or using blending weights. It embodies the information embedded within the histograms as well as the information that can be derived from the appearance models. In general, a single covariance matrix extracted from a region is enough to match the region in different views and poses. The noise corrupting individual samples are largely filtered out with the average filter during covariance computation. Covariance matrix of any region has the same size, thus it enables comparing any regions without being restricted to a constant window size. Moreover, covariance matrix can be invariant to rotations. Nevertheless, if information regarding the orientation of the points is embedded within the feature vector, it is possible to detect rotational discrepancies. We also want to point that the covariance is invariant to the mean changes such as identical shifting of color values. This becomes an

advantageous property when objects are tracked under varying illumination conditions.

3.2 Hu moments-based target tracking technique

This technique operates in three phases, namely target selection, region of interest determination in the next frame, and finding the best match. The first phase is the target selection phase. In this phase, the target is selected manually by the center of the desired target and the window size that includes the desired target in the first frame. Then, calculate the seven Hu invariant moments as a model of the desired target for the selected window. Hu [16] introduced seven functions of second and third moments that are invariant to translation, scale and rotation. The seven Hu moment invariants $(\emptyset_1, \ldots, \emptyset_7)$ are given by:

$$\emptyset_1 = \eta_{20} + \eta_{02} \tag{5}$$

$$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{6}$$

$$\emptyset_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{7}$$

$$\emptyset_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{8}$$

$$\emptyset_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{9}$$

$$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})] \tag{10}$$

$$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$+ (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{11}$$

Where η_{pq} are the normalized central moments given by:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$$

Where μ_{pq} is a central moment of order $\gamma = \frac{p+q}{2} + 1$ and $p + q = 2, 3, \ldots$

Moment invariants represent a complete set of image descriptors, and are therefore of fundamental importance for pattern recognition. The second phase is the region of interest determination in the next frame. In this phase, simply, the ROI is evaluated by summing and subtracting the ROI factor **k** with center of the desired target. Since the movement of the tracked target between two successive frames is unknown, then ROI should be determined such that the tracked target is covered in all directions. According

to the tracked target speed (# pixel/msec), the ROI size is evaluated, so that the ROI increases as the speed of the tracked target increases and vice versa. Thus, the 7 Hu invariant moments of a sliding window with the same target size will be evaluated in the ROI from its upper-left to lower-right corners for each increment step *d*. Consequently, as *d* decreases, the no. of overlapped windows inside ROI increases and vice versa. The third phase is finding the best match phase. In this phase, for a given ROI in the current frame, we calculate the 7 Hu invariant moments as a model of the object for the different windows. Then, compute the dissimilarity measure between the Hu moments measure of the first frame window and the Hu moments measure of the ROI windows using the ***Euclidean distance*** method. Whereas the Euclidean distance (D) can be calculated as:

$$
\begin{aligned}
D(\mathbf{a},\mathbf{b}) &= \sqrt{(\mathbf{a}-\mathbf{b})^T(\mathbf{a}-\mathbf{b})} \\
&= \left(\sum_{i=1}^{d}(a_i - b_i)^2 \right)^{1/2} \\
&= \|\mathbf{a}-\mathbf{b}\|
\end{aligned}
\tag{12}
$$

Where **a** and **b** are two feature vectors of dimension d. After that, we find the window that has the minimum Hu moments distance from the model and assign it as the estimated location.

3.3 Covariance / Hu moments based target tracking techniques with segmentation

The general covariance / Hu moments tracking algorithms have some problem of changing the scale of moving objects. Since the scale of the moving object often changes in time, the tracking (or object) window should be updated accordingly. In addition, the covariance / Hu moments of moving object should be adaptively changed considering the tracking window size. The covariance / Hu moments based tracking techniques with segmentation proposed a solution to this problem by segmenting the moving object from the background pixels in the tracking window by applying the region growing method. This method will improve the tracking performance of the conventional tracking techniques. Our several experiments prove the effectiveness of the proposed covariance / Hu moments based tracking techniques with segmentation.

The functional block diagram of the covariance / Hu moments based tracking techniques with segmentation is depicted in figure 1. These techniques operate in four phases, namely target selection, region of interest determination in the next frame, finding the best match and segmentation. The first, second and third phases were previously discussed in sections 3.1 and 3.2. The fourth phase is segmentation phase. In this phase, we apply segmentation method (region growing algorithm) to segment the moving object from the background pixels in the tracking window. The seed point of the region growing algorithm is selected from the window that has the minimum

dissimilarity (estimated location in the previous phase). After growing, the upper left and bottom right corners of the object are determined. Finally, the new adaptive window is determined and drawn.
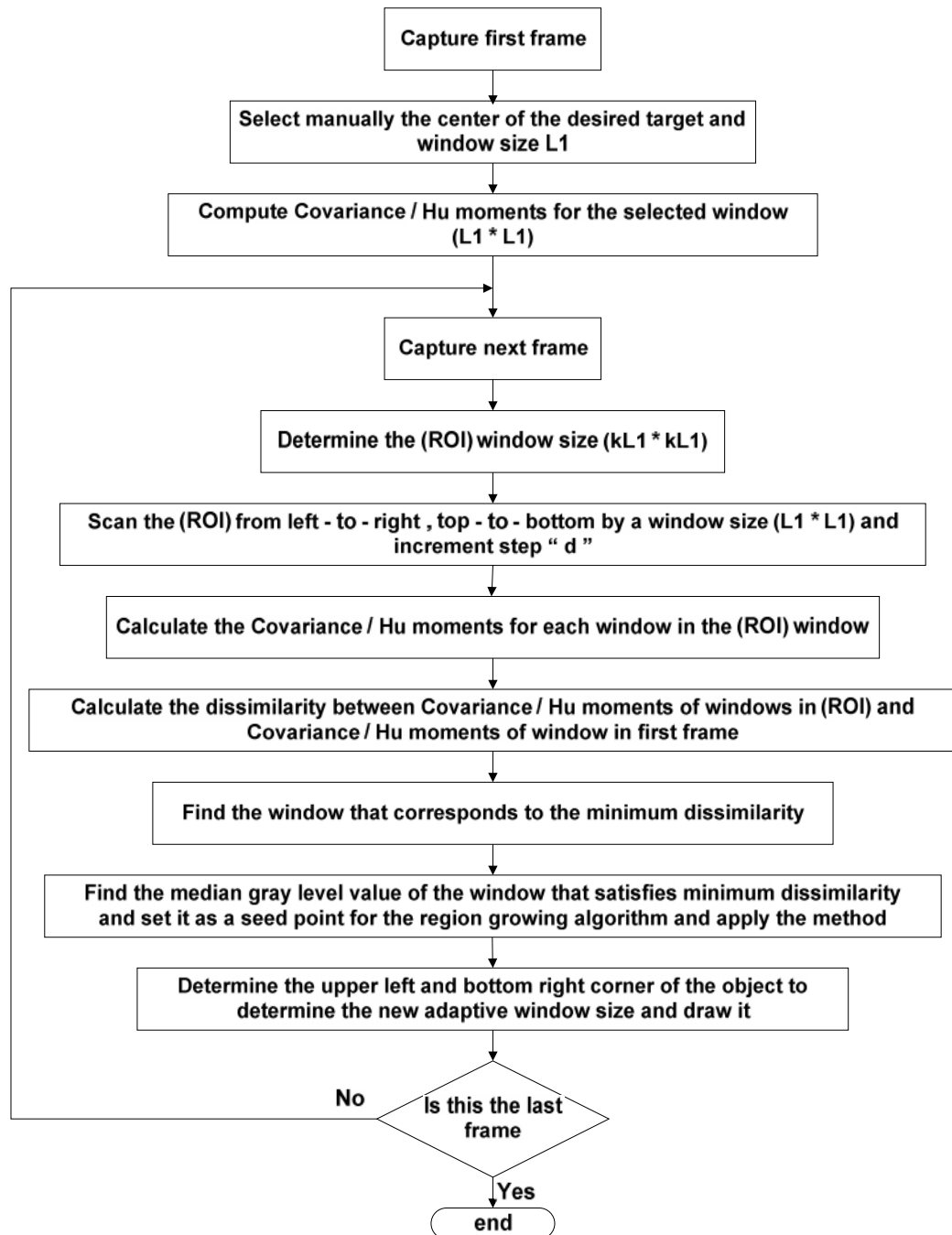

***Figure (1)** : Functional block diagram of the covariance / Hu Moments tracking techniques*
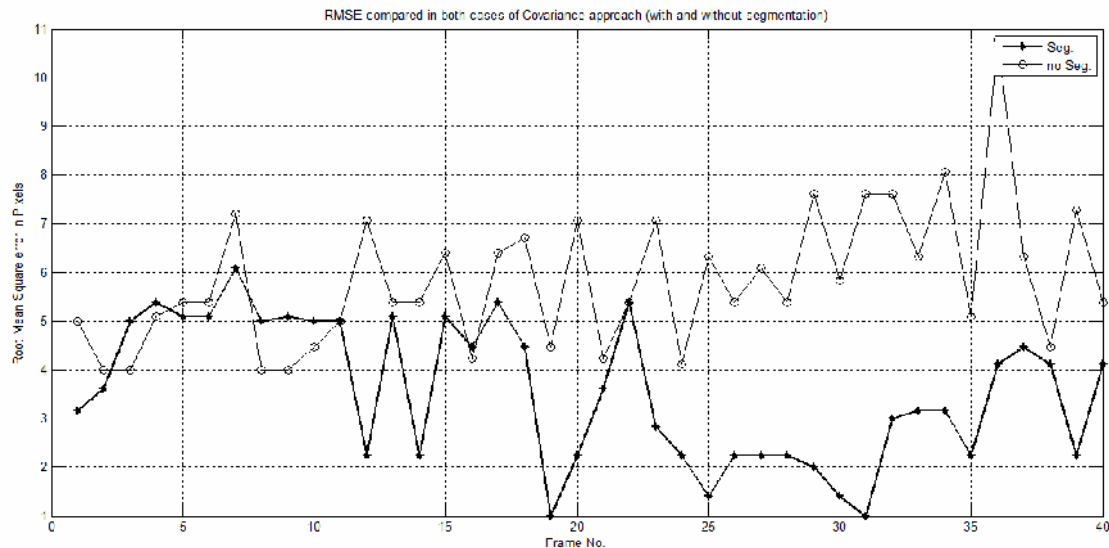
with Segmentation

4. Experimental results:

This section reports the results obtained by applying the implemented tracking techniques to several real monocular image sequences.

Experiment 1

Figure 2 illustrates the system tracked an ambulance car over a monocular sequence of 40 frames. The two techniques covariance and Hu moments with and without segmentation were applied on this sequence. The experimental results show that both tracking techniques with segmentation were much efficient and have less RMSE than traditional ones as shown in figures 3 and 4.


| (a) | (b) | (c) | (d) | (e) |
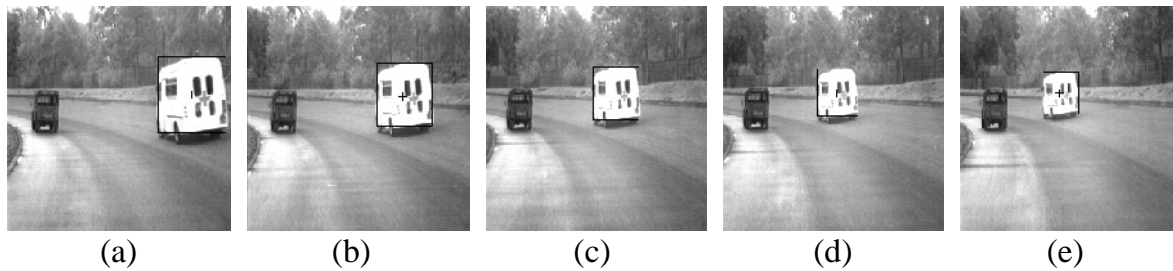
Figure (2): Selected images from "ambulance" image sequence after applying the Covariance tracking technique with segmentation. Starting from upper left to lower right:(a) First frame in sequence, (b) frame no.10, (c) frame no.20, (d) frame no.30, (e) frame no.40 last frame in sequence. The black (+) in the center of each black rectangle indicates the target detected by the tracking technique.

Figure (3): RMSE between actual target location and estimated location using Covariance tracking technique with segmentation (*) and without segmentation (o)
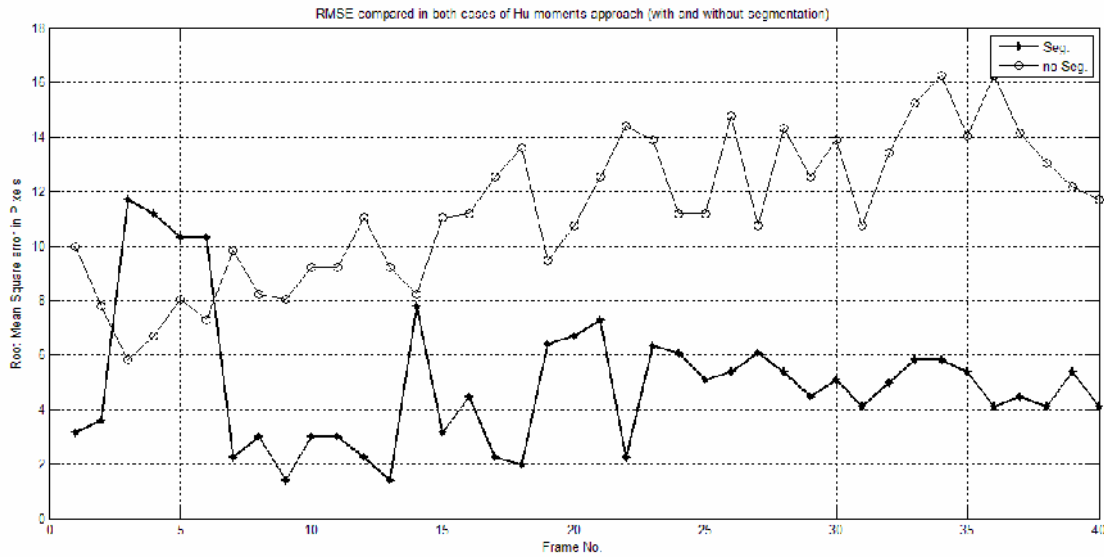
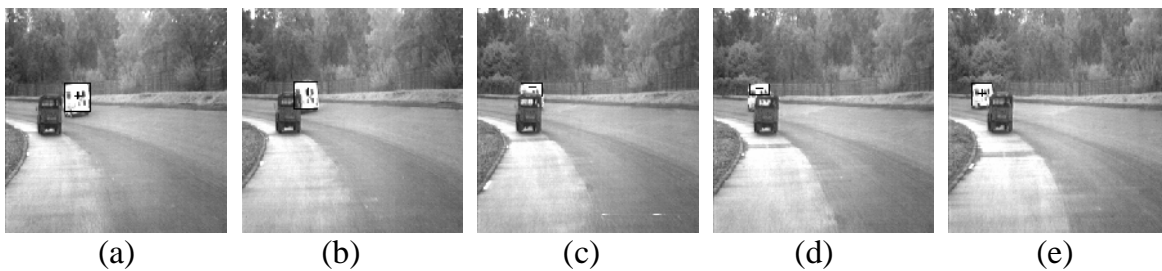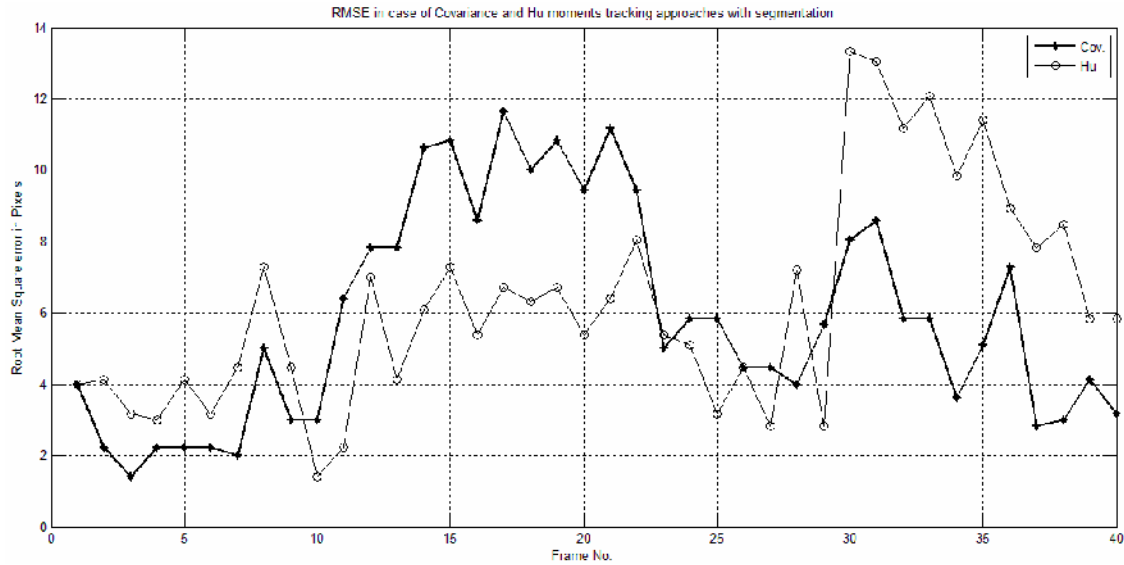*Figure (4): RMSE between actual target location and estimated location using Hu-Moments tracking technique with segmentation (*) and without segmentation (o)*

Experiment 2

As illustrated in figure 5, the system tracked an ambulance car over a monocular sequence of 40 frames. An occlusion case is introduced in this sequence, from this figure, it could be noticed that the tracking system successfully tracked the target until the end of the frame sequence. In the ambulance sequence, the desired target was selected manually in the first frame. Figure 6 contains graphs showing the errors exhibited by the system for this image sequence using the covariance-based and Hu moments-based techniques.

(a)   (b)   (c)   (d)   (e)

Figure (5): Selected images from "ambulance" image sequence after applying the tracking techniques. Starting from upper left to lower right:(a) First frame in sequence, (b) frame no.10, (c) frame no.22, (d) frame no.30, (e) frame no.40 last frame in sequence. The black (+) in the center of each black rectangle indicates the target detected by the tracking techniques.
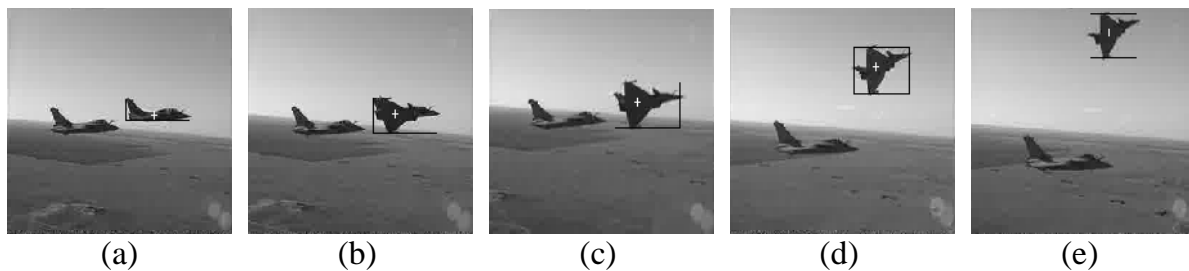
Figure (6): *RMSE between actual target location and estimated location using Covariance technique (*) compared with the RMSE between actual target location and estimated location using Hu-moments technique (o) for the tracked target in figure (5)*

Experiment 3

As illustrated in figure 7, the system tracked a fighter over a monocular sequence of 44 frames. From this figure, it could be noticed that the tracking system successfully tracked the target until the end of the frame sequence.

In the fighter sequence, the desired target was selected manually in the first frame. Figure 8 contains graphs showing the errors exhibited by the system for this image sequence using the covariance-based and Hu moments-based techniques.

(a)   (b)   (c)   (d)   (e)

Figure (7): *Selected images from "fighter" image sequence after applying the tracking techniques. Starting from upper left to lower right:(a) First frame in sequence, (b) frame no.10, (c) frame no.20, (d) frame no.37 and (e) frame no.44 last frame in sequence. The white (+) in the center of each black rectangle indicates the target detected by the tracking techniques.*
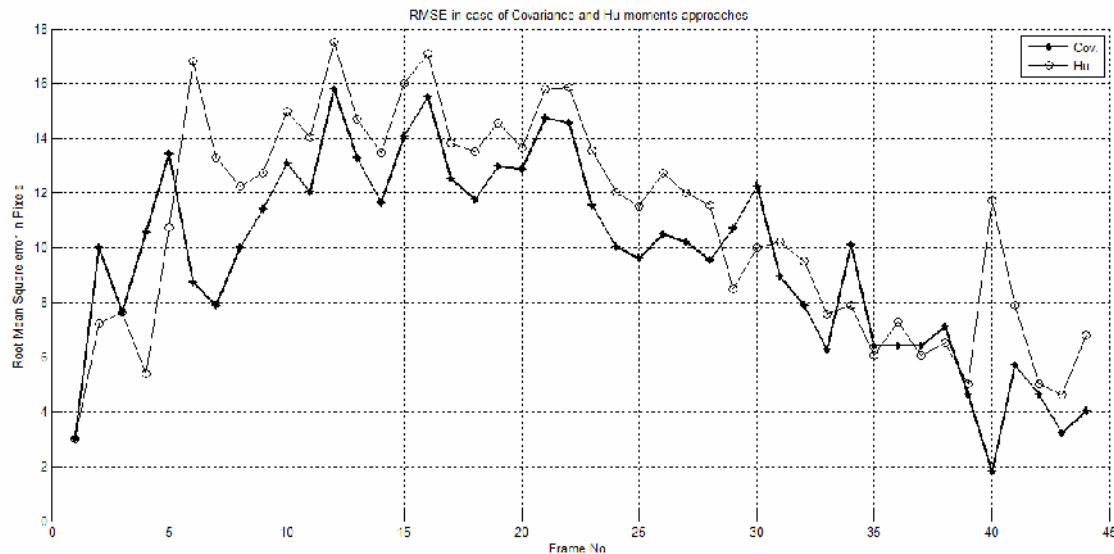
Figure (8): *RMSE between actual target location and estimated location using Covariance technique (*) compared with the RMSE between actual target location and estimated location using Hu-moments technique (o) for the tracked target in figure (7)*

5. Conclusions:

The main challenge for the implemented tracking techniques is to succeed in the tracking problem with high accuracy. The used tracking techniques overcame the occlusion problem. Both tracking techniques were implemented using a region growing segmentation technique to adapt the window size of the target. The first technique was Covariance tracking technique and the second was Hu-moments tracking technique. Different experiments were performed using the Hu-moments tracking technique against the Covariance tracking technique. Segmentation, using region growing, was applied on the two tracking techniques. The comparative analysis between these techniques proved that less tracking error (RMSE) is achieved in Covariance technique based on segmentation than the Hu moments one.

References:

[1] A. Yilmaz, O. Javed and M. Shah, "Object Tracking : A Survey", ACM Computing Surveys,Vol. 38, No. 4, Article 13, December 2006.

[2] J. Gai, "Robust target tracking : theory, applications and implementations", Ph.D. Thesis, University of Notre Dame, 2004.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 564–577, 2003.

[4]   A. Yilmaz, X. Li, and M. Shah, "Object contour tracking using level sets", In Asian Conference on Computer Vision, ACCV, pp. 1-7, 2004.

[5]   J. Lim, D. Ross, R. Lin, and M. Yang, "Incremental learning for visual tracking", In Neural Information Processing Systems, 2004.

[6]   H. Tao, H. Sawhney, and R. Kumar, "Object tracking with bayesian estimation of dynamic layer representations", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 75–89, 2002.

[7]   D. Terzopoulos and R. Szeliski, "Tracking with kalman snakes", In Active vision, pp. 3–20, 1993.

[8]   J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects", International Journal of Computer Vision, pp. 57–71, 2000.

[9]   Y. Chen, Y. Rui, and T. S. Huang, "JPDAF based HMM for real-time contour tracking", In IEEE Conference on Computer Vision and Pattern Recognition, pp. 543–550, 2001.

[10] L. Cohen, "On active contour models and balloons", CVGIP: Graphical Model and Image Processing, pp. 211–218, March 1991.

[11] B. Horn and B. Schunck, "Determining optical flow", Massachusetts Institute of Technology Artificial Intelligence Laboratory, pp. 185–204, 1981.

[12] T. Nir, A.M. Bruckstein, and R. Kimmel, "Overparameterized variational optical flow", In International Journal of Computer Vision, pp. 205-216, 2007.

[13] F. Heitz and P. Bouthemy, "Multimodal estimation of discontinuous optical flow using markov random fields", IEEE Pattern Analysis and Machine Intelligence, pp. 1217–1232, 1993.

[14] D. Fleet and Y. Weiss, "Optical flow estimation", In Handbook of Mathematical Models in Computer Vision ,Springer, pp. 239–256, 2006.

[15] F. Porikli, O. Tuzel and P. Meer, "Covariance Tracking using Model Update Based on Lie Algebra", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.728-735, 2006.

[16] M. K. Hu, "Visual pattern recognition by moments invariants", IRE Transactions Information Theory, pp. 179-187, 1962.