International Journal of Advances in Computer Science & Its Applications – IJCSIA
Volume 4: Issue 4 [ISSN 2250-3765]
Publication Date : 27 December,2014

Multiple Objects Tracking Under Occlusions: A Survey

H.A. Ali, M.A. Mohamed, M.S.El-Sayed, M.S. Diab

Abstract—**long term occlusion is a most important challenge in any multiple objects tracking system. This paper presents a literature survey of an object tracking algorithms in a fixed camera situation that have been used by others to address the long term occlusion problem. Based on this assessment of the state of the art, this paper identify what appears to be the most promising algorithm for long term occlusion that was succeeds in handling interacting objects of similar appearance without any strong assumptions on the characteristics of the tracked objects. But this algorithm failed to handle objects of too complex shapes and appearance, and the tracking results was affected by successfully of background subtraction. This paper presents a proposed solution to these failed points.**

Keywords—object trackingt; long-term occlusion; background subtraction

I. Introduction

Object tracking is becoming increasingly important component in wide range of computer vision applications such as robotics, human computer interaction, vehicle navigation, automobile deriver assistance, video surveillance, video games and industrial automation and security. Despite being classic computer vision problem, tracking is largely unsolved. This is mainly due to the fact that there are many challenges which cause the object tracker to fail such as: appearance change, scale change, distraction, illumination change, difficult motion, multiple objects, and occlusion, in this paper we are looking for a solution to the occlusion challenges. Some tracking algorithms do not deal with occlusion at all [1, 2]. Others minimize occlusions by placing the cameras so that they look down on the scene [3, 4]. Occlusion can be classified according to the causation to object self-occlusion which happens when an object changes its pose and inter-objects occlusion which means that one object is occluded by other objects. Occlusion can also be classified into partial occlusion and full occlusion or long-term and shot-term occlusion. This literature survey focuses on the cases of inter-objects occlusion, either partially or completely, long term or short term. The rest of the survey is organized as follows. Section 2 lists the main algorithms that have been used by others to solve the occlusion problems. In section 3 we identify the algorithm that appear the most promising for dealing with occlusion in general and long term occlusion in particular. Section 4, present some results of Argyros [5] methodology. In section 5, we present our proposed solution to improve the result of Argyros. Finally, Section 6 presents a conclusion of our survey.

II. Techniques Think through Occlusion

A lot of approaches have been proposed to handle the occlusion problem either implicitly or explicitly. The majority of these methods failed to handle total occlusions and assume that even partial observations of the occluded objects are possible, e.g. [6, 7]. In [6] they use the Kanade-Lucas-Tomasi (KLT) features as the representation of objects and propose a trajectory estimation algorithm with a weighting function of tracked features. This weighting function will eliminate the tracking error caused by object occlusion as long as the object is not fully occluded. [7] uses the part-based model to represent the human by providing a rich description of the articulated body; thus, it is highly discriminative and robust against appearance changes and occlusions. This part-based representation allows parts to be strictly compared to their corresponding parts. One of the most important feature of this model is avoiding the confusion from background changes by excludes most of the background within the detection window. [7] handle partial occlusions through dynamic occlusion reasoning and prediction across frames but failed in full occlusion.

Pierre F [8] divided the main approaches to solve occlusion problem into two groups: merge-split (MS) and

Hisham Arafat Ali
Faculty of Engineering/Mansoura University
Egypt

Mohamed Abdel-Azim Mohamed
Faculty of Engineering/Mansoura University
Egypt

Mohamed Salah El-Din El-Sayed
Faculty of Computers and Informatics/Benha University
Egypt

Mai Salah Diab
Faculty of Computers and Informatics/Benha University
Egypt

straight-through (ST). In all systems that are use MS approach, as soon as blobs, the things that are being detected via image processing, are declared to be occluded, the system merges them into a single new blob and characterized by new attributes until splitting again into two other blobs. When a split condition occurs the problem is to identify the object that is splitting from the group. Most of these technique that are uses MS do not address the issue of splitting and of re-establishing identity like Piater et al. [9]. As [8] conclude in his paper, ST approaches do not suffer from this problem because ST approach must be able to classify any pixel in the occlusion region as belonging to one of the occluding objects. Most systems that are using ST approach trying to build a model for each individual objects. Some model consists of the color and spatial characteristics like in [10]; other model consists of a deterministic RGB color template and a registered probability mask like in [11].

Huang [12] used the concept of object permanence, which refers to the ability of children to realize that an object exists even when it cannot be seen, to reason about full occlusion. To achieve this purpose, tracking is performed at both the region level and the object level as shown in Fig.1. Region-level tracking procedure associates foreground (FG) regions from respectively frames with each other. Object-Level tracking process used the appearance model and spatial distribution of an object to locate it by registering each individual pixel to one of the admissible objects. As shown in the structure of Huang approach in Fig.1, the occlusion relationship among neighboring objects is used also to locate unobservable objects when the corresponding region splits they emerge from existing objects.

Second there are no assumptions about motion, shape or size of objects. Third it is able to perceive the persistence of objects when they under significant occlusions and when they re-emerge from behind the occluders. And two main disadvantages, objects colors are assumed to be simple colors and limited textures, and it does not handle interacting objects of similar appearance.

Argyros [5] succeeds in treating the disadvantage of [12] by using the powerful data association mechanism that has been proposed in Argyros and Lourakis [13].

III. Argyros's Method Description

Argyros [5] needs to provide answers to the following three fundamental questions to present a robust object tracking algorithm that handle a long object occlusions. First, how are objects modeled and detected in an image? Second, what is the powerful data association mechanism that has been used to associates detected models to different objects? Third, how can make the tracking much more competent in handling long-term occlusions?

Fig.2 shows the flow diagram of the Agyros algorithm that is describing the answers of the previous questions. Which the background subtraction block has the answer of detection question, association of foreground pixels with objects block has the answer of association mechanism question, occlusion reasoning block solves the long-term occlusions and the update of object models block answer the objects model question.

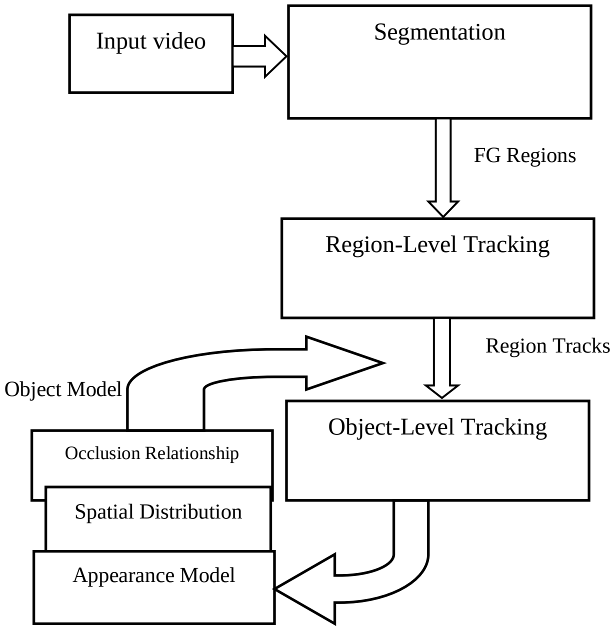


Figure 1. Architecture of Huang tracking system.

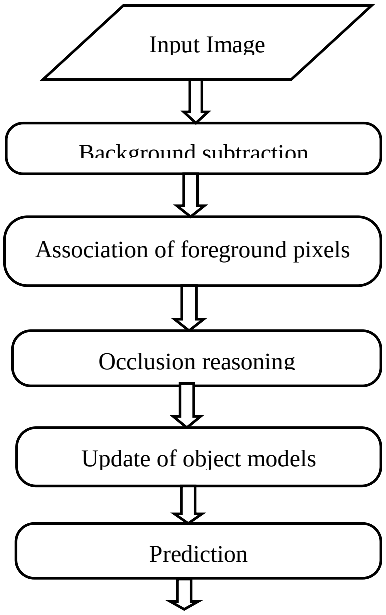


Figure 2. The flow diagram of Argyros method for tracking multiple objects in the presence of long-term occlusions.

A. Object detecting and modeling

Object detection is the basic step for every tracking method. There is a lot of common object detection methods proposed in earlier research. Yilmaz [14], provide a table of the popular methods like Point detection, segmentation, background subtraction, and supervised learning, most tracking methods for fixed camera like Argyros [5] use background subtraction methods because they are computationally efficient and have the capabilities of modeling the changing illumination, noise, and the periodic motion of the background region. Argyros [5] exploits the efficient adaptive algorithm using Gaussian mixture probability density for background subtraction that has been proposed in Zivkovic [15]. Zivkovic algorithm can reduce the processing time and improve the segmentation more than GMM (Gaussian Mixture Model) background subtraction scheme.

B. Association of foreground pixels with objects

The result of background subtraction stage is a group of distinct blobs, i.e. regions of connected foreground pixels. As shown in Fig.2, the second stage is associate foreground blob pixels with objects and to achieve that, Argyros try to build simple and generic object models then determined the relation between the outputs of previews stage (blobs) and the object models in time to define the set of pixels p(o) belonging to an object o. He decided to model an object by its appearance model, by using (GMM) that represents its color distribution, and spatial distribution to form a parametric model. More specifically, the object model (o) can be defined:

$$o \equiv (e, g) \qquad (1)$$

Where, Argyros assume that the spatial distribution of the pixels depicting an object can be coarsely approximated by an ellipse e. The appearance g of an object o is modeled as a GMM, representing the color distribution of the objects pixels. This process used YUV color representation. However, the Y-component of this representation is not employed to reduce the computational of the overall system and make a system less sensitive to illumination change.

Argyros take advantage of the powerful data association mechanism that has been proposed in [13]. The magic words in this method that's used to solve the data association problem are using both the spatial distance and appearance compatibility between a blob and an object. In [13] the morphological filtering that's applied to the result of background subtraction cannot give rise to multiple blobs for a single object, that's why a single connected object can give rise to at most one connected blob. However, the correspondence among blobs and objects is not necessarily one-to-one, two or more different objects may appear as a single blob throw occlusions.

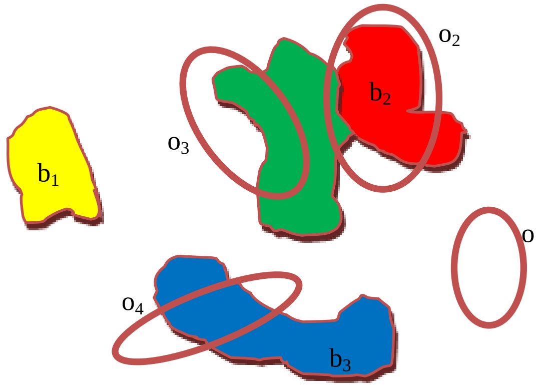


Figure 3. Possible relations between objects and blobs [5].

Fig.3 show four objects (o_1, o_2, o_3, o_4) and three blobs (b_1, b_2, b_3), it is assumed that at a given moment in time, M foreground blobs b_j,$1 \leq j \leq M$ have been detected and the N objects o_i, $1 \leq i \leq N$ are already being tracking.To find a set p(o) of pixels that's associated with object o, first, Argyroshave to find all image points in the intersection of blob b_j with objects ellipse e_i by finding the distance D(p,e) of an image point p(x,y),of bj, from an ellipse $e_j(cx_j,cy_j,\alpha_j,\beta_j,\theta_j)$ where (cx_j,cy_j) is its centroid, α_j and β_j are, respectively, the lengths of its major and minor axis, and θ_j is its orientation on the image plane as follows:

$$D(p, e) = \sqrt{\vec{v} + \vec{v}} \qquad (2)$$

Where

$$\vec{v} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \left(\frac{x - x_c}{\alpha}, \frac{y - y_c}{\beta} \right)$$

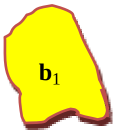
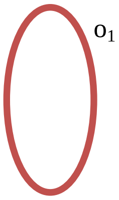
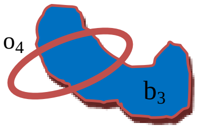
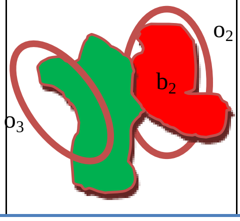
If the value of D (p, e) for (p Є b_j) is less than 1, we will conclude that the point p and the blob b_j are in intersection with object that's have the ellipse e, in other words, point p is inside ellipse e. But if the values of D (p, e) for all points p belonging to a blob $b_j$are greater than 1, we will conclude that b_j don't have intersection with all ellipses of the existing object hypotheses, like b1 in Fig.3.

Second, he tested all image points in the intersection of blob b_j with object's ellipse e_i for compatibility with the object's appearance model to find the degree C (b_j, o_i) of association between an object o_i and a blob b_j by using (3).

$$C(b_j, o_i) = \sum_{p \in (b_j \cap I(e_i))} P_A\,(p, g_i) \qquad (3)$$

Where, I (e) is a point p that is interior to the ellipse e.

TABLE I. Description of all cases that's may arise after the association between objects and blobs.

Interesting cases		Solution
Object hypothesis generation: Blobs not associated to objects. $\forall \mathbf{o_i}, \mathbf{b} \cap \mathbf{I}(\mathbf{e_i}) = \emptyset \rightarrow \forall \mathbf{o_i}, \mathbf{C}(\mathbf{b}, \mathbf{o_i}) = \mathbf{0}$ (4)		A new object hypothesis is generated and its set p(o) becomes equal to b.
Object model hypothesis removal: Objects not associated to blobs. $\left(\mathbf{U}_{\mathbf{j=1}}^{\mathbf{M}} \mathbf{b_j}\right) \cap \mathbf{I}(\mathbf{e}) = \emptyset \rightarrow \forall \mathbf{b_j}, \mathbf{C}(\mathbf{b_j}, \mathbf{o}) = \mathbf{0}$ (5)		An object hypothesis should be removed but in practice, we permit an object hypothesis to "survive" for a certain amount of time so that we account for the case of possibly poor detection or full occlusion.
Object hypothesis tracking in the presence of multiple, potential occluding objects : 1. Blobs in one-to-one correspondence with objects. 2. Blobs associated to multiple objects.	 	1. The set p(o) becomes equal to b. 2. Searching for the set p(o) of pixels to be associated with object o only within blob B(o) by using the following equation that's assigns blob pixels p to the object o* that minimizes spatial distance and maximizes appearance compatibility. $o^* = \arg\max_{o} \frac{P_A(p,g)}{D(p,e)}$ (6)

Now, each object will associate with the blob that's giving the highest degree of association. Thus, there are only four cases may arise representing three different problems needs to manage: (a) object hypothesis generation (i.e. an object appears in the field of view for the first time) (b) object model hypothesis removal (i.e. a tracked object disappears from the field of view) and (c) object hypothesis tracking in the presence of multiple, potential occluding objects (i.e. previously detected objects move arbitrarily in the field of view). All these cases described in TABLE I.

By using table 1, Argyros [5] algorithm can address all different problems described so far. But in case that a blob is associated to two objects have similar appearance (6) cannot find the correct o*. That's mean first we have to measure the appearance similarity between these two objects by employing a criterion measuring the similarity between two GMMs, this is because the presented method represented an objects appearance with a GMM. If we sure that the objects have same appearance model, there are two rules governing this association:

Rule 1: If a point of a blob is located within the ellipse of some object hypothesis then this pixel is considered as belonging to this hypothesis. In the example of Fig.4, all points in a black region associated with o_2 and all points in a blue region associated with o_3.

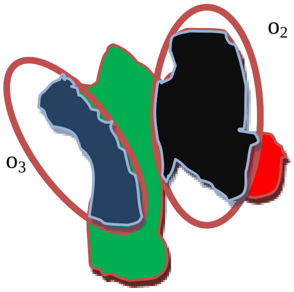


Figure 4. Blob associated with a similar appearance objects.

Rule 2: If a point is outside all ellipses corresponding to the object hypotheses, then it is assigned to the object hypothesis that is closer to it, using (7).In the example of Fig.4, all points in a green and red regions associated by using (7).

$$o^* = \arg\min_{o} D(p, e) \qquad (7)$$

C. *Occlusion Reasoning*

For handling the occlusion problem, we have to understanding whether an object is occluded or not and the identification of the occluder. There are two cases for any occlusion problem, two objects occlusions and layered occlusions.

1) Two objects partial and full occlusions

In this case, there are only two objects compete for the points of a blob. As stated earlier, one of the two objects is occluded and the other is occluder. Therefore, first we have to find the occlusion ratio R_i that's calculate the ratio between the area of object o_i at this time A_i and the area of object o_i at the last frame in which this object appeared in isolation A_i` (8).

$$R_i = \frac{A_i}{A_i^{`}} \qquad (8)$$

For the occluder object, no respectable changes in its area will be observed, that's mean $R_i \approx 1$. For occluded object, R_i can be $1 > R_i > T$, that's mean its area was decreased, or $1 > R_i < T$, that's mean o_i is disappeared because of a full occlusion.

As we mentioned before, Argyros [5] used Huang [12] method to solve the full occlusion problem by assuming that the full occluded object will stay behind its occluder object, move with it and inherits the associations of its occluder by sharing the same ellipse with it. When an occluded object start to reappears ($1 > R_i > T$), the occluded and the occluder objects are dis-associated and a new model will be constructed using the image points assigned to the occluded object.

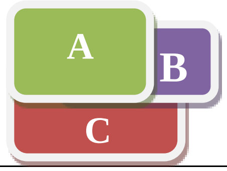
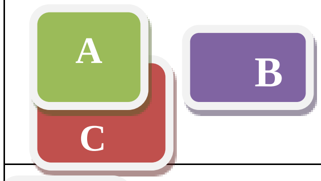
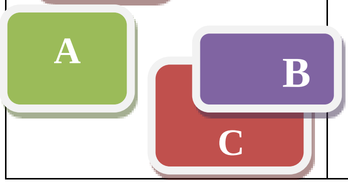
2) Layered occlusion

In this case, there are more than two objects compete for the points of the same blob as shown in Table.2. For a set of objects in a layered occlusions relation, there will always be the major occluder A and a number of occluded ones behind it B, C. All occluded objects declare all other objects as potential occluders. The reappearance of one of these B make the remaining occluded object C will be searched not only in the proximity of the original occluder A, but also in the proximity of the newly reappeared object B as described in TABLE II, and the label of the reappeared object B will be removed from the list of all of its potential occluders.

D. *Update of object models*

After having assigned blob pixels to object hypotheses, an update of the object o_i = (e_i, gi) are re-estimated based on the sets p (o_i). The parameters of e_i can be derived directly from the statistics of the distribution of point's p(o_i). More specifically, the parameter of e can be computed from the covariance matrix of the location of pixels in p(o_i), i.e. matrix whose element in the I,j position is the covariance between the i th and j th elements of a random vector. The appearance model g_i is computed through the application of Expectation Maximization algorithm over the colors of the image points in p(o_i) and it is updated only when observed in isolation, but in the case of partial or full occlusion the appearance model is stopped from being updated.

TABLE II. Layered occlusion

	Multiple objects A, B, C participates in an occlusion relationship.
	Object B is appears and object C stay around object A
	Object B is appears and object C stay around object B

E. *Prediction*

As stated earlier, data association is based on object model that have been formed at the previous time step and the detected blobs. Therefore, there is a time lag between the definitions of models and the acquisition of data these models need to represent. For that, instead of using the ellipses as those were computed in the previous frame, a simple linear rule can be used to predict the location

IV. Results

In this section, we show experimental results of Argyros [5] object tracking method. The proposed method has been tested and evaluated in a series of image sequences demonstrating challenging tracking scenarios. Results from several representative input video sequences are presented in Argyros [5]. The first such image sequence (''objects" sequence) consists of 1280 frames and shows a person manipulating several objects on a tabletop. Characteristic snapshots demonstrating tracking results are shown in Fig.5. The sequence scenario is as follows. More specifically, (Fig.5a) shows the empty desktop on which the experiment is performed and of which a background model has been built. In (Fig.5b), the human hand has already brought into the scene a box containing a few objects. Having no a priori knowledge about the scene other than a background model of it, the system identifies the constellation of the hand, the blue box and the rest of the objects as a single multicolor object, for which it builds a single object model. As soon as the hand leaves the box on the table (Fig.5c), the originally connected set of pixels becomes disconnected. The original object hypothesis (red contour) is assigned to the blue box because this is more similar to the previous box/ hand constellation. Another object (hand, green contour) is automatically generated. For the next frames, the hand color appearance model is updated. The same happens also to the appearance model of the blue box, in which the components corresponding to the previously joined hand now vanish. The hand interacts with the box again (Fig.5d). Now, the color models built assist the method in correctly assigning the pixels of the single connected blob to the two object hypotheses (hand, box). In (Fig.5e), the hand has taken the pincer off the blue box and moves it to another position on the table. For the moment, the method interprets this as a change in the appearance of the hand and, at that stage; the pincer appears as part of the hand object. This is because the pincer has never been observed in isolation but only as part of another object (box). As soon as the hand leaves the pincer on the table, the pincer is understood as an individual object (Fig.5f, blue contour). The identity of the pincer object is not lost even when the hand passes several times over it, grasps it and moves it to another place on the table (Fig. 5g–j). In a similar manner, the hand empties the basket. As shown in (Fig. 5k), the hand, the box and the pincer maintain their original identity, while the two other objects have acquired their own object identities. In (Fig. 5l), the hand has grasped the object with the purple contour and has used it to completely occlude

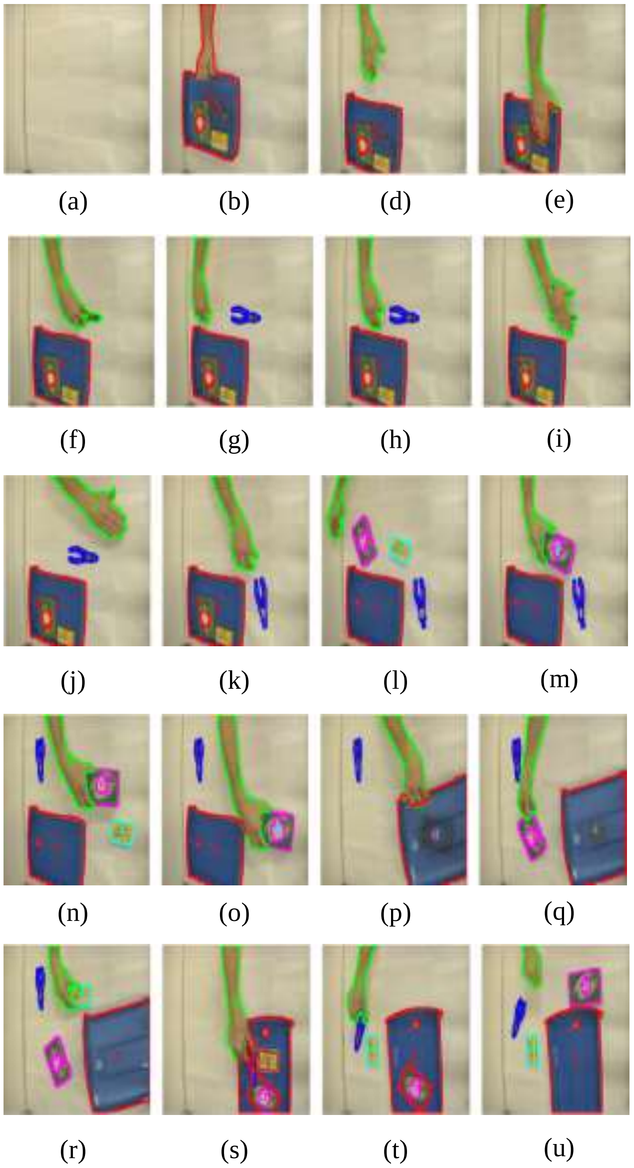


Figure 5. Characteristic snapshots from the tracking experiment on the "objects" image sequence [5].

the one with the cyan contour. The full occlusion has been signaled and both object hypotheses are maintained and tracked together with the observed region of the occluder. Both objects are transferred to a new position, the hand removes the occluding object (Fig.5m) and the correct identity for the occluded object is still maintained. The purple object is again brought on top of the cyan one, fully occluding it once more. This time, the big box is also brought on top of the purple object creating a layered occlusion (Fig.5o). When the hand brings the purple object again in sight dragging it under the big box, the purple object still maintains its original identity (Fig.5p). The same happens to the cyan object (Fig.5q).The manipulation of objects continues; the hand brings all objects again into the blue basket and starts moving the latter around (Fig.5r).The experiment ends with the hand emptying the basket once more (Fig.5s and t).

V. The Proposed Work

Argyros [5] tracking result was affect by successfully of background subtraction. If background subtraction results have many false negatives, a single object may appear as a set of disconnected foreground blobs. Moreover, if it has too many false positives, background will be mixed with objects and their appearance models may drift and fail to accurately represent them. Thence, we trying to find a robust background technique to solve this problem and produce a robust object tracking technique. Based on our research, we found a background method that uses Three Temporal Difference (TTD) and the GMM approach for object tracking presented by Luke K and Jen-Hong Lan [16]. This method uses the GMM approach as the main tracking algorithm and TDD algorithm i.e. the use of a continuous image subtraction, as spare. In other words, TDD algorithms used to fill the hole produced by GMM algorithm. The results of [16] shows that their technique combine the advantages of GMM and TTD that's shown in TABLE III, and their technique can be used to address the failed of [5] and produces a robust multiple objects tracking technique.

TABLE III. Advantage and disadvantage of GMM and TTD.

	GMM	TTD
Advantage	Complete results of the operation	Quick calculations
Disadvantage	Not a complete object tracking	A lake of complete object tracking

VI. Conclusion

In this paper, we extend the main techniques that have been used to track multiple objects under full, partial, long and short occlusions. In each case of occlusions, we references to specific systems that solve this type of occlusion. After that, we review a specialized technique that have been proposed by Argyros [5] for tracking multiple objects in the presence of long-term occlusions and a simple solution for its expected fail situation have been presented. The obtained experimental results of Argyros demonstrate that the developed tracking methodology can successfully handle occlusions in challenging situations. The tracker in [5] has very simple models of object shape, appearance and motion; this makes the tracker simple, fast and generic in the sense that no strong assumptions are imposed on the characteristics of the tracked objects. Argyros approach is expected to fail when objects to be tracked have too complex shapes and appearance or move with irregular motion patterns and if background subtraction has many false negative or positives. We expect removing these drawbacks by using a detection method based on background subtraction combined three temporal difference methods that are combined TDD and GMM approach for object tracking.

References

[1] A. Azarbayerjani, and C. Wren,"Real-Time 3D Tracking of the Human Body," IMAGE'COM 96, Bordeaux, France, May 1996

[2] T. Olson, and F. Brill," Moving Object Detection And Event Recognition Algorithms For Smart Cameras,"DARPA Image Understanding Workshop,Volume 1, Pages 159-175, New Orleans, LA, 1997.

[3] R.T. Collins, A.J. Lipton, and T. Kanade," A System forVideo Surveillance and Monitoring," Technical report, Robotics Institute at Carnagie Mellon University, 2000.

[4] W.E.L. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site, "Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, pp. 22-31, 1998.

[5] A.A. Argyros, and V.P Papadourakis," Multiple objects tracking in the presence of long-term occlusions," Computer Vision and Image Understanding, 114(7):835-846, 2010.

[6] Bing Han, Christopher Paulson, Taoran Lu, Dapeng Wu, and Jian Li," Tracking of multiple objects under partial occlusion," in: Proc. SPIE 7335, Automatic Target Recognition XIX, 733515, 2009.

[7] Guang Shu, Afshin Dehghan, Omar Oreifej, Emily Hand, and Mubarak Shah," Part-based multiple-person tracking with partial occlusion handling CVPR," cvpr, pp.1815-1821, 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[8] [8] P.F. Gabriel, J.G. Verly, J.H. Piater, and A. Genon," The state of the art in multiple objects tracking under occlusion in video sequences," in: Advanced Concepts for Intelligent Vision Systems, pp. 166-173,2003.

[9] [9] J. H. Piater, and J. L. Crowley," Multi-modal tracking of interacting targets using Gaussian approximations," Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp: 141-147,2001.

[10] [10] A. Elgammal, and L. S. Davis," Probabilistic framework for segmenting people under occlusion," International Conference on Computer Vision, Vol.8, pp: 1815-1821, 2001.

[11] [11] A.W. Senior, A. Hampapur, L. M. Brown, Y. Tian, S. Pankanti, and R. M. Bolle," Appearance Models for Occlusion Handling,"International Workshop on Performance Evaluation of Tracking and Surveillance systems, Vol.2, 2001.

[12] [12] Y. Huang, and I. Essa," Tracking multiple objects through occlusions," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1051–1058,2005.

[13] [13] A.A. Argyros, and M.I.A. Lourakis," Real-time tracking of multiple skin-colored objects with a possibly moving camera," European Conference on Computer Vision, pp: 368–379,2004.

[14] [14] Jiyan PAN, Bo HU, and Jian Qiu ZHANG," Robust and Accurate Object Tracking underVarious Types of Occlusions," IEEE Transactions on Circuits and Systems for Video Technology, pp: 223-236, 2008.

[15] [15] Z.Zivkovic," Improved adaptive gaussian mixture model for background subtraction," Proceedings of the International Conference on Pattern Recognition (ICPR),Vol.2, pp: 28-31, 2004.

[16] [16] S.C. Luke K, and W. Jen-Hong Lan," Moving object tracking based on background subtraction combined temporal difference," International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'2011) Bangkok Dec,2011.