A Survey on Occlusion Detection and Handling
Ms. Chandani Gandhi¹, Mr. Jaydeep Viradiya²

¹Department of Computer Engineering, PIET Vadodara
²Department of Computer Engineering, PIET Vadodara

Abstract - Object tracking and detection is a classical research area in the field of computer vision. Numerous kinds of applications are dependent on the area of object detection, such as advance driving assistance system, traffic surveillance, scene understanding and autonomous navigation etc. Tracking of moving targets in crowded scenes is challenging task for many applications. Object detection is a difficult job in real time tracking of multiple objects due to occlusion. When objects are occluded, some parts may not be visible to human vision. Note that, during occlusion missing object parts need to be recovered. The basic step in tracking system is to accurately detect the objects in different environments like complex backgrounds, weather conditions, cast shadows most importantly due to occlusion. Occlusion at faces may be leads to performance degradation of face recognition algorithms. So the occlusion handling is core challenge for any tracking applications. This paper represents a survey on detection of occlusion and handling it, because it’s important phase for many real time systems.

Keywords - Tracking, Face recognition, Occlusion, Human vision, Occlusion detection, Occlusion handling.

I. INTRODUCTION

Occlusion is one of the main performance reduction problems in video surveillance systems. Occlusion means that there is something you want to see, but can’t due to some property of your sensor setup, or some event. In simple words occlusion is hiding of an object by another object during multiple object tracking. [1], [2] If an objects you are tracking is hidden by another object then the hidden object is called as occluded. By which the object is hidden that object is called as occlude. Occlusion can be of three types that shown in figure 1.

![Types of Occlusion](image)

Figure 1: Types of Occlusion

Self-occlusion: Some parts of object is occluded, by other part of same object called self occlusion. This will occur frequently. [3] We can see the examples. In figure (A) the one portion of house is hidden by the other portion of same house. Same as in figure (B) we can’t see her whole face because of her own hand it’s called as self occlusion.

![Self Occlusion in Image](image)

Figure 2. Self Occlusion in Image
**Inter-object occlusion:** two objects being tracked occlude each other. [3] Figure 3 shows the inter object occlusion. Under this figure human bodies are overlapped and walking together in a scene. Because of bodies overlapping we can’t see some part of their bodies.

![Figure 3. Inter Object Occlusion in Image](image)

**Background occlusion:** This occurs if objects are hidden due to background objects in a scene. [4] Figure 4 shows that tracked object are not properly visible because of the background object.

![Figure 4. background occlusion](image)

Visual surveillance system is used, to detect, recognize, and track certain objects in a scene. This type of system was mainly used in applications such as security for human, important building, military target detection and traffic surveillance in cities. It is essentially a video recording system that is used for past-event analysis. In the earlier stages, human beings are watching the videos in such type of systems to check for any unusual activities. These systems cannot provide sufficient security by various issues like occlusion. An efficient video surveillance system must be fast, reliable and use of robust algorithms for moving object detection, classification, tracking and activity analysis. So such system has to raise warning on the occurrence of any suspicious events. Moving object detection is the important phase for further analysis of the video. Occlusion is one of the main performance reduction problems in any real time tracking systems. All automated occlusion detection system should accurately monitor occlusion. When the detected objects in a scene come behind another object, some parts in the objects become undetected due to occlusion.

In the human tracking system face recognition is one challenging task, because the faces might be masked either purposely using sun glasses or mask unintentionally like scarves or crowded places it’s called occlusion. Depending up on the places such as banks, the occlusion may be suspicious. Because of face occlusion, the performance degradation of system will occur. So researches in the last decade have concentrated on improving the performance of the human detection system under conditions like occlusion. Occlusion is the major challenges for such systems. The systems that have addressed occlusion can be classified according to how they handle occlusion are human detection without occlusion detection, with occlusion detection and localization of occlusion. The first category does not distinguish occlusion and non occlusion. Second category can detect occlusion partially. The third category can not only detect the occlusion but also locate the occluded region.

The main problem in occlusion detection is that occlusion cannot be detected directly. To detect occluded part, the pixels under goes or going to be occluded is detected. Occlusion was mainly detected for the purpose of restoring the occluded parts in an image. Occlusion detection in complex environment can be improved by using multiples view from different sensors or cameras. [4] That Explain bellow. So, accurate occlusion detection and interpretation will help the user to get the information accurately.
The remaining of this paper is as follows. Section II is a review of the existing methods for occlusion detection. Section III is a comparison table for different methods. Section IV is a different method of occlusion handling section V limitation of existing system. Section VI concludes this paper.

II. METHODS FOR OCCLUSION DETECTION

i) SVM and Block weighted
In the work [5], Zhaohua Chen, Tingrong Xu, and Zhiyuan Han have solved the problem of face recognition under occlusion due to sunglasses or scarves.[4] This method is used to improve the recognition of face occluded by sunglasses and scarves.[5],[4] Here the presence of sunglasses or scarves was detected first and the non-occluded region only was processed that shown in figure 5. Occlusion can be dealt by selecting non-occluded patches from the faces. Occlusion was detected by PCA and support vector machines (SVM). To detect the occluded region in the face, divide the images in to finite number of patches and examine each patch separately. As configuration and size of the patches are important in the performance of occlusion detection they have divided faces into six symmetrical patches. Then dimension of these patches were reduced by using PCA. While the recognition of non-occluded facial parts is performed using blocked-base weighted local binary pattern (LBP). Whole process of SVM and blocked weighted is introduced in [5].

![Figure 5. (A) Normal image (B) occluded image (C) Non-occluded image](image)

ii) Object labeling & bounding box
In the work [6], Sherin M. Youssef, Meer A. Hamza and Arige F. Fayed, was proposed an algorithm to detect and track multiple objects under occlusion. They have detected and tracked the objects using discrete wavelet transform (DWT) and identifying the objects by their color. Since discrete wavelet transform has a nice property that it can divide a frame into four different frequency bands without loss of the spatial information, it is adopted to solve this problem due to the fact that most of the fake motions in the background can be decomposed into the high frequency wavelet sub-band. A bounding box was created around each object and it was labeled.[6] The object was scanned from right to left, left to right, top to bottom and bottom to top to create bounding box.[4] The top left co-ordinates, height and width of the each bounding box was created after scanning. Occlusion was detected by analyzing the height and width of the bounding box. Here in this figure 6 objects are labeled with A and B. Each window show the different situation respect to time.

![Figure 6. Object labeling & bounding box](image)
iv) Merging and splitting approach
This method proposed a real time system for multiple human tracking in dynamic scenes. This work can deal with complete and long duration occlusion. Here they have considered object states in to three categories that is; before, during and after occlusion. System consists of object segmentation part, merging and splitting detection module for occlusion detection. [4] Occlusion was identified only after the merging and splitting of blobs in each object occurs. [7] Merging and splitting blobs is shown in figure 7.

![Figure 7. Blob Merging and splitting at the time of occlusion](image)

Merging and splitting detection procedure based on the obtained classification results. For those non-matched track, a merging detection algorithm is used to decide whether the track is merged by another measure or is missed. If a merging happens, a new group is generated. If the track is missed, the confidence of the track will be decreased, once it drops below a specific threshold, the track will be deleted. For those non-matched measures, a splitting detection module is developed to decide whether the measure is split from an active track or it is a new target. [8] When a splitting event is confirmed, a feature correspondence module is performed to labeling each object correctly.

![Figure 8. A scenario of blob merging and splitting detection. The first row contains the blob merging events and the overlapping areas. The second row contains the splitting events](image)

Merging might occur due to a non-matched track overlapped with a measure. This judgment is based on the assumption that there must be overlapped area between the initial merging bounding box and the merged object (Figure 8, 1st row). This is a valid assumption when the segmentation process is fast enough, as soon as object touches with each other at time k+1, a large bounding box contains all the merged objects will be created and it has large overlapping areas with the merged objects at time k. Fortunately the moving object segmentation method mentioned in achieves 20fps in the surveillance system, fast enough to detect merging event even with high speed objects.[8] Similar to the merging method above, splitting is detected due to a non-matched measure overlapped with a track (Figure 8, 2nd row). When a group
splits, each split object will be labeled correctly with a feature correspondence method. (correspondence and moving object segmentation method is further described in [8]).

iv) Color, texture and motion
Valtteri Takala and Matti Pietikainen [9] introduced a novel real-time tracker algorithm based on color, texture and motion information. RGB color histogram and correlogram are used to describe the object’s color properties. The merging and splitting of objects are handled using the same set of features. The main sources of descriptors are color, texture, shape, and temporal (motion) properties. But the color has gained the most of attention as it is well distinguishable to human eye and seems to contain a good amount of useful information. Tracker used in paper [9] consists of two main elements background subtraction (detection) and tracking. The subtraction on the video data, which is first processed with a Gaussian filter to remove noise, is done by an adaptive algorithm. The subtracted foreground is enhanced by filtering the artifacts caused by noise and moving background using standard morphological operations. The tracking is done by matching features extracted from the subtracted foreground shapes. These are color, texture and motion. Instead of using the bounding boxes themselves for occlusion detection, this system surrounds the boxes with circles that have the radius of the half diagonals of the boxes and use them for event detection. If the object circles are occluding each other in the previous frame n-1 a merging event in frame n is possible. If the occlusion is true in frame n + 1 and the closest occluding object is a group object 1(1&2) then a split might have happened.

III. COMPARISON OF DIFFERENT METHODS

Table 1. Comparison of different Occlusion Detection methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Type</th>
<th>Full/partial/self Occlusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM and Block weighted</td>
<td>Facial occlusion</td>
<td>Partial</td>
</tr>
<tr>
<td>Object labeling &amp; bounding box</td>
<td>Human occlusion</td>
<td>Partial/full</td>
</tr>
<tr>
<td>Merging and splitting approach</td>
<td>Human occlusion</td>
<td>Partial/full</td>
</tr>
<tr>
<td>Color, texture and motion</td>
<td>Human occlusion</td>
<td>Partial/full</td>
</tr>
</tbody>
</table>

IV. METHODS FOR OCCLUSION HANDLING

Occlusion handing is a major problem in visual surveillance. Typically, during occlusion, only portions of each object are visible and often at very low resolution. [13] This problem is generally intractable, and motion segmentation based on background subtraction may become unreliable. To reduce ambiguities due to occlusion, better models need be developed to cope with the correspondence between features and body parts, and thus eliminate correspondence errors that occur during tracking multiple objects. When objects are occluded by fixed objects such as buildings and street lamps, some resolution is possible through motion region analysis and partial matching. [13] However, when multiple moving objects occlude each other, especially when their speeds, directions and shapes are very close, their motion regions coalesce, which makes the location and tracking of objects particularly difficult.

i) Padding using occlusion-compatible scanning order
The first approach exploits the assumption that occluded pixels belong to the background. Therefore, an appropriate solution for padding occluded pixels is to fill undefined occluded regions by copying neighboring background pixels. To this end, the color of a neighboring background pixel should be determined. A proposal for doing this is to use the occlusion-compatible scanning order. The idea is to scan the destination image such that foreground pixels are scanned prior to background pixels. As a result, background pixels always overwrite occluded pixels. Note that, as opposed to the method is explain above in this Section for handling overlapping pixels, the original occlusion-compatible ordering algorithm scans the source image such that the background pixels are scanned prior to the foreground pixels. For holes, we thus scan in the opposite order.
ii) **Decomposition method**

Method provides a solution to problem of occlusion in images by removing the occluding region and filling in the gap left behind. Inpainting algorithms fail in filling occlusions when the occluding region is large since there is loss of both structure and texture. [10] In this method first decompose the image into structure and texture images using a decomposition method based on sparseness of the image. Figure 9 shows the decomposition method. Decompose image into structure and texture and inpaint the missing region during reconstruction. Decomposition and inpainting are for retaining lost structure. Decompose the original image into textural and structural sub images. The textural sub image is restored by a texture synthesis and the structural sub image is reconstructed by a structure algorithm. They are then combined to form the inpainted image.

![Image of decomposition method](image)

**Figure 9. Shows work of decomposition method** [10]

iii) **Texture synthesis**

This method is to deal with even large occlusion by incorporation a texture synthesis on the texture only image. Decomposition and inpainting for retaining lost structure followed by a patch based texture synthesis on the texture image for restoring the lost texture. In texture synthesis we copy patches of texture from known areas in the image. This method of image completion does not introduce any blur effect since actual pixel values are copied. Disadvantage of this method required large computation time. [10] Finding the most appropriate patch for a query region is fundamentally hard. The best matching patch is found always nearer to the source region. [10] Patch size also affects the result of texture synthesis. If the patch size is too small each patch will have insufficient information to perform best match. However if the patch size is large it will result in fewer patches and the system will not have sufficient number of patches to perform best match’ Large patch size has a result of reduction in computational time as more patches get filled at a time. We make patch size proportional to whole size and image content. Small patches are used for fine structures and large patches for large structures.

iv) **Patch based stereo algorithm**

In the work, Yi Deng and Qiong Yang Described the patch based framework to handle occlusion. Here the objects are divided in to many patches. This concept uses two cameras to capture the image. If any part of the image which is not visible in one camera can be obtained from the captured image in other camera. The major limitation is the co-alignment of multiple camera co-ordinates. [1] Segmentation was applied to both images and warps the segment of one image to the other by disparity. The warped segment is then divided in to small patches on the boundaries of other images. The occlusion at boundary can be treated as the occlusion at patches. A frame work using graph cuts was used to find the disparity and occlusion. A symmetric frame work using graph cut is constructed to find the disparity and occlusion at patches.[11]

V. **LIMITATIONS OF EXISTING SYSTEM**

i) **Part-based method**

This method is for face and car detection. The method consists of a set of substructure detectors, each of which is composed of detectors related to the different parts of the object. [12] The limitation of this method is that the different parts of the object need to be manually labeled in the training dataset, in particular, eight parts for face detection and seven parts for cars.

ii) **Whole-Object Segmentation method**

The method requires a hierarchical object parts design with eleven components making up the head, the torso and the legs. The edge pixels of the object that positively contribute to the part detectors are extracted and used together with the
part detector responses to obtain a joint likelihood of multiple objects. [12] In case of finding any inter-object occlusions, the occluded parts are ignored. The main limitation of this method is that it requires a manual spatial alignment of the objects, which has to be adapted to each object class. In addition, it requires a special camera set-up in which the camera has to look down on the ground-plane.

iii) Decomposition method
In this method provide a solution to problem of occlusion in images by removing the occluding region and filling in the gap left behind. Using this method decompose the image into structure and texture images. That explained above, finally the structure and texture images are combined to get an image where the occlusion is filled. Limitation of this method is however suitable only when inpainting region is small. [6]

iv) Texture Synthesis
They propose to extend decomposition method to deal with even large occlusion by incorporation a texture synthesis on the texture only image. [6] Limitation of this method is required large computation time for best patch matches.

VI. CONCLUSION
From this paper we conclude that many real time applications like Video surveillance system, Advance driving assistance system show poor result due to occlusion. For such a system most challenging issue is the proper identification of occlusion and handling it. Occlusion detection system gives better results while employing blob or bounding box creation around the objects. Many Occlusion handling methods gives better result but sometimes it required more computation time. The computation time of handling occlusion is depends on the complexity of image. Our main focus is to improve results by various occlusion detection and handling techniques.

REFERENCES

[1] Smitha Suresh, Dr.K Chitra, Deepak.P “Patch Based Frame Work for Occlusion Detection in Multi Human Tracking” IEEE ©2013
[5] Zhaohua Chen, Tingrong Xu, and Zhiyuan Han, "Occluded Face Recognition Based on the Improved SVM and Block Weighted LBP", IEEE 2011
[8] Tao Yang, Stan Z.Li, Quan Pan and Jing Li, “Real-time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes”, Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2005