2D Image Recognition of Partially Occluded Objects Based on Adaptive Window Positioning and Dynamic Time Warping

Mohammed Almuashi¹, Siti Zaiton Mohd Hashim², Dzulkifli Mohamad³ and Mohammed Hazim Alkawaz⁴

Abstract

The boundaries of computer vision are especially challenged by partial occluded object recognition. In this paper, we proposed a new algorithm that used to recognize the partially occluded objects in 2-D images. Two different cases of occlusion are considered. Firstly, when an object missing some parts. Secondly, objects are overlapping each other. This approach uses contour of an object. Adaptive Window Positioning (AWP) uses to extract features from a contour. The proposed method uses the orientation field as features of the fragment. Dynamic Time Warping (DTW) works for matching and find the similarity between sub-images. As experiment result, the proposed recognition algorithm works without prior segmentation and robust to identify missing and overlapping objects and it can works with strength occlusion.

Keywords: Partial Occluded Object Recognition; Dynamic Time Warping; Adaptive; Window Positioning; 2D images

1. Introduction

Object recognition is spread across too many fields such as industrial, image retrieval and medical models. Occlusion problems happen to be one of the biggest problems when it comes to object retrieval and recognition. The boundaries of computer vision are especially challenged by partial occluded object recognition. It is possible to divide the 2-D model-based recognition method into another two classes: (1) recognition of the occluded object and (2) recognition of the single object. For single object recognition, it is necessary for the object to be completely visible and no occlusion to be present within the image, even if there are many objects present. In the second occluded object and could occur because: (a) another object is blocking or overlapping the primary object, (b) the lighting conditions are poor or there is a reflection/glare falling on the object, (c) the object is missing a part, or (d) The angle from which the picture has been taken of the object makes identification difficult.

The single objects recognition with fully shapes and features has been studied for a long period. It can be implemented easily by using many existing algorithms. Due to the impact of occlusion, this will cause a problem in recognition of the object. That, in the event that there is more than one object was in touching or overlapping, the system will consider these objects are one object. It was decided that the best method for recognizing a partially occluded object was by pointing out the important boundary features of the unknown shape. This gave rise to a new set of problems where it was hard to determine which are the appropriate features that define the object?

There are two shape representation methods in use: contour-based technique (shape features are determined on the basis of contour only) and region-based technique (shape features are extrapolated from the total shape area). Region based techniques are not preferred because generally, the shape’s interior is not as important as the contour. We can define a ‘feature’ as a measurement function that points out a specific quantifiable object property. Its computations are performed on
the basis of the objects main features [14]. During the process, there is always a gap present between what is required and what is available.

The features of a shape can be classified into local or global to describe object. Local feature are generally at the border of an object otherwise imply a discrete small part of a section [20]. Certain usually employed local features are curves, edge fragments, and corners [5]. A lot of approaches have been evaluated by the researchers by means of several local features, the same as boundary dominant points, curve segments and wavelet descriptors. Jiayun et al. [11] suggested a way in terms of color, quality and geometry to merge local feature collaborations, into spectral coordinating so as to improve recognition under occlusion. On the other hand, a method is represented by [7] for the sake of identifying 2-D occluded object. For the attainment of the features through the object, wavelet algorithm is employed in addition to that; it is also used for matching procedure to classify. Global features generally are certain properties of areas in pictures for example area (size), perimeter, Fourier descriptors, in addition to moments [7].

Global feature requires the objects being recognized to be wholly visible, non-overlapping and not touching each other. So, the drawbacks of using global features for object recognition include sensitivity to clutter and occlusion, and difficulty in localizing an object in an image. Object recognition algorithms based on global features fail to work when partial occlusion takes place, where global features are severely contaminated. To recognize the occluded object by using features such as color and texture, we find that features are not extremely important at least in our study. In other hand, when integrating those features and use them with the local features may we get powerful features to identify the occluded object.

In the recent past years, we can say that the area of recognition of occluded object has received attention from researchers. Certain researchers [1, 10, 22] made use of recognition methods centred upon dominant point so as to identify partly occluded objects. To solve the high computation load problem in occluded recognition dominant points are used. An amalgamation of dynamic programming and neural network was employed by Gupta and Upadhye [9] to recognition of occluded object. A group of dominant points is identified on the way to divide the contour into sections. Each section, which is composed of three successive points, is categorized as a local feature. It is not suitable to use dominant points based object recognition to recognize occluded objects.

The most popular representation to objects of partly occluded recognition is the polygonal approximations [3, 17]. Repetitive, consecutive and optimum algorithms are frequently used for such approximations [6]. Ayache and Faugeras [2] used a technique to match plain descriptions of the models and occluded objects which name is H.py. This method centred the problem of occluded recognition domin. Ayache and Faugeras [2] used a technique to match plain descriptions of the models and occluded objects which name is Hypotheses Predicted and Evaluated Recursively (HYPER). They resorted a polygonal approximation method to represent the 2-D shapes to excerpt the feature primitives that form the boundaries of the object. Schwarts and Sharir [18] found the shortest paths near polygonal curve and put forward an algorithm that helped partly identifying occluded objects in either two or three dimensions. Polygonal approximation doesn’t hold importance for all types of objects but is confined to only a few objects. Polygonal objects are the ones for which polygonal identification methods work and this is considered the main drawback of this approach.

In addition, there is also an effective technique which will be viewed here quickly and concisely. An algorithm was presented by [12] for the extraction and location of a partly visible entity from an input image. They utilized Genetic algorithm to deal with the problem of identification of partly occluded object. The image object is identified by the algorithm through comparison between polygonal fragments of the image and polygonal fragment of the model. Lim et al. [16] and Lichun et al. [15] addressed the problem of identification of partly occluded objects by using curve
moment invariance, Ansari et al. [1] used technique based on landmark to deal with the same problem. By the landmarks of an object we mean the regions of interest having necessary features of the shape. Landmarks include holes, corners, high curvature points and projections. To execute matching, sphericity is employed which is a measure resemblance between two landmarks. Sreyasee et al. [4] used significant edge of an object as feature to recognize occluded in 2-D object. Relationship between geometrical configuration, texture and color employed to detect the occluded object and represented by Fan [8]. Vimina and Jacob [21] used texture features and local color of the object region and global features of the image. In this approach, object is segmented into constant partitions and edge is determined by using edge thresholding.

In this paper, we proposed a new algorithm that used to recognize the partially occluded objects in two different cases: an object missing some part and objects are overlapping each other and to determine as to how to take out features from an occluded object. With taking into account the effect of the percentages of the occlusion. This includes the Adaptive Window technique for extracting the features from an object. In addition, Dynamic Time Warping (DTW) works for matching between objects. Moreover, the orientation field is used to calculate the angle of a window’s fragment.

2. Proposed Algorithm for Occluded Object

The proposed algorithm consists of several processes as shown in Figure 1.

![Proposed Algorithm for Occluded Object](image)

**Figure 1. Proposed Algorithm for Occluded Object**

2.1. Dataset

Our dataset that we used is MPEG-7 taken from the internet [13]. The silhouettes image is a candidate to be the dataset in this study. In fact, the dataset MPEG-7 is not for the occluded objects. Due there is no dataset (binary) for the occluded objects yet. We generate occluded objects from the MPEG-7 dataset by merge two
single objects manually to build dataset of occluded objects. Images that we have in the dataset are not in same size; thus, we have to re-size it. We opted for the missing objects a fixed size which is 256×256 pixels and 350×350 pixels for overlapping objects. MPEG-7 dataset contains over 1407 images, we did not use it all, but we selected images randomly to be employed in our study. Dataset outline shows in the following Table 1. The dataset was built on the basis of the type of occlusion the missing objects and overlapping objects, in addition to the rate of occlusion in an object.

<table>
<thead>
<tr>
<th>Missing</th>
<th>Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ≤ % ≤ 10</td>
<td>22</td>
</tr>
<tr>
<td>11 ≤ % ≤ 20</td>
<td>12</td>
</tr>
<tr>
<td>21 ≤ % ≤ 50</td>
<td>15</td>
</tr>
<tr>
<td>% ≥ 51</td>
<td>11</td>
</tr>
</tbody>
</table>

The total images for the missing objects are 60 images which divided into four sections of occlusion percentages. In the overlapping objects, contains 36 images; each image is composed of two objects and the total of those objects that composed for each image is 72 objects.

2.2. Contour Detection

The lines at the edges of geometric shapes in case of digital images are called contours. In this study representation of object is interest to local feature with a view to recognition of occluded object. Our proposed algorithm uses object’s contour to extract features. Basically, the binary images represent as black “0” value which denotes as foreground while white “255” value for background. To determines and extract the contour or edges present in an object we use Sobel edge detector algorithm.

2.3. Features Extraction Using Adaptive Windowing Positioning (AWP)

The Adaptive window positioning algorithm is a practical component, which eventually results in to small sub-images in a division of an object. However, the Adaptive Window position technique has been used in the recognition of handwriting [19].

According to our recognition, division helps in creating the fragment of an object, which brings the significant comparison of these fragments. However, we have to measure the angle $\theta$ for every fragment in each window. We have denoted the square windows of size $n\times n$ thus have selected this division, as we have already operated in a fixed size ($13\times13$). Initially, we need a division of an object in to small sub-images or fragments if we want to excerpt the features or local features and distinguish the object. Thus this division takes place by putting the small windows. Each window contains a fragment as shown in Figure 2.

![Figure 2. Fragment inside the Window](image-url)
For our recognition, we have created an adaptive division technique that is an adaptive technique for the contour trace. We requisite an adaptive method so we can regulate the position and location of the window which is to trace the contour. However, for each window they have figured the feature orientation and positioning. The following illustrates the steps of the algorithm. The outlined of the algorithm is shown in Figure 8.

i. Find the large number of pixels on the object contour.

![Figure 3. Find Large Number of Pixels on the Contour](image)

ii. Place the first window to find the pixel on the contour.

![Figure 4. Start To Trace Window on the Contour](image)

iii. Choose the next window by define four directions according to the following conditions:
   a. If the contour locates from East or West, place the next window to right (East) or left (West) respectively of the current window.
      - Move the window up and down (vertical direction) to get the good position of window on the contour.

![Figure 5. Select the Direction and the New Position of the Window (North Direction)](image)

   b. If the contour locates from North or South, place the next window to top (North) or down (South) respectively of the current window.
      - Move the window left and right (horizontally direction) to get the good position of window on the contour.
c. If the contour locates from more than one direction, we deal with each of the branches separately.

iv. Final greatest positioning for the new window.

v. Repeat step 3 until the contour of the object covered by the windows.

Figure 6. Select the Direction and the New Window (East Direction)

Figure 7. Placed a New Window in the Greatest Positioning

Figure 8. Algorithm for Position Windows on the Contour
2.4. Compute Orientation in Window

The orientation is the feature of the fragment that finds in every window on the object contour. The local variance of gradient magnitude of each window is calculated in terms of the subsequent steps:

i. For every single pixel of the standardized fingerprint image \( N(i, j) \); estimated gradients in horizontal and vertical directions, which are signified by \( G_x(i, j) \) and \( G_y(i, j) \), respectively are calculated by using the following Sobel mask \( 3 \times 3 \) matrix.

\[
\begin{bmatrix}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1
\end{bmatrix}
\]

(a) (b) (1)

(a) Horizontal Sobel mask matrix \( S_x(m, n) \) and (b) Vertical Sobel mask matrix \( S_y(m, n) \). Then,

\[
G_x(i, j) = \sum_{m=1}^{n} \sum_{n=1}^{m}(S_x(m, n) \times N(i + m, j + n))
\]

(2)

\[
G_y(i, j) = \sum_{m=1}^{n} \sum_{n=1}^{m}(S_y(m, n) \times N(i + m, j + n))
\]

(3)

Let \( V_x(u, v) \) and \( V_y(u, v) \) be as in Eq.

\[
V_x(u, v) = \sum_{i \in \Omega} \sum_{j \in \Omega} (G_x^2(i, j) - G_x^2(i, j))
\]

(4)

\[
V_y(u, v) = \sum_{i \in \Omega} \sum_{j \in \Omega} 2G_x(i, j)G_y(i, j)
\]

(5)

ii. Compute gradient angle \( \theta(u, v) \) for every window using the equation:

\[
\theta(u, v) = \frac{1}{2} \arctan \left( \frac{V_x(u, v)}{V_y(u, v)} \right)
\]

(6)

Now, the standardized form of \( \theta(u, v) \) (i.e. angle beyond 180˚ is set to 0 – 180 consequently) is provided by:

\[
\theta(u, v) = \begin{cases} 
\theta(u, v) 
& \text{if } V_x(u, v) \geq 0 \text{ and } V_y(u, v) \geq 0 \\
\theta(u, v) + \pi/2 
& \text{if } V_x(u, v) \leq 0 \text{ and } V_y(u, v) \geq 0 \\
\theta(u, v) + \pi/2 
& \text{if } V_x(u, v) \leq 0 \text{ and } V_y(u, v) \leq 0 \\
\theta(u, v) + \pi 
& \text{if } V_x(u, v) \geq 0 \text{ and } V_y(u, v) \leq 0
\end{cases}
\]

(7)

2.5. Matching Windows Using Dynamic Time Warping (DTW)

After the sub-images represented by a set of features, we want to select a similarity measure that permits us to compare two sub-images. DTW algorithm is used to discover the corresponding points of two sequences or signals. Dynamic Time Warping (DTW) is one of the new techniques that were developed for pattern similarity matching for two time series sequences. To compute the DTW algorithm and get the matching value, there are three steps must be followed:

a. Build the distance matrix by using:

\[
\gamma(i, j) = d(X_i, Y_j) + \min \begin{cases} 
\gamma(i - 1, j - 1) 
\gamma(i - 1, j) 
\gamma(i, j - 1)
\end{cases}
\]

(8)
where, $x_i$ and $y_j$ both are elements of this matrix contains the distance between two sub-images which is articulated by \( d(x_i, y_j) \).

b. Compute the short path in the distance matrix and find the best warping or alignment path by using dynamic programming:

\[
\gamma(i,j) = \min \left\{ \gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1) \right\}
\]  

(10)

c. Calculate the similarity between two sequences of sub-images as following:

\[
DTW(X,Y) = \min \left\{ \sum_{i=1}^{K} d(W_k) \right\}
\]  

(11)

we use the minimum distance for selecting the distance of \( DTW \) sequences of \( X \) and \( Y \) as shown in equation (10).

3. Experimental Result

In this experiment, the query image consists of two sections: (1) missing parts of object as shown in Table 2 and (2) overlapping objects as shown in Table 3. The objects in both are not in one style as it can take different cases of the ratio of the occlusion.

<table>
<thead>
<tr>
<th>Table 2. Recognition Results for the Missing Part of the Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query image</td>
</tr>
<tr>
<td>_____</td>
</tr>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image 5" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image 7" /></td>
</tr>
</tbody>
</table>

Through the above table, there are four images have been tested in different percentages of occlusion. Moreover, all the images that used in this experiment are the in same size (256x256). Those images have been tested depending on the strength of the occlusion. The number of features extracted from the image is associated with the occlusion.
percentage. That’s mean, whenever the occlusion percentage was tiny, the number of features will be large. On the contrary, the higher percentage of the occlusion, the number of features will be few, and this leading to an imbalance in the performance of our algorithm. In this context, when the object is under severe occlusion such as the percentage is more than 50%, there is only a very small portion of features detected from the object.

Given that all these features are correctly matched, the percentage of matched features for the object is too low for it to be considered recognized, leading to a low confidence of recognition. Therefore, whether the object is recognized is related to its occlusion percentage. In our proposed algorithm, only features that are highly associated with the object are considered for matching, which can be interpreted as taking the occlusion percentage into consideration.

The experimental results demonstrate that this occluded object recognition system is capable to retrieve an occluded object in case of a missing part under different percentages of the occlusion. In addition, the system is capable to identify and recognize the object, even if it is in a different position. Moreover, it is able to identify the object even the percentage of the occlusion is too high. This is due to several factors: the location of the missing part of the object and its type, whether it is distinguished or unique in the object. Table 3 shows overlapping objects recognition result.

### Table 3. Recognition Result of the for the Overlapping Objects

<table>
<thead>
<tr>
<th>Query image</th>
<th>Image feature</th>
<th>Obj1 Sim</th>
<th>Obj1 Sim</th>
<th>Occluded strength</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td>28%</td>
<td>18%</td>
<td>simple</td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td>45%</td>
<td>14%</td>
<td>severe</td>
</tr>
</tbody>
</table>

According to Table 3, there are two images have been tested as a query under the effect of occlusion strength which is one of the important objectives of this research. Moreover, all the images that used in this experiment are the in same size 350×350.

Approach to find similarities for the missing and overlap are not the same. The missing image is consisting of a single object and one matrix of features for this object while the overlapping image consists of two objects and one matrix of features for both objects. So, we face difficulty in obtain the similarity in overlapping image; what are the features that related to which object? Therefore, the following illustrates how the overlapping image processing and find similarities.

i. Compared an image in the query with all images in the dataset.
ii. Sort the Results.
iii. The first image of the result is the first object.
iv. Remove shared windows between the queries.
v. Compared the rest windows of the query with all images in the dataset.
vi. Sort the results.
vii. The first image of the result is the first object.

The performance of the algorithm is evaluated on the basis of the existence of two overlapping objects in the image as well as a difference in the strength of the occlusion.

Table 4 and Table 5 respectively show the performance of the algorithm for missing parts (consisting of 60 queries with different ratio of occlusion) and overlapping objects (consisting of 36 queries with different ratio of occlusion).
### Table 4. Missing Parts Recognition Matching Results and the Accuracy

<table>
<thead>
<tr>
<th>%</th>
<th>10%</th>
<th>20%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top</td>
<td>Top 1</td>
<td>Top 5</td>
<td>Top 1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>85%</td>
<td>65%</td>
<td>85%</td>
</tr>
</tbody>
</table>

### Table 5. Recognition Matching Results and the Accuracy

<table>
<thead>
<tr>
<th>Object in scene</th>
<th>Tree and carriage</th>
<th>Horse and classic</th>
<th>Fork and octopus</th>
<th>Cup and guitar</th>
<th>Face and hat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows</td>
<td>83</td>
<td>92</td>
<td>198</td>
<td>86</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object recognize</th>
<th>Tree and carriage</th>
<th>Horse and classic</th>
<th>Fork and octopus</th>
<th>Cup and guitar</th>
<th>Hat</th>
</tr>
</thead>
</table>

Accuracy for 36 objects = 79.16

### 4. Conclusion

The target of this paper is to partially occluded object recognition. This technique is contour-base; the features extracted from a contour. Adaptive Window uses to extract features form an object's contour. While, Dynamic Time Warping (DTW) employed for matching and get the similarity between sub-images in both query and dataset objects under different ratio of occlusion. Experiments were carried out on two types of occlusion, missing parts and overlapping objects. As experiment result, the proposed recognition algorithm works without prior segmentation and robust to identify missing and overlapping objects and it can works with strength occlusion. The possible future works which improvements the algorithm includes; enhance the algorithm to deal with scale and rotation invariant. Integrate local features and global features. Enhance the algorithm to deal with recognition if occluded object in real time.

### Acknowledgments

I would like to express my special thanks of gratitude to the Saudi Ministry of Higher Education, Cultural Mission in Malaysia, UTM-Big Data Center as well as University Technology Malaysia (UTM) who gave me the golden opportunity to do this paper on the topic ”2D image recognition of partially occluded objects based on adaptive window positioning and dynamic time warping”, which also helped me in doing a lot of research and I came to know about so many new things I am really thankful to them.

### References

Acknowledgments

Mohammed Almuashi, he was born in Saudi Arabia. He received the Diploma in Programming from Technical & Vocational Training Corporation (College of Technology) in 2006 and B.Sc in computer science and information system from Universiti Teknologi Malaysia, Johor Bahru, in 2011. In addition, M.Sc. degree in computer science from the Universiti Teknologi Malaysia, Johor Bahru, in 2013. He is a current Ph.D. student in Universiti Teknologi Malaysia, Johor Bahru. His research interests include Computer Vision, Image Processing and Multimedia Database.

Siti Z. Mohd Hashim, she is an Associate Professor at the Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia (UTM). She received her B.Sc. Degree in Computer Science from University of Harford, USA, M.Sc. in Computing from University of Bradford, UK and Ph.D. in Soft Computing from The University of Sheffield, UK. Her research interests are Soft Computing techniques and applications, System Development and Intelligent System. She is now the member of UTM Big Data Centre of

Authors

References
Dzulkifli Mohamad, he is a senior professor in the Faculty of Computing Universiti Teknologi Malaysia (UTM) Johor Malaysia. His keen research interest are facial animation, intelligent features mining, Pattern Recognition and Security. He has supervised several Master and PhD students in the field of Computer Science and Information Security. He is author of more than 300 publications that are published in world class conferences and journals of international repute.

Mohammed Hazim Alkawaz, he was born in Mosul, Iraq, in 1989. He received the B.Sc. degree in computer science from University of Mosul, Mosul, Iraq, in 2010, and the M.Sc. degree in computer science from Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia, in 2013. He received the Ph.D. degree in computer science from Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia in 2015. He won the Best Postgraduate Student Award in the faculty in 2013 for his M. Sc and for his Ph.D. as well in 2015. He is currently a lecturer in Department of Information Science and Computing, Faculty of Information Sciences and Engineering, Management and Science University, Selangor, Malaysia.