

PROBLEMS FINDING OBJECTS AND MONITORING THEM IN CASE OF OBSTACLES

Narine Kabakulakyan

Ph.D. in Economics

Robert HAKOBYAN

Master student at NPUA

Keywords: Object detection, object tracking, computer vision, Haar cascade

Introduction. In today's world, as technologies evolve with exponential speed, almost every aspect of our lives becomes more and more dependent on AI. One of the most growing AI fields is computer vision. Using computer vision tasks like image processing, object detection, object tracking, and text recognition makes tremendous achievements. One of these achievements is the autonomous security systems, which are heavily used in our daily lives. These use-cases vary from simple home door lock systems and/or smartphone AR camera toys to highly secure government applications and national bank security systems. Object detecting and tracking in an environment with a presence of multiple obstacles is a very common task, but the existing solutions sometimes can be very expensive in computational resources and expensive in general. This paper addresses several concepts of image processing, object detection, and object tracking and suggests optimal use cases for each of the spoken concepts, and also suggests the best general solution for this task.

Methodology. Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images, videos, or live footage, for example, from security cameras. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do. Computer vision is concerned with the automatic extraction, analysis, and understanding of useful information from a single footage frame or a sequence of frames. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multidimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems [Roshtkhari, Levine, 2016, 239-255]. Computer vision suggests solutions for a variety of problems some of which are listed below:

Recognition.

- Object recognition
- Object classification
- Object localization

- Frame segmentation
- Object detection

Recognition. In computer vision, recognition is a general term that describes the process of analyzing, classifying digital images and identifying different objects in them. These three main problems of object recognition in the digital frame often create confusion about the meaning and difference between these terms.

Object recognition. The term object recognition is mainly used to encompass both image classification (a task requiring an algorithm to determine what object classes are present in the image), as well as object detection (a task requiring an algorithm to localize all objects present in the image) tasks [Sonka, Hlavac, Boyle, 2008, 131-188].

Object classification. Take a look at Figure 1 shown below:

It's obvious that there is a dog in the image. Let's pause for a moment to analyze how we recognize it instantly. As we already have knowledge about how a dog looks, we can classify the object type shown in the picture above as a dog. And this is a basic concept of the purpose of object classification (Figure 2).



Figure 1

This example describes a case where there is only one main object shown in the image. In the case of having multiple objects in the image, the object recognition algorithm must iterate over all possible predicted classes for every object to correctly classify all the objects. After all classification jobs are finished, the object recognition algorithm stumbles upon determining objects' locations in selected frame of the current content.

Object localization: Object localization refers to identifying the location of one or more objects in an image and drawing a bounding box around their extent (Figure 3).

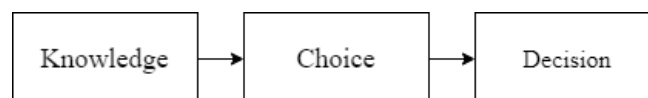


Figure 2. The basic concept of object classification

Frame Segmentation: When we have an object recognition problem, before applying the object detection, classification, and even localization algorithms, it is crucial to

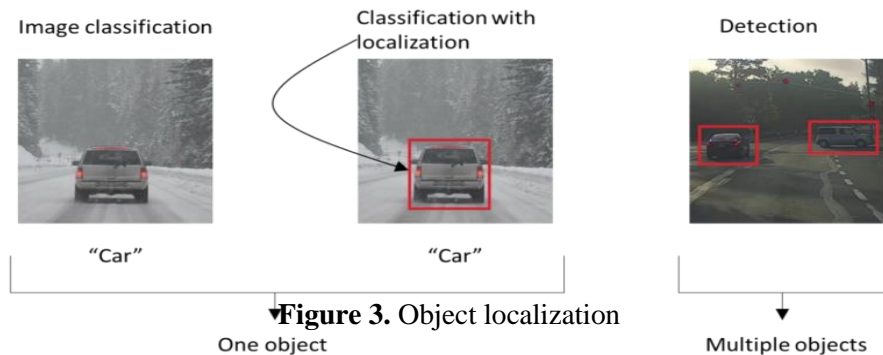


Figure 3. Object localization

understand the consistency of the selected frame firstly. To solve this problem, we rely on the concept of frame segmentation. The algorithm of frame segmentation is basically dividing the selected frame into segments and individually analyzing them to get the essential data for the particular problem. The reason for this procedure is to get rid of the unnecessary data that the selected frame possibly can content.

It is common knowledge that any image consists of a set of pixels. The way the frame segmentation works is by grouping together the frame’s pixels that have similar signatures: color range, pixel location, etc. (Figure 4).

Object Detection: As described above, the object localization task is responsible for the single object localization in the selected frame. In the case when the object recognition algorithm must deal with multiple object localization problems, object detection concepts can be used. Object detection does multiple object localization in a single frame [Luo, Xing, Milan, Zhang, Liu, Kim, 2021, 99-129]

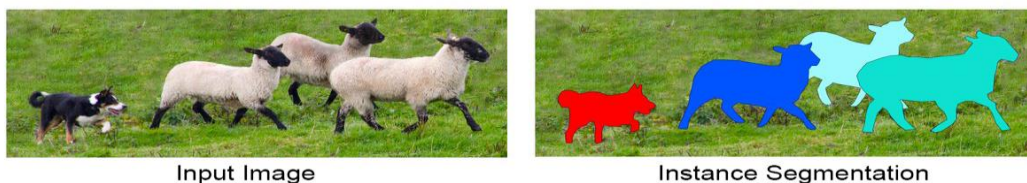


Figure 4. Frame segmentation example.

Literature review: Summing up the computer vision recognition problem. Now we have a theoretical understanding of how each of its main components work.

- Object classification - classifies the content of the current frame
- Object localization - locates a single object in the current frame
- frame segmentation - creates a pixel-wise mask for each object in current frame
- Object detection - locates multiple objects in the current frame

In Figure 5 you can see examples of each problem.

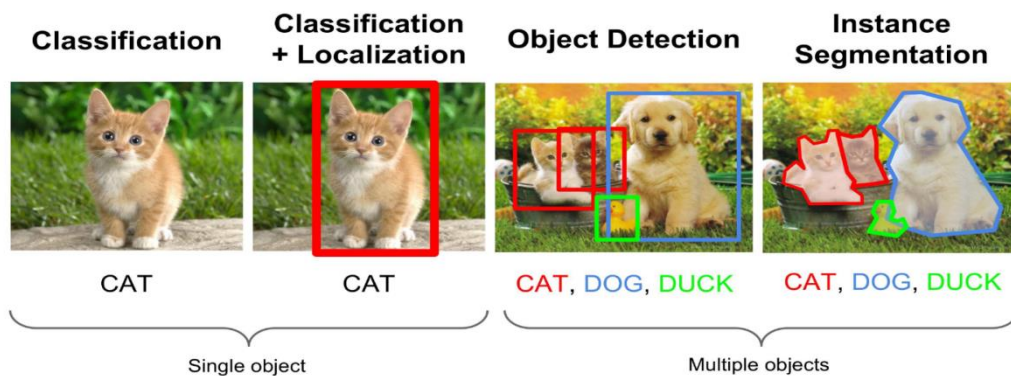


Figure 5. Examples of each computer vision problem

Economic importance: In modern organizations, it is very important to ensure the security of the organization's premises, which is why it is often necessary to hire a large number of security personnel. By replacing them with cameras and the possibility to follow the trespasser, in the long run the organization will avoid the consequences of human error, which will result in minimization of the security-related losses. Moreover this article is offering multiple solutions for decreasing the possibility of the intruder to hide behind different objects from the cameras and ensuring the security of the restricted area.

Analysis. In an environment with a presence of a lot of obstacles, the object detection and tracking problem can be very difficult and in some cases even impossible. To solve this problem some solutions are described below. These solutions may vary in different scenarios.

- Separation of the object of interests from the entire frame
- Generating rough 3D map by one frame and descriptions for each detected object
- Switch the view if it is applicable for the problem

Possible solution 1: This is a very simple solution for this problem, but it is very handy in many scenarios. The separation of the object of interest can be done in many ways. One of these ways is using object detection based on object color and shape. This method can be used in cases when the object of interest has one main color and has a non-complex shape (for example, “red ball” - the object has a circular shape in a 2D frame and it has the one main color, which is red). For separation, the object detection algorithm needs to be initialized with the object of interest, then the tracking algorithm needs to extract that object from the mainframe and consider the case with an empty rectangle and the “shadow” of the tracking object (Figure 7). As you can see, the tracking algorithm (left side) is using the white shadow of the actual red object. The other way to consider only one object from the whole frame is by using Haar cascades. A Haar cascade is a machine learning-based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. Let’s discuss an example with a face tracking problem. To train the

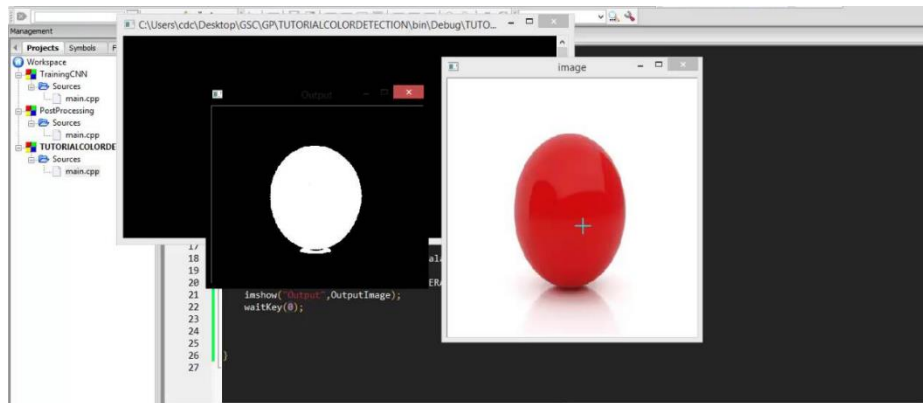


Figure 7. Creating shadow of the red ball.

classifier, the algorithm requires a large number of positive images (images of faces) and negative images (images without faces). After that, we must collect features from it.

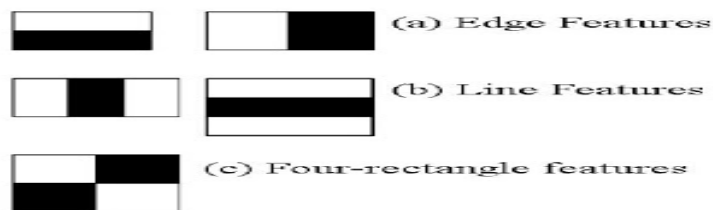


Figure 8. The Haar features

Haar features, as seen in Figure 8 below, are used for this. Each feature is a single value calculated by subtracting the number of pixels beneath the white rectangle from the sum of pixels beneath the black rectangle. (Just imagine how much computation does it need? Even a 24x24 window results in over 160000 features). But usually, the non-face area of an image makes up the majority of the image. As a result, using a simple method to verify whether a window is not a face region is a better idea than analyzing the whole image in order to find the face. If it isn't, throw it out in one go and don't bother processing it again. Instead, concentrate on areas where a face may appear. We may spend more time testing potential face regions in this way. To avoid this let's use the cascade of classifiers. Instead of applying all 160000 features on a window, the features are grouped into different stages of classifiers and applied one by one. If a window fails the first stage, discard it. So, the window, which passes all stages is a face region [Jalled et al., 2016, 1-6]. After determining the face region, the tracking algorithm needs to just cut it from the whole frame and use it like it was used in the above example with a ball.

Possible solution 2: The main goal of this solution is to generate information about the 3D objects in 2D video footage. For example, the video frame shows the big empty room with a square floor. The camera that captures all is mounted in the nearest

top right-hand corner of the room. In the middle of the room, there is a solid box. Also, there is a moving blue ball in Figure 9.

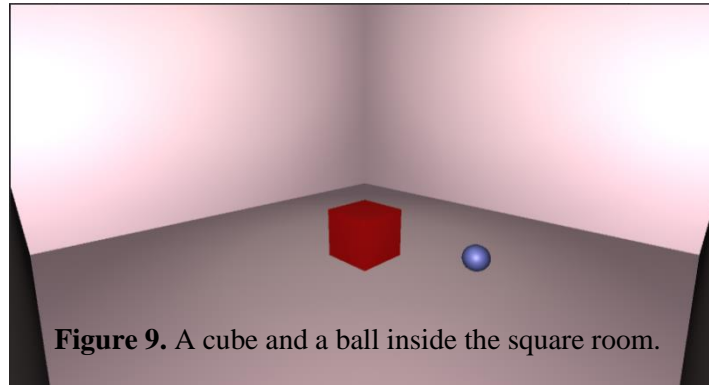
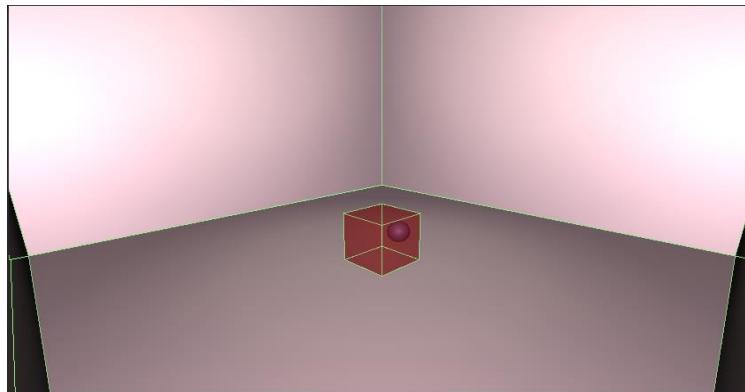


Figure 9. A cube and a ball inside the square room.

When the ball gets behind the cube, the object tracking algorithm will lose the target and it can't be recovered automatically. But what if the algorithm can determine that the obstacle is a cube with certain dimensions and the ball is behind it (Figure 10).



For this task, the algorithm must use several techniques. First of all, it must determine the 2D cube representation as a 3D object to calculate the size of edges. Then the algorithm must use egomotion to determine possible positions of moving the ball after it comes out from the behind of the cube. After that, it must initialize several threads for motion tracking algorithms for each possible position of the frame in which the ball can appear after coming out. When one of the algorithms catches the ball, the other threads must be killed to save the processing resources.

Possible solution 3: Now let's discuss the case when the ball goes behind the cube but the object tracking algorithm can't determine the 3D sizes of the cube and can't perform an egomotion algorithm. The simplest and obvious solution to continue object tracking is to find another viewpoint so the cube doesn't overlap the ball in the selected frame (Figure 10). The main problem of this solution is to calculate which viewpoint to

choose. For this purpose, the object tracking algorithm must use a very well-trained neural network because the object can have a non-symmetric shape.

Conclusion. Computer vision has many use cases and many algorithms for each possible problem. They are based on human vision and decision techniques. For object detecting and tracking in an obstacle-full environment, the algorithm must be

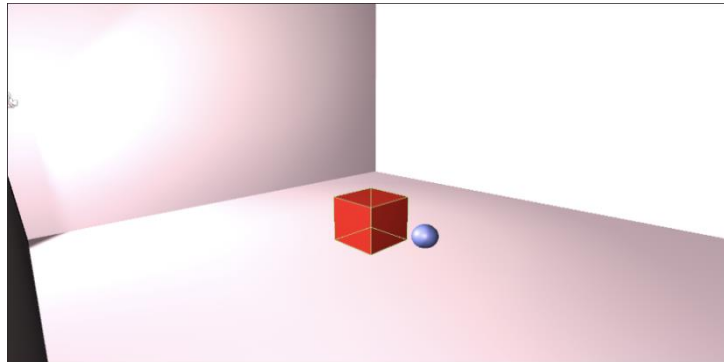


Figure 10. The other viewpoint in the room.

dynamically adjustable and must combine several solutions. It must handle as many corner cases as it can, but it doesn't need to be very complex to save some computing time and resources. We have spoken about 3 simple solutions for this problem, but there can be more solutions depending on selected environment and object of interest.

References

1. M. Roshtkhari, M. Levine. Handbook of pattern recognition and computer vision, 2016, p.239
2. Milan Sonka; Vaclav Hlavac; Roger Boyle. "Image Processing, Analysis, and Machine Vision." Thomson. ISBN 978-0-495-08252-1. 2008, pp. 131-188
3. W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, T. Kim, Multiple object tracking: A literature review, Volume 293, 2021, pp. 99-129
4. Fares Jalled, Iliia Voronkov, Object Detection using Image Processing, 2016, pp. 1-6

Narine Kabakulakyan, Robert HAKOBYAN

Problems finding objects and monitoring them in case of obstacles

Keywords: object detection, object tracking, computer vision, Haar cascade

Object detection and tracking in an environment with obstacles is a common problem, but existing methods can be costly in terms of processing resources and expensive in general. This paper investigates the object detecting and tracking techniques using object recognition, object classification, object detection etc. Moreover it suggests a possible solution for object tracking with obstacles in the frame content using computer vision tools. Automation of tracking people who entered restricted areas in any organizations can lead to long-term profits, as the need to have a large number of security personnel will be lost. As a result this approach may prevent serious consequences from happening. Furthermore, this article provides different approaches for reducing the intruder's ability to hide behind various objects from the cameras and maintaining the prohibited area's security.