Advanced Engineering Technology and Application
An International Journal

http://dx.doi.org/10.12785/aeta/020301

Moving Objects Detection in Intelligence Video Surveillance

Atuahene Samuel[*], *Qin Shihong and Yuan Fating*

Department of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan, China

Received: 3 Feb. 2013, Revised: 21 Apr. 2013, Accepted: 3 Jun. 2013
Published online: 1 Sep. 2013

Abstract: Multiple driving systems (in reference to target recognition and tracking) is a critical element in electronic warfare (EW), advanced avionics, smart weapons, and many other additional fields. In this article, we introduce a new technique for multiple systems operation for cars. Our system enables a vehicle, with its artificial intelligence, to recognize another vehicle as a target. The target takes full access in controlling the trajectory of the vehicle and enables the following car to track it exactly and continuously from one point to another as far as the recognition area is not blocked by other objects. To begin with, after capturing the image to obtain the reference image, the binarization process continues in order to get the target information with the coordinates of "x" and "y", then the width (w) and height (h). The third process is to search the image area by using a reference block to match the obtained target in the current image. Next, it computes the different size of the last image and current image of target to get speed instructions of the following vehicle to be able to control its speed. Again, the direction becomes an important factor to consider, according to the difference of the last coordinates and current coordinates of the target's centre point. The system will have components same as a robot. A movable physical structure to move the body structure, a motor of some sort, a sensor system that receives information about the body and the surrounding environment, a power supply to activate the muscles and sensors and a computer "brain" that processes sensory information and tells the muscles what to do that controls all of these elements. Of course, also some intangible attributes, such as intelligence and morality, but on the sheer physical level, the list above about covers it. An analog camera, this will serve as the eyes for the car. It will capture images which will be manipulated by a main board (computer "brain") which has been fixed on it for current location and position of target for successful tracking.

Keywords: Intelligent Systems, Real-Time Surveillance, Security Systems, Object Detection

1 Introduction

Attempts to scientifically examine natural technology have increased remarkably in recent years. Before modern technology came into existence, different forms of transportation existed. Some of these were Chariots, which were found towards the end of the 18th century, and entered the Vatican Collections in 1804 after being sold by Antonio Pazzaglia, who was a famous engraver of precious stones and had restored it by assembling original and heterogeneous (by chronology and origin) parts. A recent restoration has permitted reconstruction of the chariot starting with a few original elements of which we are certain. Another example would be Palanquins, also known as palkis, were luxurious methods used by the rich and noblemen for travelling. This was primarily used in the olden days to carry a deity or an idol belonging to a god, and many temples have sculptures of them being carried in a palki [1].

The purpose of this investigation is to extend earlier research on driving systems. A reasonable person might question the need for additional investigation since a rich body of evidence already exists on this topic. However, previous evidence has resulted in contradictory conclusions, although a growing number of analyses indicate a positive link. Furthermore, almost all of the previous evidence was derived from samples of firms from multiple industries. This research bets to differ since it concentrates on the evidence from the higher technological manufacturing industries. .

Finally, this investigation is believed to make a contribution in this heated debate by providing empirical evidence from a single industry that has a set of unique characteristics by offering additional insights into the question and to also mitigate some of the measurement

[*] Corresponding author e-mail: samfozzzy@gmail.com

problems of earlier research. According to the findings of this study, a driving system is one of the most important, yet one of the most overlooked and misunderstood concept in car technology. Nowadays, car manufactures rather concentrate on features involved in making money like interior comfortability, the technology needed to create fame for their company, or other aspects instead of driver fatigue and other minor factors that causes road accidents.

Activities using objects in motion using a video is a challenging sciatic problem. It brings to our attention, like many other researchers, a whole number of institutions and other commercial companies. Our motivation in researching this problem is to enhance an earlier study on recognition and tracking system based on real-time intelligent systems. The system has been presented to sum up almost all of the above methods

This article has been organized as follows:

The first part introduces the good work formal researchers put in place for driving systems and motivation.

Secondly, the introduction of our algorithm used to conclude the project (even though it seemed difficult), which was done successfully

And finally, discuss the tools used, experimental results, conclusion, and future predictions.

2 SYSTEM ARCHITECTURE

A) Target car

First, let's take missiles into consideration. There are a lot of ways to guide a missile to reach its given target. The missile's target accuracy is a critical factor for its effectiveness. Most missiles use heat from the target to track it when both the missile and the target are in the air, while others may use guidance system such as INS, TERCOM or GPS when target is in motion or static. This guidance mentioned above guides the missile by knowing the current position of the missile and the position of the target, and then calculating a course between them. With this technique, the following car must be guided accurately in other not to target and track other objects in the background. There wouldn't be enough heat from the target or a significant sound to be tracked. A specified shape and color was used as the target. Unlike the missile, which must crush and destroy a target, our aim is to only track the target preventing colliding and destruction, because of this, the movement of the following car will be controlled and at an assigned distance between the target and the following car, there should be a repelling force. The speed will be control by the size of the target given. If the shape appears to be smaller, instructions will be sent noticing that the target is distant so acceleration will be needed. The closer the target, the slower the following car must move indicating the closeness of the target. To make this possible, at the tail of the target car is a red rectangular shaped flag for the following car to pursue. It serves as a distinct color which could be followed by ignoring other colors that appears in the real-time captured image. The target car is controlled by remote or could be manually operated to insure movement at a constant speed. The target car could be any remote controlled car as it's not the main object of the study.

B) Following car

The following car consists of features like a normal car, (steering wheel, braking system, ignition, clutch, accelerator, etc.), but will also have components same as a robot. A movable physical structure to move the body structure, a motor of some sort, a sensor system that receives information about the body and the surrounding environment, a power supply to activate the muscles and sensors and a computer "brain" that processes sensory information and tells the muscles what to do that controls all of these elements. Of course, also some intangible attributes, such as intelligence and morality, but on the sheer physical level, the list above about covers it. An analog camera, this will serve as the eyes for the car. It will capture images which will be manipulated by a main board (computer "brain") which has been fixed on it for current location and position of target for successful tracking.

C) Video Capturing System Architecture

The video capturing system is composed of video decode model, DSP (TMS320DM642) and external memories. The signal from the camera will be connected to the codec chip TW2804. The analog video signal is converted into a digital signal and the image signal will be separated from the synchronize signal. The digital image signal from the decoder goes through the video port of the DSP and then to the FIFO, then it will be sent to the external memory of the DSP. The processed image will be saved in the output cache.

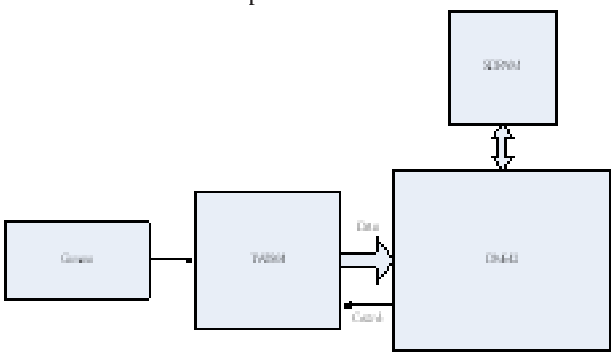


Figure 1: Video Capturing System Architecture.

3 SYSTEM ALGORITHM

A) Binarazing The Captured Image.

After capturing the image, unlike humans with natural intelligence to identify a given target, a computer may not, so the first step is to separate the target which appears in the image from the background for the given target to be detected from the current scene. Image binarization has been a topic of interest in image processing and computer vision during these last decades. This is understandable as binary images can be much easily exploited by classical pattern recognition techniques. Digital images have two possible vales, since each pixel can be classified as a binary image. Black and white is the two basic colors normally used in binary images even though other two colors could be used. One color is used to identify the foreground whiles the other color for the background. There are many binarization techniques which have been used in the past few years. But two main approaches exist, both being global approaches and local ones. Global technique exploits the information provided in the entire image, while the local technique uses local variations in small windows to extract a local threshold that varies from one window to the other [2]. If we use Otsu's technique to binarize [3], being a threshold selection method from gray-level histograms, we can divide the image into 16*16 blocks, without overlap, overlays the image. We binarize the entire block one after the other with Otsu's method to search for the threshold that minimizes the intra- class variance defined as the weighted sum of variances of the two classes [4].

$$\sigma_{\omega}^{2}(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (1)$$

Where ω_1 is the prospect of the two categories independently detached by a threshold t and σ_1^2indicating the discrepancies of the two classes. Here, the intra-class variance is as same as the maximizing interclass variance. First, we computed the histogram of each intensity value with its probabilities, then we set the initial value of $\omega_i(0)$ and $\mu_i(0)$. For possible thresholds, we then loop and update$\omega_i(t)$ and$\mu_i(t)$. Also, we compute and chose the threshold $t*$ corresponding to the maximum of $\sigma_b^2(t)$. So we can express out binarized image as shown in equation 2.

$$I_B(x,y) = \begin{cases} 0, I(x,y) < t* \\ 1, I(x,y) \geq t* \end{cases} . \quad (2)$$

Let's take *figure 2* to be our supposed captured image. The rectangle in the middle of the image is our proposed target and the surroundings (also known as the background). As we can see, there are different colors in the background and when a target is set to be recognized, it may be difficult to identify the specific target. Human vision can identify three specific colors being white, gray and black. The picture indicates a black road with white lines differentiating the direction of incoming and ongoing vehicles. The sky shows white and gray colors which may be justified by human vision as either cloudy or evening, but a computer may not be able to identify all these so the best thing we can do is to change the values for each of the colors. As we started from 0 to 255.

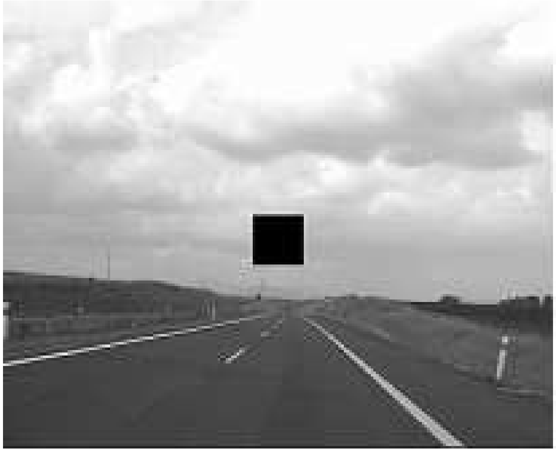


Figure 2: Captured image with supposed target.

To distinguish between the target and the background environment, binarization has to be made. Here in this case the target is given by separating the target to match the color black as close as possible, and then the background as being as white as possible. Doing that, the computer or camera will ignore any other object found in the frame and only recognize the presented target. The binarization threshold chosen for this distinct color was 125. Because the target is purely black, the other colors which are closer to gray than black will all be seen as white.

```
private void Form1_Load(object sender, System.EventArgs e)
{
icImagingControl1.ShowDeviceSettingsDialog();
if( !icImagingControl1.DeviceValid )
{
Close();
return;
}
icImagingControl1.MemoryCurrentGrabberColorformat = TIS.Imaging.ICImagingControlColorformats.ICY800;
icImagingControl1.LiveDisplay = false;
icImagingControl1.LiveStart();
```

This passing source code extract clarifies how binarization was made on for image processing algorithms. The program starts by initiating the built-in dialog in which the following car uses to select a video capture device (.ShowDeviceSettingsDialog). at the final point of the function or to say the end of it, Form1_Load(), the real-time image figure stream from the video capture device is shown,

using .LiveStart. .MemoryCurrentGrabberColorformat= ICY8. Figure 3 confirms that the binarization algorithm obtains an 8 bit graylevel image:

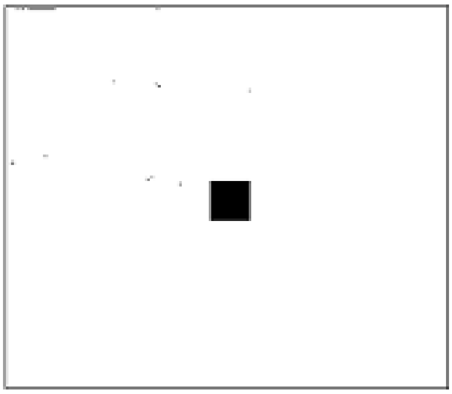


Figure 3: Binarization result

After binarization, we could get a result of two specific colors given in *Figure 3* as a black rectangle for the target and a white background with little back spots being the noise in the camera. In order to enhance the quality of image, which results in the previous step, we first proceed to a conditional dilation with a 3x3 4-connected structuring element [5]. The condition we apply for is that a foreground pixel is added only if its corresponding gray value has no large difference (<5%) with the gray value of the already existing foreground pixel in the 3x3 neighborhood. This condition ensures that character thickness remains the same while several gaps in the character body will diminish. Finally, we proceed to a successive application of shrink and swell filtering in order to further improve the quality of the text regions and preserve stroke connectivity by isolated pixel removal and filling of possible breaks, gaps or holes.

B) Target Recognition

Recognition is the ability for a camera to use its artificial intelligence to detect and identify a specific object given as the target. There are various ways for recognizing objects appearing in cameras. For instance, there is the tendency for an object to be tracked using the heat it produces. In this problem, our task is to recognize or identify a given object or image as a target and then trace it. This section introduces the procedure in which an image can be recognized automatically, which explains how objects are detected by the camera. It also introduces the tools used in making this project a success. In a method of recognizing and tracking an image which corresponds to a real image in at least one sensing zone of at least one sensor for electromagnetic radiation on the basis of images of the sensing zone sensed by means of a sensor, succeeding images are evaluated in succeeding cycles and a search is made for at least one respective object found in a cycle on the basis of a corresponding image in a later cycle to track the given target. At least, one part of an earlier image with respect to an earlier state of the object or a real object corresponding to it, which is determined using a corresponding earlier image in the current image determined in the current cycle or with an image recognized in the current cycle during at least one current cycle on the basis of the results of at least preliminary evaluation of a current image.

To find the target from the current image, we match the target image given to the target that appears in the current image after binarization. The target is positioned on the upper left corner of the current image and slide horizontally, from one line to the next line and from one pixel to the next pixel.

$$S = \sum_{x=0}^{m} \sum_{y=0}^{n} \left| f(x.y) - f'(x.y) \right|^2 \quad (3)$$

The matching procedure is done after each movement. The reference block is first used to match pix by pix horizontally. The movement direction is from the horizontal left to the right from the upper left corner to the upper right corner. Then it moves to the next line horizontally. It goes the same way until it gets to the bottom of the frame. When the target is found, the reference block matches to find equality. If same, we presume our target has been found.

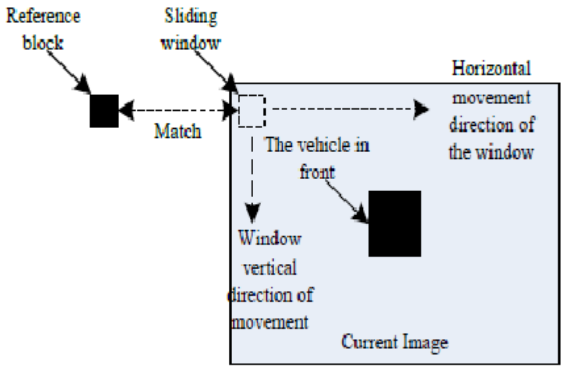


Figure 4: Matching result and current image.

The next thing is to concentrate on the first flag track. If the tracking process yields "*I*", it indicates success and the reference image can be updated before searching for a new area image to match, but supposing

the first flag should not yield "*I*", directly we have to search for a new area image for reference matching. But how can this be done? First, we check or calculate for differences in the average template.

C) Best Match

The best match applies to data files that store records with several real valued keys or attributes. The problem is to find those records in the file most similar to a query record according to some dissimilarity or distance measure. Formally, given a file of N records (each of which is described by k real valued attributes) and a dissimilarity measure D, find the m closest records to a query record (possibly not in the file) with specified attribute values [5]. It should be known that the target and the reference template are identical. The shape and side of the reference template saved in the memory already would be used to match the supposed target. After getting the color specified to be tracked, the next thing needed to calculate is the difference between the luminance of the current area and the reference template, and if the difference is smaller than the threshold we presume to have a set. This leads us to believe that we have found the target and gotten the starting position of the target.

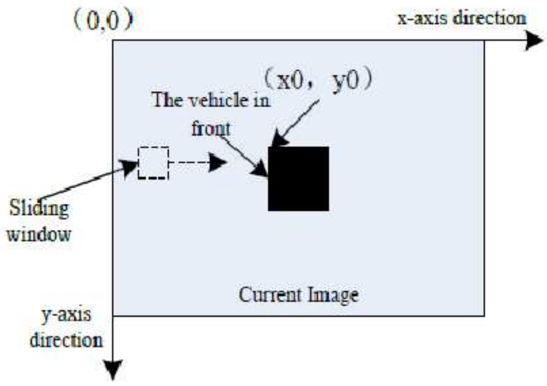


Figure 5: Coordinate(x0,y0) at starting position of target.

D) Calculating the size of the Target.

After getting the starting position of the target, we continue to slide the window along the *x* axis and *y* axis and calculate, so we can get the last matching area

point's coordinates along the *x* axis which is(*x1, y0*) and along the *y* axis which is (*x0, y1*).

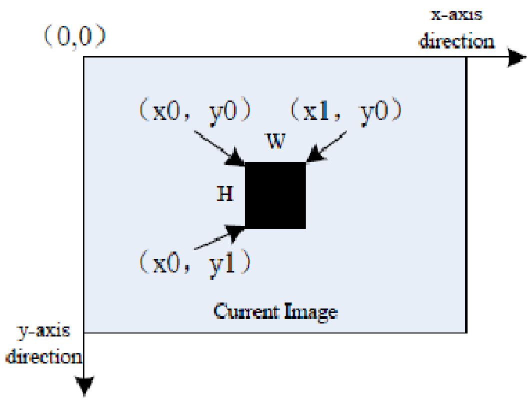


Figure 6: Calculating target size (W=x1-x0, H=y1-y0)

E) Validating Target Size

If the result of recognition is correct, the area in which the starting position is(*x1,y1*)the matching area of the reference template, so we can calculate to validate the result. If the result is not correct, the most likely thing is that the position of the rectangle is not regular. That means the edges of the rectangle are not horizontal or vertical. To solve this problem we can give an offset on the center we have found, which is just a few pixels, then calculate the size of the target and validate the result again. Or we can change the threshold in order to increase the fault tolerance of the system

4 Tracking

After recognizing the target as indicated above, the image data of the domain containing the rectangle will be saved and be taken as the reference template which is used for following tracking. The height and width of the rectangle are the results of learning. And the height is taken as the reference and used for estimating the distance between the two vehicles shown in *figure 7*. Focusing on the tracking algorithm, the first one hundred (*100*) frames which will be taken are classified and are not to be needed. The camera may or may not be stabled as needed. If the camera is not stable and the frames are recorded, we are liable to get the wrong information and the recognition process will fail. When the two cars are positioned and the camera becomes stable, those hundred frames are disposed as the camera may not be stable for correct result. The next frames after that (for instance fifty (**??**) frames) are then saved, for learning or modification. Within the fifty frames

saved, recognition process will begin as in knowing the coordinates of "*x*" and the coordinates of "*y*", the size being the width and height.

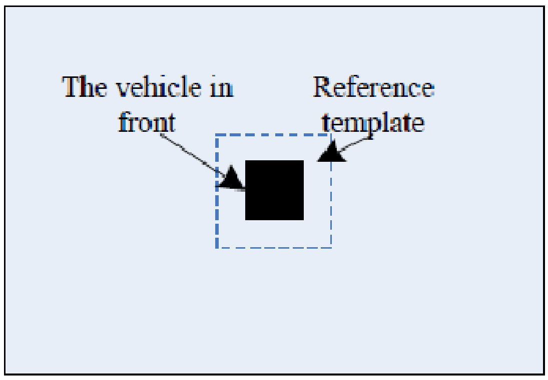


Figure 7: reference template

From the captured digital images, using the image matching algorithm, we then search and find the area that best matches the reference template. Then find the best match area of the target's position in the current image.

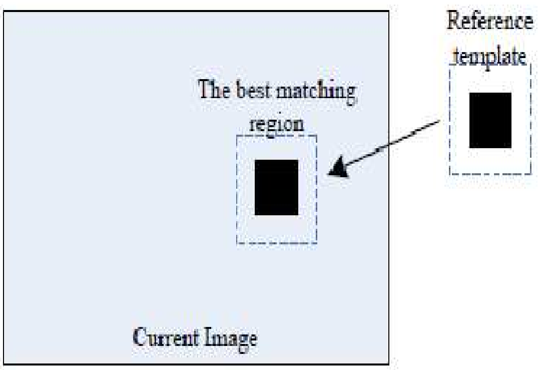


Figure 8: The most matching area is the target's position

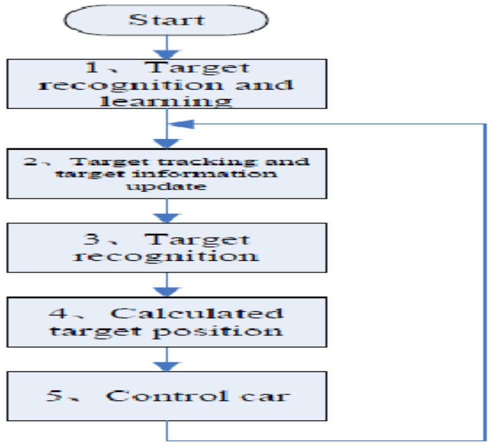


Figure 10: flowchart for tracking targeted car

During the target's forwarding, because of the light's changing and the target's turning and so on, these factors will cause the changing of the target vehicle's image. Keeping using the original reference template, this will let us never find the best matching area or find the wrong area. And it means we have lost the target. So during the tracking, we need update the reference template according to the current situation. And that means the best matching area contained in the current image should take place the original reference template and be taken as the reference template used in the following tracking.

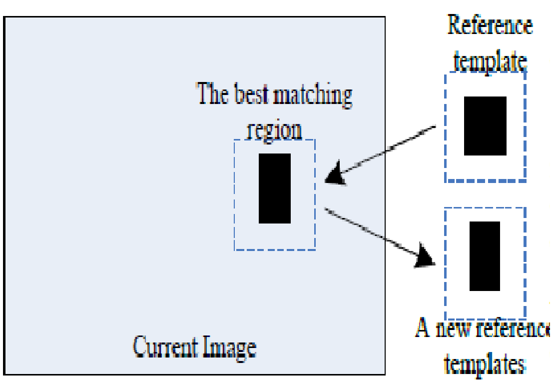


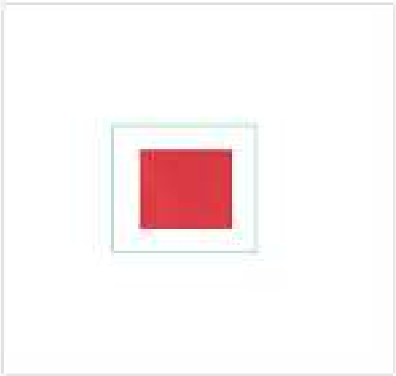
Figure 9: New reference template

Due to a greater change in the target image information, there will arise the need to update the reference template and to calculate best match using the current image as a new reference template region

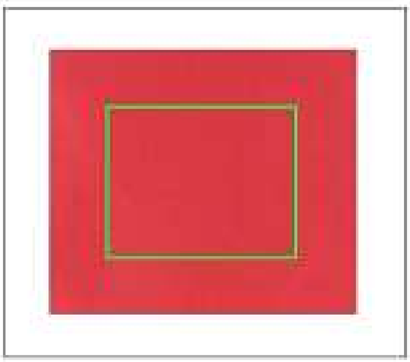
5 Experimental results.

Two radio remote control cars, a virtual board and a video camera were used for the experiment. A red rectangular shaped object which was pasted on the back of one remote control car being the target. The virtual bored was encoded with a program which defines the movement procedure of the target.

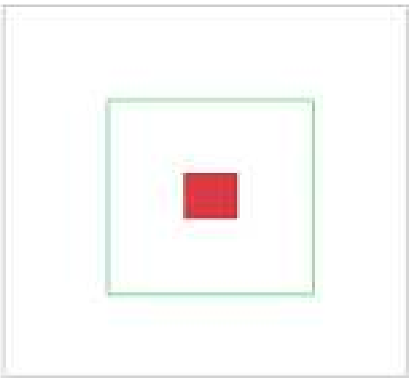
The following car was controlled by the movement of the target. The distances between the two cars were specified. Before tracking, a reference template was saved and used to control the changes with time. The first one hundred (*100*) frames which were taken were discharged as the camera may not be stable for calculating the target position. Figure 11*(a)*,*(b)* and *(c)* shows how the speed was obtained. After video has been captured, the whole recognition process will be done again to obtain the current position of the target. After obtaining the reference template as shown in *Figure 11 (a)*, it's then necessary to control the distance as the target moves. When the target is closer to the camera as indicated in *Figure 11 (b)* below, instruction will be sent to the following car after calculating the difference between the reference template and the new template. This indicates that the car is too close to the target and to avoid colliding must reduce speed.



(a)



(b)



(c)

Figure 11: the location of target after comparing reference template to current template.

If the target appears to be smaller than size saved when calculating the reference template as shown in *Figure 11 (c)*, it shows the target is quite distant and that there following car has to accelerate to the distance at start. This process continues until the target becomes static or blocked by other object. If the target disappears from the current template, the following car stops tracking until the target appears again in the captured video. *Figure 12* and figure *13* below shows the target car and the following car.

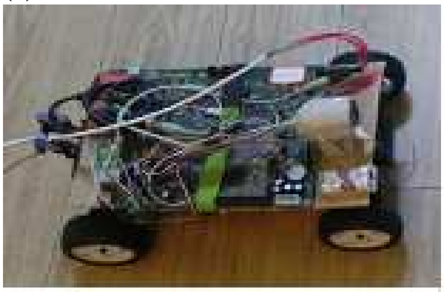
(a)

(b)

Figure 12: Target car with target at the tail

(a)

(b)



(c)

Figure 13: following car.

6 Difficulties in Recognition.

During the tracking procedure, if the information of the target and the background are more, the result will be better. But in the current image it just contains two things, one being a white background and a black target. The information is very little and drab, which means it's not enough. Whether we choose the background or a certain part of the target as the reference template, so there are some influence on the tracking effect. So we should choose a template that is convenient to track. Until now we can envisage that if we choose the top left corner that contains a part of the white background and one corner of the target. And in this chosen certain area, the value of luminance of the pixels can give us the most contrast. Then we can practice more precisely matching during the following tracking procedure.

Acknowlegment

Firstly, I would like to thank the almighty God for given me the strength, knowledge and power to complete this project. Without Him, this work wouldn't have been successful.

I take immense pleasure in thanking **Prof. Shihong Qin,** Dean of Electrical and Information Department of Wuhan Institute of Technology for having permitted me to carry out this project work.

I wish to express my deep sense of gratitude to my Internal Guide**, Doctor Lou,** Numerical Analysis lecturer of Wuhan Institute of Technology for his useful suggestions, a source of inspiration and for his timely guidance in the conduct of this project work which helped me in completing the project work, in time.

And finally,I would like to thank the National Natural Science Foundation of China (NSFC) 60975011/F030403 for partly supporting this project

References

[1] http://en.wikipedia.org/wiki/Transport_in_India
[2] Neural based binarization techniques,Hatem HAMZA*, Eddie SMIGIEL**, Abdel BELAID** LORTA. University Nancy 2. ** LICIA-INSA de S t r a s b o u r g.
[3] Image Binarization using Otsu Method, XuLiang,Monday. April 20th. 2009NLPR-PAL Group. CASIA.
[4] Fast Document Image Binarization Based on an Improved Adaptive ,Otsu's Method and Destination Word Accumulation
Yudong ZHANG , Lenan WU 1553-9105/ Copyright ©2011 Binary Information Press June, 2011
[5] Improved Document Image Binarization by Using a Combination of Multiple Binarization Techniques and Adapted Edge Information **B. Gatos, I. Pratikakis and S.J. Perantonis, 978-1-4244-2175-6/08/$25.00 ©2008 IEEE
[6] AN ALGORITHM FOR FINDING BEST MATCHES ** IN LOGARITHMIC EXPECTED TIME ** SLAC-PUB-1549 (Rev.) STAN-CS-75-482.
[7] Y. Chen, J. Bi, and J. Z. Wang. Miles: multiple-instance learning via embedded instance selection. InPAMI, 2006. 3282
[8] noindent [8]R. Fergus, H. Bernal, Y. Weiss, and A. Torralba. Semantic label sharing for learning with many categories. In ECCV, 2010. 3282
[9] A. Gupta, A. Kembhavi, and L. S. Davis. Observing human-object interactions: using spatial and functional compatibility for recognition. In PAMI, 2009.3282
[10] N. Ikizler-Cinbis and S. Sclaroff. Object, scene and actions: combining multiple features for human action recognition. In ECCV, 2010. 3282.
[11] C. Liu, J. Yuen, A. Torralba, J. Sivic, W. Freeman. SIFT Flow: Dense correspondence across different scenes. Proc. of 10th European Conference on Computer Vision, Oct. 2008.
[12] D. Chekhlov, M. Pupilli, W. Mayol-Cuevas, and A. Calway. Realtime and robust monocular SLAM using predictive multi-resolution descriptors. In 2nd International Symposium on Visual Computing, November 2006.
[13] D.Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In Proc. ISMAR, Cambridge, UK, 2008.
[14] Skrypnyk and D. G. Lowe. Scene Modelling, Recognition and Tracking with Invariant Image Features. In Proc. of International Symposium on Mixed and Augmented Reality (ISMAR), 2004.
[15] Y. Li, A. Goshtasby, and O. Garcia, Detecting and tracking human faces in videos, Proc. ICPR'00 vol. 1, pg. 807-810 (2000).
[16] F. Marqus and V. Vilaplana, Face segmentation and tracking based on connected operators and partition projection, Pattern Recognition 35, pg. 601-614 (2002).
[17] D. Comaniciu and V. Ramesh, Robust detection and tracking of human faces with an active camera, Proc. Visual Surveillance 2000, pg. 11-18 (2000).
[18] H. Roh, S. Kang, and S.-W. Lee, Multiple people tracking using an appearance model based on temporal color, Proc. ICPR'00 vol. 4, pg. 643 -646 (2000).