

CROWD ANALYSIS WITH FISH EYE CAMERA

Huseyin Oguzhan Tevetoglu¹ and Nihan Kahraman²

¹Department of Electronic and Communication Engineering, Yıldız Technical University,
Istanbul, Turkey

¹Netaş Telekomünikasyon A.Ş., Istanbul, Turkey

²Department of Electronic and Communication Engineering, Yıldız Technical University,
Istanbul, Turkey

ABSTRACT

Nowadays crowd analysis, essential factor about decision management of brand strategy, is not a controllable field by individuals. Therefore a technology, software products is needed. In this paper we focused on what we have done about crowd analysis and examination the problem of human detection with fish-eye lenses cameras. In order to identify human density, one of the machine learning algorithm, which is Haar Classification algorithm, is used to distinguish human body under different conditions. First, motion analysis is used to search for meaningful data, and then the desired object is detected by the trained classifier. Significant data has been sent to the end user via socket programming and human density analysis is presented.

KEYWORDS

Fish-Eye, Human Detection, Haar Classification, Crowd Analysis, Machine Learning, Motion Analysis, Density Analysis

1. INTRODUCTION

Trends in video processing and human tracking systems have become popular today. Particularly for security purposes, analysis of data from IP cameras used in various locations has gained importance. However, the camera array, which is designed to completely cover the scene, is disadvantageous both in terms of the number of the images and the size of the data to be processed.

The fish-eye cameras function as both a fixed camera and a moving camera. A fixed and restricted area can be displayed with stationary cameras. Thanks to moving cameras which are manually controllable cameras, areas outside the area where you want to view still images can also be displayed. In this case, it is very likely that the movements in the region to be displayed constantly or continuously are missed. Based on this information, we can say that fish-eye cameras are very advantageous.

Fisheye lenses are created by the do not correction of normal lens defects. These lenses, which we can obtain data in wide angle, provide the creation of hemispherical images. It can be said that firstly it is used in the field of metrology and its usage also increases in security systems today. Depending on the barrel bending error in the lens production, it is possible to obtain two types of images, 360 and 180 degrees.

Resolution performance from a fish-eye camera is not equivalent to a fixed camera. This is why a rectangular 360 degree image is obtained from a rectangular image sensor. It is necessary to make a mathematical explanation; the resolution of 5 megapixels is 2592 x 1944 pixels. The rectangular image that is tried to be obtained is cropped when the circle is fitted because of a diameter of 1944 pixels. The actual resolution from a 5 Megapixel camera is about 3 megapixels, which is the number of pixels left in the field.

A more important disadvantage of fisheye cameras is that the edges of the images are distorted. These distortions are particularly prominent in moving objects and human images.

2. REFERENCE STUDIES

In the literature, there are human detection studies with both fixed cameras and fish-eye cameras. Below is a brief description of the methods used in these studies.

Pfinder [1] is a study on real-time human body tracking. As in almost all work, this work also has background subtraction to get rid of constant background noises. YUV space conversion and normalization operations were applied to avoid the distortions caused by light and similar reasons. Pixels were updated using adaptive filtering at every image and neighbours relationships with nearby pixels were evaluated. As a result, the most suitable pixel center is selected and the similarity of the distribution of the human figure is reached. Pfinder is a work done using a fixed camera.

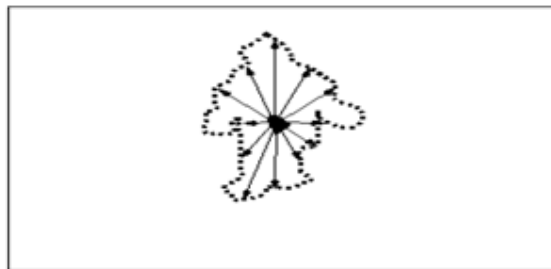


Figure 1. Pfinder morphological growing process

In another study with a fish-eye camera [2], background extraction has been done using Gradient based edge detection technique. In the remaining foreground regions, head and shoulder shapes were searched and tried to match the templates that created earlier. Below is a diagram of working.

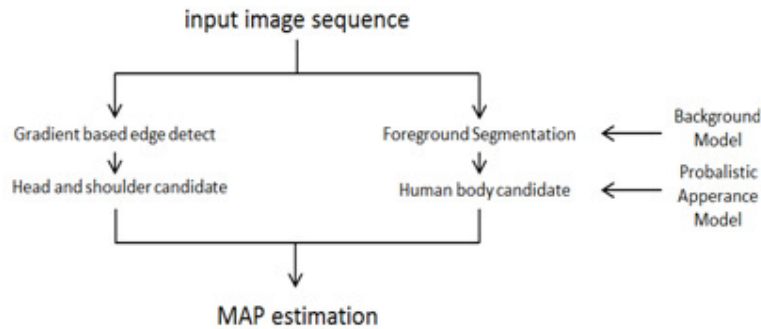


Figure 2. Working diagram[2]

In another work done, it was aimed to prevent occupational accidents by performing human detection with a fish-eye camera placed in front of heavy duty machines [3]. Object distortions from fish-eye cameras used to obtain a wide angle were the biggest challenge they faced. They created databases for object recognition and trained classifiers with high number of positive and negative samples.

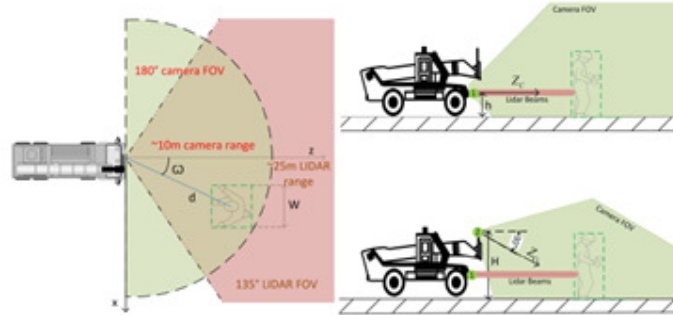


Figure 3. Picture of working[3]

In this study, the extraction of Haar attributes proposed by Viola and Jones and the object classification with enhanced adaboost classifier has been used [4].

3. STEPS OF DESIGN

Object detection methods are used to determine moving objects in video sequences using various methods and to extract motion direction. This process is regarded as the first part of image monitoring systems and has a great impact on system performance.

Finding the moving object by taking the background difference is a method used when the camera that receives the image is fixed and the target object is moving. The background image is determined by the image areas that remain in motion after the system starts running or before the system starts. And then the difference between two background frames is detected. The absolute value of the difference obtained is compared with a predetermined threshold level value. If the difference value is above this threshold value, it is decided that it is a change in the related coordinates. When this method is applied to sequential image frames in video images, the following formula applies.

$$image_i(x, y) - image_{i-1}(x, y) \quad (1)$$

As in other studies, Gaussian operation must be performed on image frames before background subtraction. In this way, the change of light and various distortions of the obstacles are reduced.

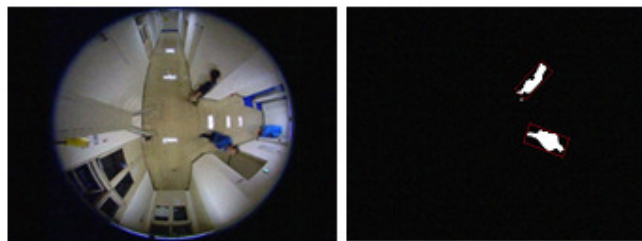


Figure 4. Background subtraction [2]

After background subtraction is done, it should be determined whether moving objects are searched data. Haar classifier will be used for this. Trained classifiers are created by giving the marked samples to the algorithm for the determination of the objects, and by the learning process in the result of this.

When training data set is being prepared; the data in which the specified object is found are marked as positive, and the data not found are marked as negative. In Haar Classifier training, the following frames and the like are used. Frames scan over positive data to detect dark / bright differences and create target values where the differences between black and white regions are obvious.

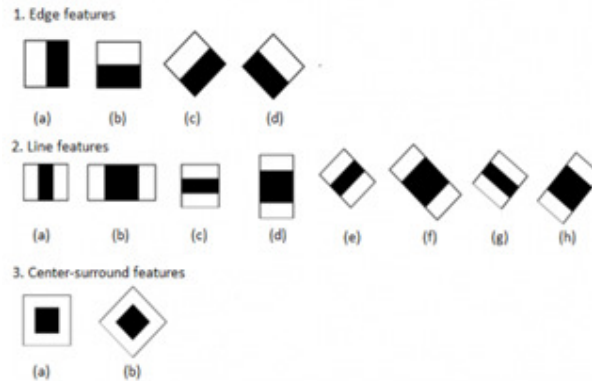


Figure 5. Haar weak classifier

During the detection of the object being searched, the small classifiers specified in the training are scanned on the frame and fixed in the regions that match the target values. In an input frame, many small frames may provide target values for different regions, but since the object size to be searched is approximately known, an object is detected in the area where a range of frames are densely mapped.

In educational classrooms, the data used in the training set must be consistent with each other. Otherwise the faulty rate will be high. In the study performed, the human silhouette was seen in different shapes in the different coordinate sections because of the fish eye camera. To avoid this, the camera is positioned at a height of 5 meters. This time, as the pixel size of the human silhouette decreased, the number of detected features decreased and the educational success was adversely affected.



Figure 6. Dimension and shape differences in image

As a solution, training was generating with different data sets for different coordinate zones of the same frame. Training was carried out using a total of 5000 positive, 3000 negative data sets.



Figure 7. Training regions

The coordinates of the detected people are sent to the client with the generated socket programming code. The UDP protocol is used because high data transfer is a matter of concern. The client presents the data it receives in a visual form that end user can understand. The following illustration shows the example.



Figure 8. Crowded heat map

The pseudo code and flow diagram of the generated algorithm is presented below.

Table 1. Pseudo Code.

1	Start
2	Get Frame
3	Apply Gaussian Filter
4	Background Subtraction
5	Divide Frame to 3 Different Regions
6	Clear Remaining Remnants
7	Search For Moving Objects
8	Give Object to Classifier
9	Send Meaningful Data Via Socket
10	View Heat Map With Web Browser

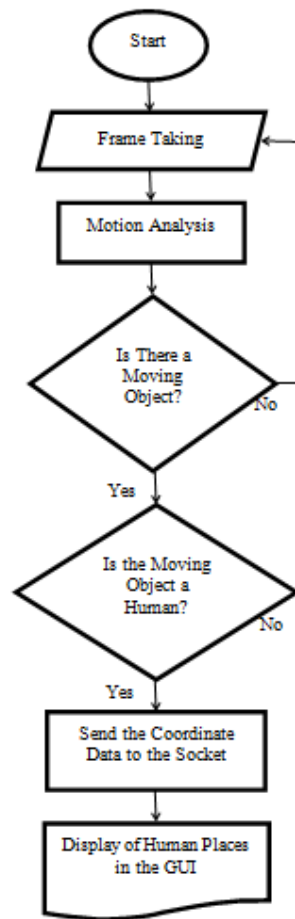


Figure 9. Algorithm Flow Diagram

4. RESULT

Within the scope of the study, a solution is proposed for an affordable human detection and intensity map of the area that does not contain blind spots and covers the entire area.

In the present case, classifying moving objects reduces the processing power. However, when the detected object stops moving, the previous position continues to be marked as positive. If there is no movement within 3 seconds, the region is given a reclassifier and it is confirmed that the detected object is human. Algorithm and classifier optimization studies are continuing to reduce the specified time threshold.

It was mentioned above that the training was done with different data sets according to the separated regions in the image frames taken from the camera. However, it was observed that the detection success rate decreased in the transition locations between the two different regions. The work is ongoing to resolve this problem.

The height of the position of the camera is affected by the algorithm performance ratio. With the created front ace, the user is able to enter the integrated height of the camera parametrically. It is planned to increase the performance ratio by changing the algorithm and classifier values according to the entered value.

REFERENCES

- [1] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell and Alex Paul Pentland., "Pfinder: Real-Time Tracking of the Human Body ", IEEE Trans.on Pattern Analysis and Machine Intelligence. Vol 19, No 7, July 1999.
- [2] Mamoru Satio, Katsuhisa Kitaguchi and Gun Kimura, Masafumi Hashimoto, Human Detection from Fish-eye Image by Bayesian Combination of Probabilistic Appearance Models, IEEE Systems Man and Cybernetics, Oct. 2010.
- [3] Vincent Fremont, Manh Tuan Bui, Djamel Boukerroui and Pierrick Letort., "Vision-Based People Detection System for Heavy Machine Applications" Jan. 2016
- [4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," presented at IEEE Conference on Computer Vision and Pattern Recognition, 2001

AUTHORS

Nihan Kahraman is Assist. Prof. Dr. at Yıldız Technical University Electronic and Communication Engineering Department. She has completed many projects throughout her academic career. Her interests are listed below.

- Analog and digital IC design
- Microelectronics
- Neural networks
- Electronic design automation
- Image processing



Oguzhan Tevetoglu is a software architecture designer at Netas and also graduate student at YTU. He has taken many roles in embedded system projects. His interests are listed below.

- Image processing
- Embedded Systems
- Neural networks
- Computer network

