SMART VIDEO SURVEILLANCE FOR BACKGROUND ANALYSIS

Andrada Cirneanu

Phd, Military Technical Academy "FERDINAND I", andrada.cirneanu@mta.ro

Abstract: *This paper presents a method for motion detection and form recognition from a video stream generated by a video surveillance system. The purpose of this paper is to propose an algorithm that differentiate a person form from other types of forms. Motion detection is realized based on the comparison of 2 consecutive frames extracted at a specific time interval from the real-time video stream and based on the resulted comparison, a binarization algorithm and a region growing algorithm are used to extract all independent regions in case of*

Keywords*: person detection, motion detection, surveillance.*

1. Introduction

Nowadays computer vision is an ultramodern field that has been fully developed in recent years. Applications developed in this domain are based on a real-time analysis of images provided by one or more video camera. The possibilities of this technology are almost unlimited and reach different areas of the industry. The variety of applications in which images are used on the computer is expanding gradually as users discover unquestionable advantages of this technology. The main areas where computerized imaging can be used are control application, automation, recognition or accurate measurement.

Before analyzing the images, the most important thing is to capture the image. This process is called the acquisition of the image..

The image acquisition process is carried out in several stages. Initially, the subject of interest reflects an energy. Then an optical system concentrates that reflected energy and finally a sensor measures that amount of energy (Fig.1).

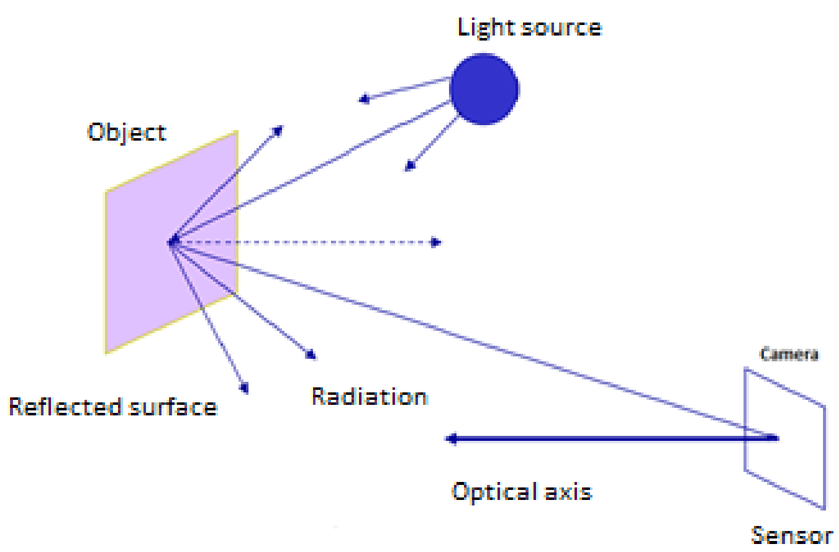

Fig. Image acquisition

The field of motion detection cameras is monopolized by two different types of technologies depending on the type of image sensor used. These two types are CCD (charge-coupled device) and CMOS (complementary metal-oxide semiconductor). Inside both types of cameras, motion detection is realized by comparing each captured image with the previous one and the cameras will interpret any noticeable changes between consecutive images as motion. Two important settings characterize almost all surveillance cameras: sensitivity and percentage. Sensitivity represents a contrast setting that determines when a change between consecutive images represents motion in the monitored area. On the other hand, percentage is set in the case of detecting motion in a sub-area of the monitored area.

Even though these motion detection cameras are an affordable option for a regular security system, there is a potential need for security systems that are also based on surveillance cameras, but the motion detection is based on an auxiliary image processing system that can also accomplish other processing stages if required. The reason for implementing such systems is represented by the scope of this paper which is motion detection in a monitored area and forms differentiation like humans from animals.

Despite the fact that great progress has been made in recent years, people detection continues to be a challenging task in the field of computer vision.

To this purpose, the authors from [1] propose a methodology which is focused on people detection and recognition. The study consists in an efficient image descriptor that is inspired from the mathematical statistics of the Brownian motion. The obtained results proved to be promising for pedestrian classification and person identification tasks. In [2], the authors present an approach for pedestrian detection based on a combination of Haar and LBP features in conjunction with AdaBoost machine learning algorithm for the learning process and Meanshift algorithm for performance improvement.

Authors in [3] present a method for human detection using a cascade-of-rejectors approach integrated with Histogram of Oriented Gradients (HoG) features. First, the detection system identifies the most appropriate set of blocks and then uses a rejection cascade to speed up the computation. The algorithm proves to have a high accuracy level. In [4], the authors came up with a solution of people detection based on RGB-D sensor. The algorithm browses through two steps: at the beginning twoRGB and depth feature fusing strategies are proposed and compared and then an improved non-maximum suppression algorithm is proposed to further boost the performance of detection without increasing the time consumption.

Paper [5] comes with an approach of people detection from depth maps produced by sensors mounted in a zenithal position. The algorithm finds the object of interest in the scene by adopting a dynamic background modelling strategy after which the noise from the foreground image is filtered out in order to determine the position of persons. The experimental results proved the efficiency of the algorithm.

The authors in [6] follow an approach based on machine learning features. This video person tracking algorithm computes a simple motion detection framework through motion blobs and has an advantage of detecting people who are stationary for a period of time.

Our methodology for people detection and differentiation from other types of forms from a video stream generated by a video surveillance system considers the development of an algorithm that detects motion by comparing two consecutive frames extracted at a specific time intervalfrom a real-time video stream. Further, on the resulted comparison,a binarization is applied and a region growing algorithm is used to extract all independent regions from the resulted comparison. Over the extracted regions, a series of features like area of the detected region, dimensions of a box that would include the detected region or the Euler number of the detected region are computed to differentiate people form from other detected forms.

2. Methodology

The video surveillance system performs real-time monitoring of a perimeter and generates an alarm only when persons are detected in the supervised area.The system must trigger the alarm to notify the supervisors of the monitored area. Thus, when the alarm is triggered, the proposed algorithm must present in a clearly way the scene in which a person was detected.

The proposed methodology for motion detection and people recognition in a specific area supervised by a surveillance camera is based on frames extraction at specific time interval from the real-time video stream and a series of processing steps that are applied over the extracted frames.

The proposed system goes through two stages.In the first stage, the system detects motion in a certain scene and identifies the form of that object (whether person or something else).In the second stage, the system triggers an alarm if the identified form is a person form based on proposed characteristics.

Motion detection from the video stream is realized by making a subtraction on grayscale level between two consecutive frames extracted at a specific time interval (one second interval was used in the current algorithm). Assuming frame F_1 is extracted at t_1 timestamp and frame F_2 is extracted at $t_2=t_1+1$timestamp, a difference frame is obtained, like in (1):

$$F_D = F_1 - F_2 \qquad (1)$$

in which the absolute value of each pixel subtraction is memorized.

Even though this subtraction of consecutive frames will expose all pixels for which there is a difference in value from one frame to the other and thus, a potential motion in these pixels, a series of processing tasks for the difference frame F_D are required to eliminate the pixels that do not belong to a significant region from the supervised area.

First processing task is represented by the global binarization of the difference frame F_D. To compensate the illumination variations caused by weather conditions that can change instantaneous and a possible inadequate contrast between background and the detected moving regions and due to the fact that a fast binarization algorithm is required, the Otsu binarization method [8] is used in the proposed algorithm.

Otsu binarization method performs histogram form-based image thresholding to reduce a gray-level image to a binary image. This method takes into consideration two classes of pixels (e.g., foreground and background) and it calculates an optimum threshold *t* value that would separate those 2 classes in order to obtain a minimal combined spread (intra-class variance). Thus, the required threshold *t* is the one that minimizes the intra-class variance defined in (2) as the weighted sum of variances of the two classes [7].

$$\sigma_w^2(t) = q_1(t)\,\sigma_1^2(t) + q_2(t)\,\sigma_2^2(t) \qquad (2)$$

In (2), q_1 and $q_2$represent the class probabilities of different gray-level pixels and they are computed based on image histogram *P* as it follows:

$$q_1(t) = \sum_{i=0}^{t} P(i) \qquad q_2(t) = \sum_{i=t+1}^{255} P(i) \qquad (3)$$

Further, σ_1 and σ_2 represent the individual class variances for the 2 defined classes (e.g., foreground and background) [9] and they are computed as in (4).

$$\sigma_1(t) = \sum_{i=0}^{t} \left(i - \mu_1(t)\right)^2 \frac{P(i)}{q_1(t)} \qquad \sigma_2(t) = \sum_{i=t+1}^{255} \left(i - \mu_2(t)\right)^2 \frac{P(i)}{q_2(t)} \qquad (4)$$

where μ_1 and μ_2 represent the class means for each class (5).

$$\mu_1(t) = \sum_{i=0}^{t} \frac{i \cdot P(i)}{q_1(t)} \qquad \mu_2(t) = \sum_{i=t+1}^{255} \frac{i \cdot P(i)}{q_2(t)} \qquad (5)$$

After computing the optimum threshold *t* based on Otsu technique, the difference frame F_D is binarized at pixel level based on the threshold *t* and the binary difference frame F_{Dbin} is obtained, as in (6) where each pixel of F_D is compared with the Otsu threshold.

$$F_{Dbin}(i,j) = \begin{cases} 0, F_D(i,j) < t \\ 1, F_D(i,j) \geq t \end{cases} \quad (6)$$

The second processing task is the labelling of all independent regions from F_{Dbin} with the aid of a region growing algorithm[9] [10]. The input parameter of the region growing algorithm is represented by F_{Dbin}image and the output is a similar image $R_{Dbin}$with the same dimension in which each white pixel from F_{Dbin} has a value corresponding to the number of the independent region to which it belongs.

In order to compute R_{Dbin} image, the region growing algorithm starts from the top left corner pixel of F_{Dbin} and it iterates on line and column each pixel of F_{Dbin}until it finds the first white pixel $P_{i,j}$positioned at row *i* and column *j*. This means that the first element of the first independent region was found and the value of this element is set to 1 in R_{Dbin}image.

After setting the first pixel $P_{i,j}$ of the first independent region, the algorithm must check for other pixels that might belong to this region. Thus, it checks each white neighbour of $P_{i,j}$ if it is set in R_{Dbin}image and if not, the algorithm sets that the value of those specific neighbours to 1 in R_{Dbin}image. This process of neighbours check is done repeatedly until there are no white neighbours available.

Further, after labelling the elements of the first regionR_1, the region growing algorithm will iterate all the other pixels of $F_{Dbin}$and it applies the same steps as for region R_1.

Finally, a set of independent regions *R* is obtained as in (7).

$$R = \{R_1, R_2, \dots, R_n\} \quad (7)$$

where R_k,k=1,...,nrepresents a collection of white pixels from F_{Dbin}image that are labelled with *k* value in R_{Dbin} image and *n* represents the number of independent regions labelled in R_{Dbin} image. Consequently, all R_k,k=1,...,n regions are disjoint as in (8) meaning that any pixel from F_{Dbin}image is found in only regionR_k ,*k=1,...,n* extracted by the region growing algorithm.

$$R_{k_1} \cap R_{k_2} = \emptyset with k_1 \neq k_2;\ k_1 = 1, \dots, n; k_2 = 1, \dots, n; \quad (8)$$

After identifying all independent regions from F_{Dbin}, the proposed algorithm calculates the following characteristics for each region:

- the dimensions of the smallest rectangle that contains the region
- area of the region
- the Euler number of the region.

These geometric features create an identity card of each region in order to uniquely differentiate the regions.

The smallest rectangle that contains region $R_k$is defined as a rectangle type sub-image of F_{Dbin} image that includes all pixels of region R_k. For this characteristic, 2 properties are of interest: the width W_{rec} and the height H_{rec} of the rectangle.

The area of a region R_k is computed as in (10) where *p* represents the number of pixels that belong to R_k region.

$$A_{R_k} = \sum_{l=1}^{p} l \quad (9)$$

On the other hand, the Euler number [11] which establishes the homogeneity of the image is computed as below since there is one compact object present in R_k region.

$$E_{R_k} = 1 - H_{R_k} \quad (10)$$

where H_{R_k} represents the number of holes from the R_k region.

The decision part of the proposed algorithm is based on the 3geometric characteristics and it establishes that a region R_k is a person if:

- ✓ $H_{rec} > W_{rec}$
- ✓ $A_{R_k} > 4000$
- ✓ $E_{R_k} > -40$

The alert module is the process of alerting the operator that the system detects a person who enters the perimeter which is surveillance with the camera. Practically, the alarm should be triggered if the form that was detected by the proposed algorithm is classified as being a person. The alert process can be replayed as a warning message or by marking the region of the detected form.

3. Experimental results

The proposed paper for the person detection that enter in a specific perimeter is based on a TCP/IP type network with specific features for acquiring, viewing and storing images obtained with a surveillance camera.

The equipment includes as it is shown in Figure 2:

- 1 IP camera
- SwitchPoE
- PC
- Power source 24 V
- Ethernetcable

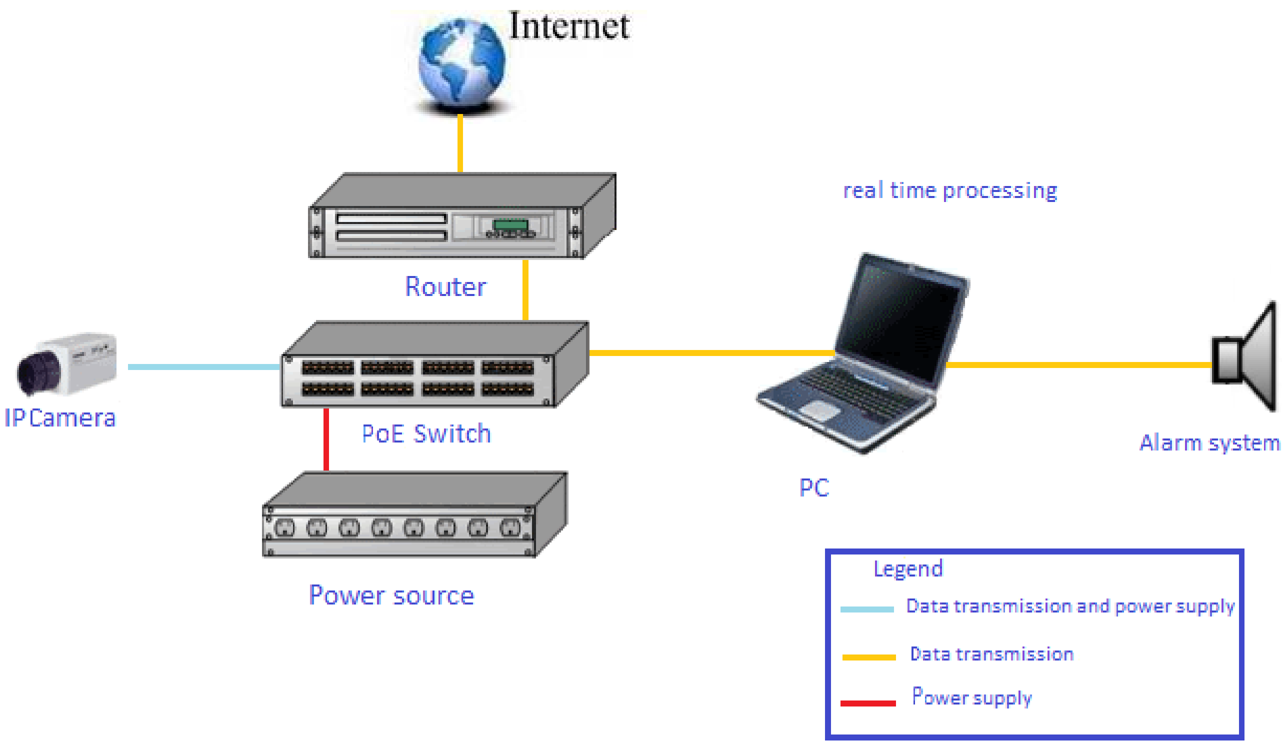

Fig. 2. Proposed System Architecture

The proposed algorithm was tested on a series of monitored areas.

In Figure 3, there are 2 frames extracted from an area where no person is present, but motion is still detected because of the weather conditions. However, after achieving the difference between the two frames extracted at some random timestamps from the video stream, motion was detected, andthe proposed algorithm computed 14 regions suitable for analysis. By computing and applying the 3 proposed characteristics for each region, all the regions are eliminated, meaning that none of regions contains a person form and in this case, the alarm is not triggered.

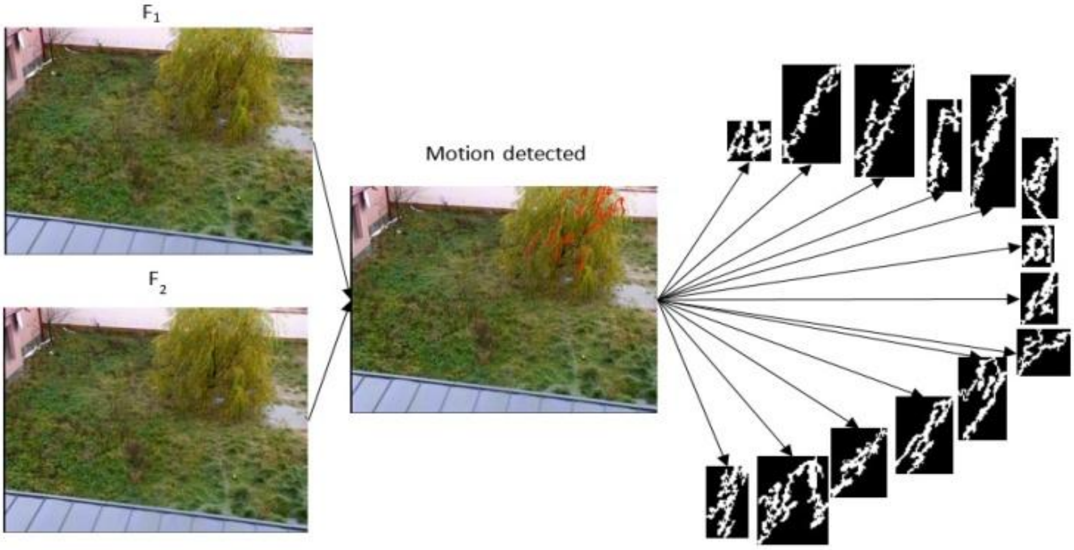

Fig. 3. Motion detection in a non-person area

Several results of the proposed methodology for the people detection in 2 different areas are presented in figures 4 and 5. In both cases, 2 frames were extracted at consecutive timestamps from the video stream of the 2 areas in which a person generates the detection of motion.

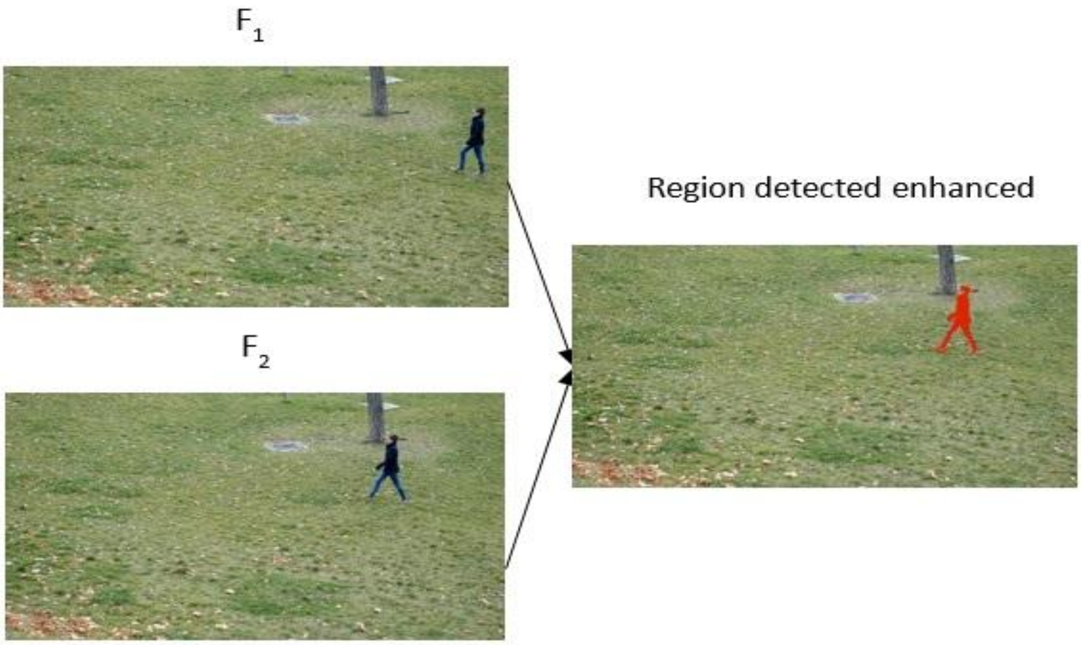

Fig. 4. Motion Detection in a person area(1)

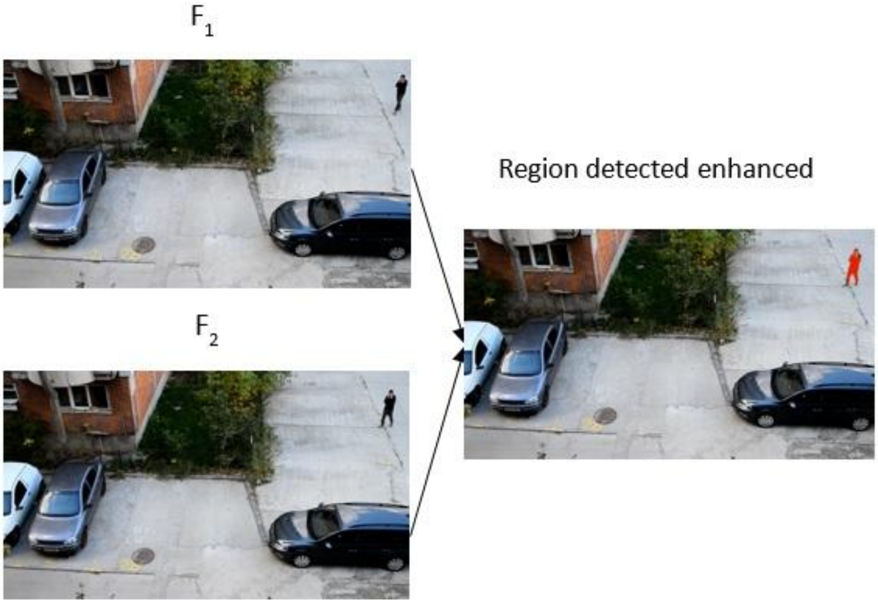

Fig. 5. Motion Detection in a person area(2)

6. Conclusions

The algorithm developed in this paper demonstrated reliable results in people detection using a video surveillance system and it also provided a satisfactory delimitation in form recognition (if is a person who entered in that perimeter or not). The algorithm was used on different environments (areas with a lot of vegetation or parking zones) and every time the results were correctly classified.

References:

[1] Zhengyan Zhang, Jingjie Yan, Guanming Lu, Haibo Li, Ning Sun, Qi Ge, "Multiscale Adaptive Local Directional Texture Pattern for Facial Expression Recognition", KSII Transactions on Internet & Information Systems, Vol. 11, Issue 9, pp. 4549-4566, 2017.

[2] Wilbert G. Aguilar, Marco A. Luna, Julio F. Moya, Vanessa Abad, Humberto Parra, Hugo Ruiz, "Pedestrian Detection for UAVs Using Cascade Classifiers with Meanshift", Semantic Computing (ICSC), pp. 509 – 514, 2017.

[3] Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, S. Avidan, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients", Computer Vision and Pattern Recognition, pp. 1491 – 1498, 2006.

[4] Qi He, Kuixiang Liu, Lei Qu, "Efficient and robust indoor people detection based on RGB-D camera", Chinese Automation Congress (CAC), pp. 1063-1068, 2017.

[5] Vincenzo Carletti, Luca Del Pizzo, Gennaro Percannella, Mario Vento, "An efficient and effective method for people detection from top-view depth cameras", Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6, 2017.

[6] Imran Ahmed, Awais Ahmad, Francesco Piccialli, Arun Kumar Sangaiah, Gwanggil Jeon, "A robust features based person tracker for overhead views in industrial environment", IEEE Internet of Things Journal, Vol. PP, Issue: 99, 2017.

[7] Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, Issue 1,pp. 62 – 66, 1979.

[8] Ningbo Zhu, Gang Wang, Gaobo Yang, Weiming Dai , "A fast 2d Otsu thresholding algorithm based on improved histogram", CCPR, pp. 1-5,2009.

[9] N. Ikonomatakis, K.N. Plataniotis, M. Zervakis, A.N. Venetsanopoulos, "Region growing and region merging image segmentation", Digital Signal Processing Proceedings, Vol. 1, pp. 299 – 302, 1997.

[10] Weihong Cui, Zequn Guan, Zhiyi Zhang, "An Improved Region Growing Algorithm for Image Segmentation", Computer Science and Software Engineering, Vol. 6, pp. 93 – 96, 2008.

William K. Pratt, "Digital Image Processing", Second Edition, John Wiley & Sons, New York, 1991.