Detection and Tracking of People's group

Jorge Paullada, Ana Herrera, Hugo Jiménez, Maribel Bueno, Consuelo Frías, Felipe Estrada

Abstract— This paper presents a method for monitoring human activities in real time on video sequences. The monitoring process consists in detecting people in video sequences by motion detection using a background model, where each pixel is modeled as a mixture of Gaussian. The detection of people is performed by skeletonization of the movement objects. The segmentation of groups of people is obtained by analyzing the minimum vertical histogram of the moving objects. The results show that the scheme is suitable to detect activity in specific scenarios, even when lighting changes have on the scene or the video quality is low it acquired.

Index Terms—Detection, segmentation, background, tracking.

I. INTRODUCTION

Monitoring and surveillance systems are a useful tool for taking decisions. Depending on the events on a video sequence, some of the activities that can occur might be of interest to us or not. Lately, the area of computer assisted video monitoring and surveillance has been growing, due to the fact that people can't be monitoring video streams for long periods of time non-stop because of tiredness or other physiological needs. These computer assisted tools for video monitoring, can identify events that are of interest in a video stream, allowing people to focus on other activities and avoiding problems related with fatigue or other problems that can arise from watching video streams for prolonged time.

In recent research, some algorithms have already been developed [1-3]. One of the most widely used algorithms was proposed by Kanade [4] where the author developed vision systems based on image differences. Subsequently, other authors such as Zhang et al. [5], have proposed methods of detection based on extraction of image features, which are grouped together and assigned a meaning according to the

Manuscript received Oct, 2014.
Jorge Mateos, *Instituto Tecnológico de Querétaro, Querétaro, México*
Ana Herrera Navarro*, Instituto Tecnológico de Querétaro, Querétaro, México.*
Hugo Jimenez*, Centro de Investigación y Desarrollo Industrial, Querétaro, México.*
Maribel Bueno, *Instituto Tecnológico de Querétaro, Querétaro, México.*
Consuelo Frías, *Instituto Tecnológico de Querétaro, Querétaro, México.*
Felipe Estrada*, Instituto Tecnológico de Querétaro, Querétaro, México.*

observed dynamic. Additionally, Ivanov et al. [6] proposed an approach based on logic level modeling of states and transitions between them, where the expressions LP0 represent the actions at the scene. Oliver et al. [7] and Buxton [8] proposed an approach based on probabilistic connectivist models where a set of states (more movement areas) define the probabilistic relations of the transition object due to its motion in the scene.

In this paper, a model for recognition and monitoring of activities of people is presented. This work proposes a method to identify moving objects and people through skeletonization.

This paper is organized as follows. In an initial part, segmentation of moving objects are discussed. Moving objects is detected by a model of background subtraction, where each pixel is modeled as a mixture of Gaussian [9] of the observable intensities. Tolerance to lighting changes in the scene is achieved by incorporating a set of constraints to the process of estimating the parameters of the Gaussian analysis and spatial connectivity through a morphological filter. In section 2, individual persons or groups are detected. The detection of people is performed by skeletonization of the movement objects. The segmentation of groups of people is obtained by analyzing the minimum vertical histogram of the moving objects. The skeletonization process is performed by analyzing the peaks of the distance from the center of the moving object to its edges. In a third stage is the analysis of the activities through temporal and spatial information about the person. Finally, activities are identified through the evolution of the skeleton of the people detected in the scene.

II. BACKGROUND SEGMENTATION

Given a sequence of images $I = \{I_1, I_2, I_3 \dots . I_K\}$ stationary objects maintain constant levels of intensities. Then for an x arbitrary position, the probability density function (pdf) of the color intensity for all images of the sequence $f(v) \backsim v + N$ is modeled as a random variable, where v is the color depth and N is the measurement noise with mean 0. Then, it is assumed that the noise is a normal distribution, the distribution of each point in the image can be modeled by a Gaussian. In general, when there is more than one value then observed intensity has a mixture of Gaussian. $I(\mathbf{x}) \backsim \sum_{i=1}^{n} \alpha_i G(v_i; \mu_i, \sigma_i) = 1$, where n is the number of Gaussian. $G(v_i; \mu_i, \sigma_i)$ represents the ith Gaussian for stability and value v_i . Parameters μ_i, σ_i and α_i, is a value that is associated with the probability of observing the ith Gaussian. Consequently, given an arbitrary image $I_i(\mathbf{x})$and a x position, intensity belongs to the background if it corresponds to any of the estimated Gaussian.

$$B(\mathbf{x}) = \begin{cases} 1 & \exists_{i\in[1,\dots,n]}\, d\big(I(\mathbf{x}), G(v_i;\mu_i,\sigma_i)\big) \le \lambda_{thr} \\ 0 & \text{en otro caso} \end{cases} \qquad (1)$$

where $d\big(x, G(v_i;\mu_i,\sigma_i)\big) = \frac{(x-v_i)^2}{\sigma_i^2}$ is the normalized distance between the x measurement and the Gaussian particular. The λ_{thr} belonging threshold is a function of likelihood that the sample belongs to the observed Gaussian and typically it takes a value of 2, which represents certainty of 0.95. Finally, the process of estimating the parameters can be estimated by several ways [10-11]. However, when an estimate in real time is required, this task is considered as a problem of optimization. A computationally efficient approximation is the *EM* algorithm [12]. This method iterative update the mean and standard deviation for each time t as follows

$$\mu_{i+1} = \rho\mu_{t+1} + (1-\rho)x; \quad \text{and} \qquad (2)$$
$$\sigma_{i+1}^2 = \rho\sigma_{t+1}^2 + (1-\rho)(x-\mu_{t+1})^2;$$

Where ρ is a convergence constant that define how fast new objects are similar to background. Finally, when a scene is constant for intervals of time, the Gaussian changes slightly and the dispersion of the Gaussian (see Fig. 1) is minimum. However this behavior is not ideal when sudden changes in lighting conditions such as reflections, external lights, etc. Hence, that the estimate of the standard deviation is given by

$$\sigma_{i+1}^2 = \begin{cases} \sigma^2 \ge \sigma_{\min}^2 & \rho\sigma_{t+1}^2 + (1-\rho)(x-\mu_{t+1})^2 \\ \text{other case} & \sigma_{i+1}^2 = \sigma_{min}^2 \end{cases} \qquad (3)$$

where σ_{min} is the minimum standard deviation that tolerates the luminance changes in the scene.

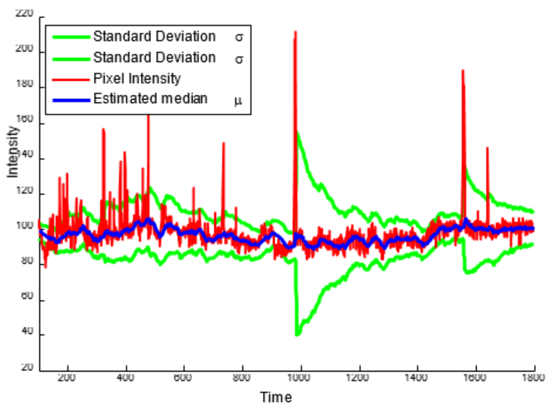


Fig. 1. Convergence and changes in the mean and standard deviation. When the intensity is constant, the dispersion is minimal.

Finally, when the Eq. 1 is applied, a morphological opening [13] is applied to remove areas, which are not related to the image. So the motion map for a particular image is given by $M = \delta_\lambda \varepsilon_\lambda (B(X))$, where ε and δ are the erosion and dilation, respectively, and λ a structuring element. To illustrate the process, Fig. 2 shows an example of the detection of images is presented. The size and shape of the structuring element can discard those regions that represent noise or very small objects that are not relevant to this work.

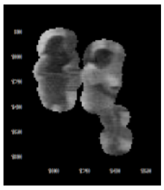


Fig. 2. Example of detection of objects in movement.

III. TRACKING OBJECTS

The skeletonization process consists in detecting the maximum of the histogram distances of the contour detected objects. To determine objects in a frame, the concept of neighboring points is defined.

Two points $I(x_i)$ and $I(x_j)$ are related if and only if $I(\mathbf{x}_1) = I(\mathbf{x}_\mathrm{J}) = 1$ and $|\mathbf{x}_1 - \mathbf{x}_\mathrm{J}| < c_r$, where c_r is the radius of connectivity. Consequently, *path* is the set of connected points that binds two consecutive non-related points, where the *path length* is the number of points needed to unite the two points. Finally, a related object o_i is defined as the set of points such that there exists a route r between all pairs of points that make up the object.

The center of the border is calculated as:

$$\bar{\mathbf{x}} = \frac{1}{|o_i|}\Sigma_{\mathbf{x}_i \in o_i}\mathbf{x}_i \qquad (4)$$

Therefore the distance from the centroid to border in a direction θ is given by:

$$d(\bar{\mathbf{x}},\theta) = \left\{ \begin{matrix} \max(|\bar{x} - x^m|)\ \text{tal que}\ x^m \in o_i\ \ and \\ \theta = \operatorname{atan}\left(\frac{\bar{x}_2 - x_2^m}{\bar{x}_1 - x_1^m}\right) \end{matrix} \right\} \qquad (5)$$

where $d(\bar{\mathbf{x}},\theta)$ represents the maximum distance from the centroid to the farthest element in the θ direction. Finally, the histogram of the maximum distance from a point to a resolution of 2π / k is defined as $h_d(\bar{\mathbf{x}}, o_i) = [h_1,\dots,h_k]$ where $h_i = d(\bar{\mathbf{x}}, \frac{2\pi}{i})$. This histogram represents the distance from the centroid to the edge of the object. The histogram of distances provides information on the object shape. Now, the skeletonization process consist in determining the possible of extremities of found objects.

An efficient approach is to detect the maximum of the histogram of distances on the edge of the object.

To determine the local maxima of the histogram can be applied the maxima transform, as shown in the following: $E_{max}(h_d) = R_{max}^{\gamma}(H_{max}(h_d))$; where $R_{max} = M - R_I^{\gamma}(h_d\text{-}1)$, $H_{max} = R_{max}(I-1)$ and h_d is distance histogram, H_{max} the maxima transform, R_{max} the local maxima and $R_I^{\gamma}(I-1)$ [13] the geodesic dilation of distance histogram of the image I. Finally $h_d(\theta_i)$ positions for which has a maximum at θ_i, represent the possible ends of the object.

People Detection

The process of detecting people, consist of to analyze and to classify the different skeletons of the detected objects. The detection process involves comparing the distribution of the skeletons of people found and interior angles. When people appear walking a silhouette of movement occurs in most cases, as any of the structures shown in Fig. 3. It can be seen the main forms of the skeletons people. Note, the silhouettes of the people present in most cases have this four basic configurations.

Fig. 3 Typical structure of the skeletons that represent people.

Identifying people is made possible through detecting the configuration that identifies the lower extremities and the trunk of the person. The trunk of the person is always represented by an axis that is approximately at right angles (Fig. 3 d). The lower extremities by a shaft (Fig. 3 d) or two axes (Fig. 3 c) are represented. In the first case, the shaft has three right angles. In the second case, each shaft is in the 3rd and 4th quadrant and each have a maximum opening of the right angle. The remaining axes (arms) provide information on the position of the person. So if a skeleton has at most 5-axis and two or three of them (as applicable), are by their angle axes that represent a head and lower extremities which are assumed to be of a person. Formally it is expressed as,

$$(\{\theta_1, \dots, \theta_n\}) = \begin{cases} 1 & \exists_{\theta_i}\theta_i \approx \frac{\pi}{2} \wedge (\exists_{\theta_j}\theta_j \approx \frac{3\pi}{2} \vee \exists_{\theta_j,\theta_k}\theta_j \in [\pi, \frac{3}{2}\pi] \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where classification function takes a 1 value if a person is detected or 0 value in otherwise for all angles θ_i representing maximum.

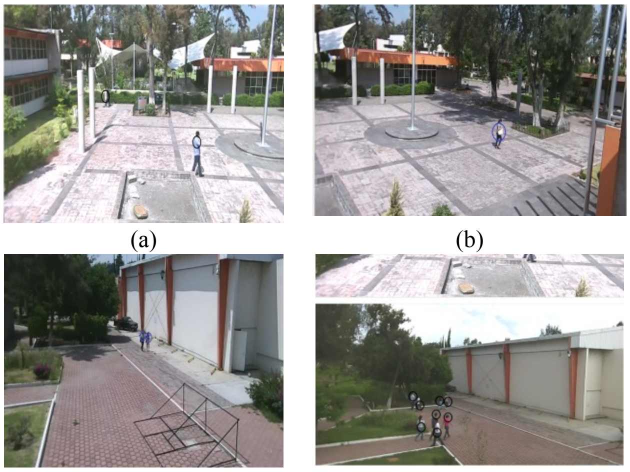
Groups of People

The process of separation of people is to analyze the vertical histogram of the detected object in motion. The vertical histogram of a binary image of size mxn is defined as $h_v(I) = [h_1, \dots, h_n]$, where $h_i = \Sigma_{j=1}^{n} I(i,j)$ and $h_v(I)$ is the vertical sum of the binary digits on which motion has been detected. In this case the local minima were estimated using the maximum transformed by calculating the complement of the vertical histogram as $E_{\min}(h_v) = \max\{h_1, \dots, h_n\} - [h_1, \dots, h_n]$ .Using the positions of the minimum, the original object is divided. Then, for each new object applies Eq. (4) and it is checked whether or not the detected object is indeed a person.

IV RESULTS AND DISCUSSION

The experimental model consist in a set of different videos in outdoor scenarios, we using Outdoor HD wireless IP cameras to test our approach.

The testing process consists of a sequence corresponding to monitoring a full week of a schedule of 9 am to 8 pm. The scenario presents different light conditions caused by shadows, reflections and the different positions of the sun. The video presents interference noise. This interference has been stopped explicitly to assess the level of robustness of the proposal. The noise is caused by two sources: Interference sources of electrical energy and a high level of compression to save bandwidth. The camera uses a capture rate of 15 images per second at a resolution of *320 × 240*. Fig. 4 shows the results of screening of people, although there are other moving objects. Those are marked with a black circle. The images show that the scheme is able to detect people even in great lighting changes caused by external light sources (vehicle lights and sunlight). An example of the segmentation of groups of people with the vertical histogram is shown in Fig. 4 d. This scheme to segment groups of people is effective so long as people do not occlude the camera perspective. The results show that the scheme is suitable to detect activity in specific scenarios, even when lighting changes have on the scene or the video quality is low it acquired. Furthermore, the proposal is suitable for systems real-time response.



(a) (b) (c) (d)

Figure 4. Detected people

Figure 5.Graphical interface

Conclusion

In this paper a method for automatic monitoring of people is presented. The proposed method is computationally efficient and suitable for real time surveillance systems. The method presented for detecting people from object skeletonization is efficient, and moreover, the criterion of segmentation groups of people is appropriate, as long as a side perspective is taken, and there are no occlusions of people. In the experimental stage it shown to be able to identify people, although there are significant changes in the scene like lighting and the quality of the video signal is not optimal.

ACKNOWLEDGMENT

The author Jorge Mateos wishes to thank Concyteq for financial resources and the Tecnológico Nacional de Mexico Campus Querétaro for the installations and resources provide to the development of this work

REFERENCES

[1] M. Piccardi, "Background subtraction techniques: a review," in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '04), pp. 3099-3104, October 2004.

[2] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08), December 2008.

[3] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08), December 2008.

[4] B. Lucas, T. Kanade,T "*An Iterative Image Registration Technique with an Application to Stereo Vision*," DARPA Proceedings on Image Understanding Workshop, pp. 674–679, 1981.

[5] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan; "*Semi-supervised Adapted HMMs for Unusual Event Detection",* Computer Vision and Pattern Recognition, vol. 1, pp. 611–618, 2005.

[6] Y. Ivanov, Ch. Stauffer, A. Bobick, Grimson, "Video Surveillance of Interactions", In CVPR Workshop on Visual Surveillance, pp. 82-89, 1999.

[7] N. Oliver, B. Rosario, A. Pentland, "Graphical Models for Recognizing Human Interactions", Proceedings of Neural Information Processing Systems, pp. 924–930, 1998.

[8] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08), December 2008.

[9] R. J. Radke, S. Andra, O. Al-Kofahi, B. Roysam, Image change detection algorithms: a systematic survey, IEEE Transactions on Image Processing, v.14 n.3, p.294-307, March 2005.

[10] O. Duda, P. Hart, Pattern Classification, Wiley-Interscience; 2 edition 2000.

[11] S. Panahi, S. Sheikhi, S. Hadadan, and N. Gheissari. Evaluation of background subtraction methods. International conference on Digital Image Computing: Techniques and Applications, pp 357–364, 2008.

[12] C. Stauffer, W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking", IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 252–259, 1999.

[13] P. Soille, Morphological Image Analysis: Principles and Applications, Springer-Verlag, 1999.

Jorge Mateos Paullada, currently pursuing a B.S. degree in Computer Systems Engineering with specialization in Distributed Systems in the Tecnológico Nacional de México Campus Querétaro.

Ana Herrera received the B.S. degree in Computer Engineering and her M.S. degree in Computer Science from Informatics Faculty of the Queretaro Autonomous University, Mexico. She received his Ph.D. from Queretaro Autonomous University. Her research interests include morphological image processing and computer vision.

Hugo Jimenez received the B.S degree in Computer Science Engineering at the Queretaro Regional Technological Institute; his M.S. in Computer Science at the Center for Computing Research of National Polytechnic Institute (IPN), Mexico. He received his Ph.D. degree from CICATA Queretaro-IPN. His research includes automatic activities detection, associative memories and time series analysis.

Maribel Bueno, Computer System Engineer, currently teaches at the Technological Institute of Queretaro **currently working as Head of** Department of Computer and Systems, coordinating the implementation of programs of study related to the areas of systems and computer career computer Systems Engineering and the development of research projects and links with the productive sector derived above program also teaches at the subject line programming

Felipe Estrada Rojas, Computer Systems Engineer, currently teaches at the Technological Institute of Querétaro, teaching courses of Data Structures and Database Administration. He was Head of Computer Center, Academic Assistant of rectory, among other administrative positions in the institution, implementing the plans and curricula of educational competency model.

