REAL TIME MOTION DETECTION FOR FAST HUMAN IDENTIFICATION BASED ON FACE RECOGNITION

[1]Pablo Rivas, [2]Mario I. Chacón
Digital Signal Processing & Vision Laboratory,
Chihuahua Institute of Technology
Ave. Tecnológico # 2909 Colonia 10 de Mayo
C.P. 31310 Chihuahua, Chih.
[1]pablorp80@ieee.org, [2]mchacon@itchihuahua.edu.mx

ABSTRACT

The problem of real time in motion human identification based on face recognition has attracted the attention of the scientists and researchers. Because of real time face recognition in non-cooperative environments is a complex task, more tools and algorithms are needed in order to reduce the processing time required for a complete system to operate in an appropriate way. In this paper is presented an algorithm for motion detection that has low complexity and it's designed to operate in real time applications. The algorithm presented was designed specially for a human identification based on face recognition. The algorithms consist in low level digital image processing tools, like binarization, 2D filtering, dilate and erode operators. An auto-adaptive background subtraction based on motion energy method is also presented. The algorithm performs very well on real time applications developed with high-level programming language.

1 INTRODUCCIÓN

Nowadays there is motion detection and analysis methods, which have many applications, like object tracking through unobserved regions [1], automatic pedestrian counting [2], human tracking [3], non-rigid object tracking [4], motion detection on a desktop interface [5], circular motion analysis [6], reactive motion analysis [7], motion analysis for intelligent transportation [8], and also for image reconstruction [9].

CIECE – CITEE 2006
Ponencia presentada en el XVI Congreso Interuniversitario de Electrónica, Computación y Eléctrica – II Congreso de Innovación Tecnológica en Eléctrica y Electrónica
Cd. Obregón, Sonora; del 5 al 7 de Abril de 2006

The applications mentioned previously utilize motion detection or analysis techniques. Some of the existing methods for motion detection vary depending on its application, as an example, in [10] is shown a method to detect motion using motion coherence, other utilize code arrays with correlation and convolution properties [11]. Other authors, for their applications, utilize the background extraction technique [12-16]. Neural or Bayesian networks can be utilized as well [17-18]. Probabilistic methods are suggested in [19], the frame differencing technique is also well utilized, this technique is also called image difference [20-21]; other algorithms work dividing a video sequence in small sub-blocks for independent analysis [13] [22-23], other algorithms were designed for mobile cameras [18] [24], in [14-15] Gaussian techniques were utilized but they are computationally expensive. Some of these methods are non-causal systems, they depend on further inputs of a signal, in terms of digital video this means that the video is already recorded or the video is delayed on purpose, for this reason real time implementation of the algorithms is not possible [23] [12].

Real time face recognition systems require being dynamic and dealing with many other processes. To be dynamic means an exhaustive facial search on every single frame when motion is detected on a scene. After an exhaustive research that there is only one specialized motion detection technique for human identification based on face recognition [25], but this algorithm can be improved. Therefore, the main task is to find an appropriate algorithm for motion detection.

The algorithm must fulfill the following 3 requirements; first, the algorithm must treat the video as causal signal in order to be implemented on a real time system. Second, the method must avoid computational complexity because there are other main tasks like face detection, feature extraction, and face recognition, not addressed in this paper. Third, one of the outputs of the system must include the image of the object in motion in order to be processed searching a human face in a further work. Some of this metrics were taken from [25-26].

2 MOTION DETECTION

In real time face recognition systems the process of the entire system begins with video acquisition followed by a motion detection and segmentation sub-process. After the segmentation of the object in motion, the sub-process of face detection is activated to extract the facial image if detected. When facial image is available, characteristics are extracted from the facial image, and finally the sub-process of face recognition classifies the identity of the input face. This process is described graphically in Figure 1.

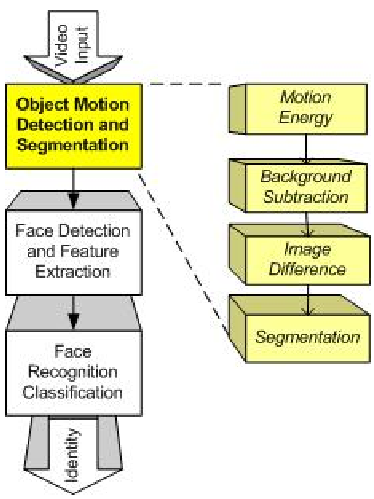


Figure 1. Real time face recognition system. Motion detection sub-system stages.

In figure is shown the stages of the motion detection sub-system proposed, and will be explained in the following sub-sections.
For this project a standard gray scale camera and a standard acquisition card via RCA were utilized for video acquisition.
Most of this work is based on [25] which is a previous research focused on evaluate and find an optimal motion detection methods for face recognition applications.

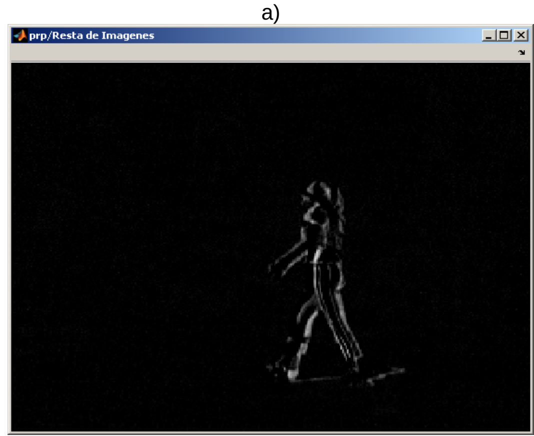
2.1. Image difference.

The image difference is a technique that aims to show any change between two images. Frame differencing is also utilized for motion detection, but for face recognition systems has poor results [25]. This method is also known as frame differencing or image differencing and it is widely utilized because of its easier comprehension and implementation. It is described as follows

$$I_{diff}(x,y,t) = \left| I(x,y,t) - I(x,y,t-1) \right| \quad (1)$$

were $I_{diff}(x,y,t)$ denotes the result of subtract the current image $I(x,y,t)$ and the previous image $I(x,y,t-1)$. The frame reference is t.
The Figure 2 show the results of applying the image difference on two images captured in sequence while a person is walking at normal speed. In a) $I(x,y,t)$ is shown, and b) shows the result $I_{diff}(x,y,t)$.

a)



b)

Figure 2. Image difference result. a) Original image. b) Difference between the original image and a previous image in the sequence.

2.2. Auto-adaptive threshold based background subtraction.

After video acquisition, every frame is grabbed into $I(x,y,t)$ where x,y are the spatial coordinates of every pixel of I. t is the current frame grabbed. The size of an $n \times m$ image is also represented by $x \times y$.

Given $I(x,y,t)$ we have to estimate the energy of the motion described as follows:

$$E(x,y,t) = \sum_{n=0}^{x}\sum_{m=0}^{y}\left[I(x,y,t) - I(x,y,t-1)\right]^2 \quad (2)$$

where $E(x,y,t)$ describes the energy of the motion, and is a scalar real value. The previous frame is denoted by $I(x,y,t-1)$.

Calling the equation (1) the energy equation can be transcribed as

$$E(x,y,t) = \sum_{n=0}^{x}\sum_{m=0}^{y}\left[I_{diff}(x,y,t)\right]^2 \quad (3)$$

If the motion energy is less than the threshold Th which also is a scalar real value, then the image $I(x,y,t)$ is stored at $I_{Bg}(x,y,t)$. Because of the low energy, the current frame is considered as part of background and the motion is considered as minimal. Otherwise the background remains the same. Such cases are defined by the following equation

$$I_{Bg}(x,y,t) = \begin{cases} I(x,y,t) & \forall \quad E(x,y,t) < Th \\ I_{Bg}(x,y,t) & Otherwhise \end{cases} \quad (4)$$

At this point the background is defined and can be obtained at anytime. The frame differencing technique is used as follows

$$I_{diff}(x,y,t) = \left|I(x,y,t) - I_{Bg}(x,y,t)\right| \quad (5)$$

Note that the image subtraction is realized between the current frame and the background approximation in order to segment the objects in motion.

The threshold Th is updated looking for the minimum energy value with the equation 3. If $t=0$, the first value of energy calculated is set to Th. Otherwise every time Th is greater than the current value of energy this is set as the new threshold. The ideal value of the energy when there is no motion in the scene is zero, however in real applications zero values in the motion energy means probably a problem with the video capture device or any other kind of problems.

For this reason if a zero value is reached by the motion energy the threshold value remains the same until energy value changes. The threshold is defined as

$$Th = \begin{cases} E(x,y,t) & if \quad t = 0 \\ E(x,y,t) & if \quad Th > E(x,y,t) \\ Th & if \quad E(x,y,t) = 0 \end{cases} \quad (6)$$

In the Figure 3 a) is presented a frame grabbed when a person is walking through a natural scene. In b) is shown the background calculated with equation 4, with a threshold obtained with the equation 6. The image difference calculated with the equation 5 is shown in c). Finally in d) is shown a plot of the motion energy and the threshold calculated while time changes.

2.3. Image segmentation.

After background subtraction the image needs to be processed to extract the objects in motion. First the image is converted to a binary format in order to be utilized as a kernel or filter to multiply it by the original image in a further step. To convert an image to binary format a threshold is needed, in this case for this project a threshold of 15 is used to convert the image. The binary conversion is defined as follows

$$I_{Bw}(x,y,t) = \begin{cases} 1 & if \quad I(x,y,t) \ge Th_{Bw} \\ 0 & Otherwhise \end{cases} \quad (7)$$

where $Th_{Bw} = 15$.

The second step in the segmentation process is to eliminate as much noise as possible, in not complex processes in terms of CPU time. For this purpose the 8-connectivity theory is utilized to eliminate the camera noise, or nature noise generated by the trees moving because of the wind, especially in outdoor environments. The 8-connectivity is denoted by

$$\begin{aligned} N_8 = \{ & I_{Bw}(x-1,y+1,t), I_{Bw}(x,y+1,t), \\ & I_{Bw}(x+1,y+1,t), I_{Bw}(x-1,y,t), I_{Bw}(x,y,t), \\ & I_{Bw}(x+1,y,t), I_{Bw}(x-1,y-1,t), \\ & I_{Bw}(x,y-1,t), I_{Bw}(x+1,y-1,t)\} \end{aligned} \quad (8)$$

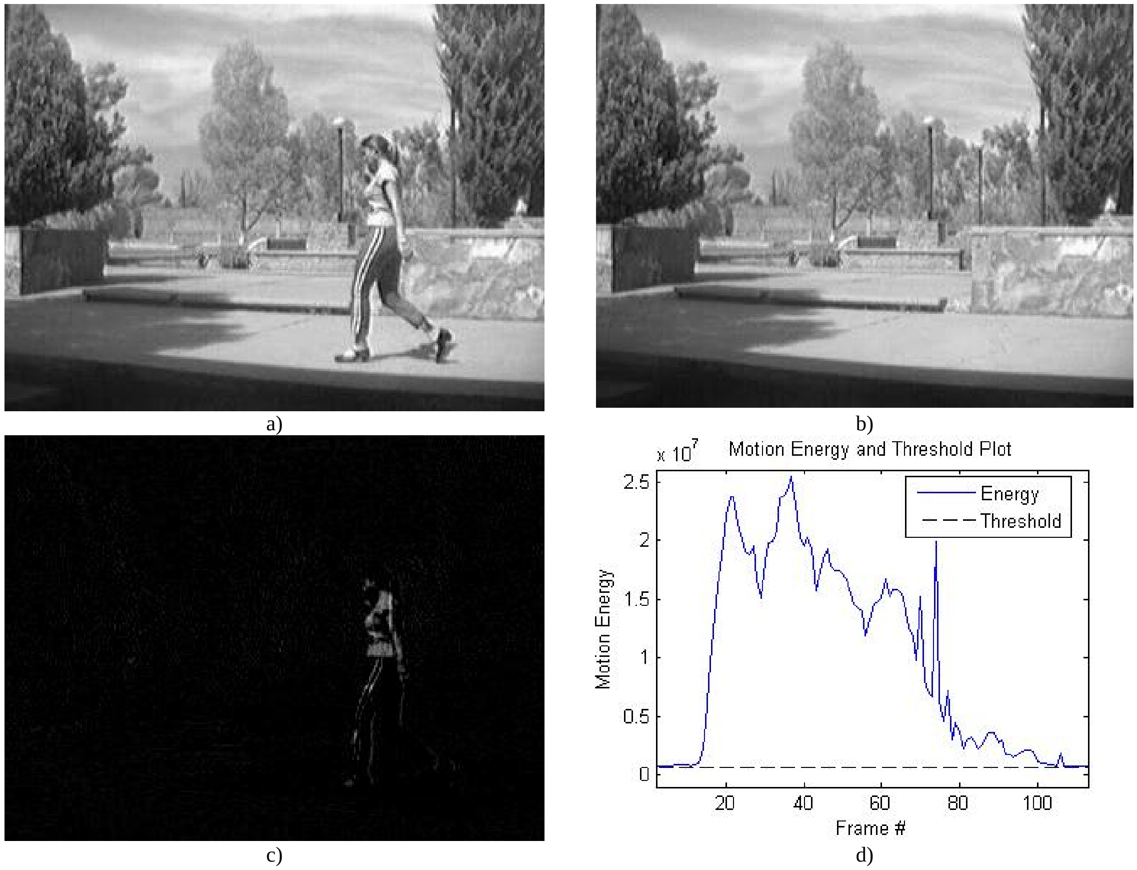


Figure 3. Auto-adaptive threshold based background subtraction. a) A person walking in a natural scene sequence. b) Background subtracted with the proposed method. c) Image difference between a) and c). d) Motion energy (in lighter color) and threshold (in darker color) plot.

where N_8 is evaluated at all the spatial coordinates of $I_{Bw}(x, y, t)$.

The third step tries to connect the pixels of the region of the object in motion via dilate operator. The dilate operator dilates the binary image $I_{Bw}(x, y, t)$ via 2-D convolution with a kernel of ones of size $n \times n$. The convolution of an image with a kernel is denoted by

$$I_c(x,y,t) = I_{Bw}(x,y,t) * h(x,y,t) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} I_{Bw}(m,n,t)h(x-m,y-n,t) \quad (9)$$

$$h(x,y,t) = \begin{bmatrix} 1_{1,1} & 1_{1,2} & \dots & 1_{1,n} \\ 1_{2,1} & 1_{2,2} & \dots & 1_{2,n} \\ \dots & \dots & \dots & \dots \\ 1_{n,1} & 1_{n,2} & \dots & 1_{n,n} \end{bmatrix} \quad (10)$$

$n = 9$

Alter connecting regions dilating the objects, the inverse process is required to get the image with the connected regions of the original sizes. For this, the erode operator is required, and it is applied the same way as the equation 9, but the kernel changes as follows

$$h(x,y,t) = \begin{bmatrix} 0_{1,1} & 0_{1,2} & \dots & 0_{1,n} \\ 0_{2,1} & 1_{2,2} & \dots & 0_{2,n} \\ \dots & \dots & \dots & \dots \\ 0_{n,1} & 0_{n,2} & \dots & 0_{n,n} \end{bmatrix} \quad (11)$$

$n = 9$

The final step for segmentation is simple, after substitute the equation 11 in equation 9, multiply the result of this last step by the original image as follows

$$I_{Obj}(x,y,t) = I_c(x,y,t) I(x,y,t) \quad (12)$$

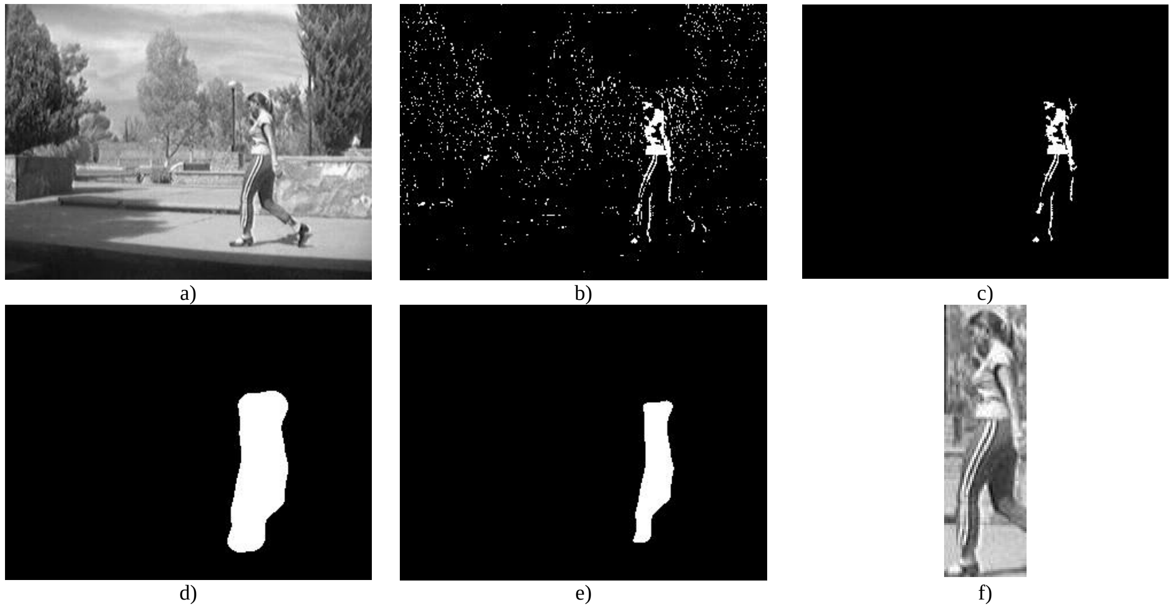


Figure 4. Image segmentation results. a) A person in motion. b) Image difference binarized with threshold of 15. c) Image with lees noise, result of the applied 8-connectivity. d) Result of dilate with a kernel of [9x9]. e) Result of erode with a kernel of [9x9]. f) Results of the segmentation, the person walking.

where $I_{Obj}(x,y,t)$ contains the segmented objects. In Figure 4 is presented the results of the segmentation steps, in a)is the original image with a person walking normally. The image difference between the background and a) after the binarization process with a threshold of 15 is shown in b), and contains noise. c) is the result of apply 8-connectivity into b). Applying convolution to dilate and erode with a kernel size of 9×9 we get the results shown in d) and e) respectively. As expected, the regions were connected in the desired way. Finally in f) is shown the object extracted after the whole process. As can bee seen, the result is optimum for face recognition because it segments the entire object including the face area, which is the most important thing regarding this research.

3 RESULTS AND CONCLUSIONS

In a high-level programming language the algorithm presents real-time operation capabilities. This algorithm process 20 frames per second on a Pentium IV 2.4 GHz, with 512MB RAM.

In this paper it's presented an algorithm to work with video as a causal signal. The algorithm has low computational complexity. And the algorithm outputs an image containing the face area as required in section I. We've presented a method that deals with noise in natural environments which are known as the most subject-to-noise environments, because of the variable illumination conditions, also the wind that moves objects or trees, etc. This method has proved to be robust to detect objects in motion designed for human identification, based on face recognition.

4 ACKNOWLEDGMENTS

The authors want to thank to CONACyT for the support of this research with record #:193324. The authors also want to thank to SEP-DGEST for the support of this research under grant 445.05-P.

5 REFERENCES

[1] Stauffer C., "Learning to track objects through unobserved regions", IEEE Workshop on motion and video computing. 2005.

[2] Biliotti D., Antonini G., Thiran J. P., "Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people on video sequences", IEEE Workshop on motion and video computing. 2005.

[3] Lim J., Kriegman D., "Tracking humans using prior and learned representations of shape and appearance", IEEE International conference on automatic face and gesture recognition. 2004.

[4] Jacquot A., Sturm P., Ruch O., "Adaptive tracking of non-rigid objects based on color histograms and automatic parameter selection" IEEE Workshop on motion and video computing. 2005.

[5] Sato Y., Oka K., Koike H., Nakanishi Y., "Video-based tracking of user's motion for

augmented desk interface", IEEE International conference on automatic face and gesture recognition. 2004.
[6] Jiang G., Wei Y., Quan L., Tsui H., Shum H. Y., "Outward-looking circular motion analysis of large image sequences", IEEE Transactions on pattern analysis and machine intelligence. Vol. 27 pp. 271–277. February 2005.
[7] Choi W., Jeong S. Z., Hashimoto N., Hasegawa S., Koike Y., Sato M., "A development and evaluation of reactive motion capture system with haptic feedback", IEEE International conference on automatic face and gesture recognition. 2004.
[8] Wu Q., Cheng H., "Motion detection based on two-piece linear approximation for cumulative histograms of ratio images in intelligent transportation systems", IEEE International conference on networking, sensing & control, March 2004.
[9] Zhang Y., Xiao J., Shah M., "Motion layer based object removal in videos", IEEE Workshop on applications on computer vision. 2005.
[10] Enzweiler M., Wildes R. P., Herpers R., "Unified target detection and tracking using motion coherence", IEEE Workshop on motion and video computing. 2005.
[11] El-Khamy S., El-Ganainy N., "New techniques for image change and motion detection based on complete complementary code arrays", 21st National radio science conference. March 2004.
[12] Haga T., Sumi K., "Human detection in outdoor scene using spatio-temporal motion analysis", IEEE International conference on pattern recognition, 2004.
[13] Seki M., Fujiwara H., Sumi K., "A robust background subtraction method for changing background", IEEE pp. 207–213. 2000.
[14] Zivkovic Z., "Improved adaptive gaussian mixture model for background subtraction" IEEE International conference on pattern recognition, 2004.
[15] Stauffer C., Grimson W. E. L., "Adaptive background mixture models for real-time tracking", IEEE, 1999.
[16] Thongkamwitoon T., Aramvith S., Chalidabhongse T.H., "An adaptive real-time background subtraction and moving shadows detection", IEEE International Conference on multimedia and expo, pp. 1459–1462. 2004.
[17] Chen M., Ma G., Kee S., "Pixels classification for moving object extraction", IEEE Workshop on motion and video computing. 2005.
[18] Jia Z., Balasuriya A., "Motion based image segmentation with unsupervised Bayesian learning", IEEE Workshop on motion and video computing. 2005.
[19] Veit T., Cao F., Bouthemy P., "Probabilistic parameter-free motion detection" IEEE Computer society conference on computer vision and pattern recognition, 2004.
[20] Sangi P., Heikkila J., Silven O., "Motion analysis using frame differences with spatial gradient measures", IEEE International conference pattern recognition, 2004.
[21] Tian Y., Hampapur A., "Robust salient motion detection with complex background for real-time video survelliance", IEEE Workshop on motion and video computing. 2005.
[22] Reddy V., Sengupta S., "A new predictive full-search block motion estimation", IEEE International conference pattern recognition, 2004.
[23] Latecki L. J., Miezianko R., Pokrajac D., "Motion detection based on local variation of spatiotemporal texture", Proc. Of CVPR Workshop object tracking and classification beyond the visible spectrum. Washington, July 2004.
[24] Agrawal M., Konolige K., Iocchi L., "Real-time detection of independent motion using stereo" IEEE Workshop on motion and video computing. 2005.
[25] Rivas Perea P., Chacón Murguía M. I., "Evaluación de métodos de detección de movimiento para un sistema de identificación de personas basado en reconocimiento facial" International Congress on Electronics, ELECTRO 2005.
[26] Schlogl T., Beleznai C., M. Winter, H. Bischof, "Performance evaluation metrics for motion detection and tracking" IEEE International conference on pattern recognition, 2004.

CURRICULUM VITAE

Pablo Rivas Perea, received the BS in Computer Systems Engineering degree from the Nogales Institute of Technology, Sonora Mexico in 2003. Actually he is pursuing the MS in Electrical Engineering at the Chihuahua Institute of Technology, Chihuahua Mexico since 2004. He has national and international publications on the fields of computer science and electrical engineering. He is a student member of the IEEE and of the following societies: Aerospace and Electronic Systems, Broadcast Technology, Circuits and Systems, Communications, Computational Intelligence, Computer, Microwave Theory and Techniques, and Signal Processing, Society of Industrial and Applied Mathematics. He is chair of the IEEE Chihuahua Student Branch.

Mario I. Chacón Murguía, received the BS in Electrical Engineering and the MS in Computer Engineering degree from the Chihuahua Institute of Technology, Chihuahua Mexico in 1982 and 1985 respectively and his Ph.D. in Electrical Engineering from New Mexico State University , USA, in 1998. He has made research for the industry and other universities. Dr. Chacón has several scientific and technical international publications. He is a Senior member of the IEEE and member of the Pattern Recognition Society. His current research interests include Computer Vision, Digital Signal Processing (DSP), Pattern Recognition, Neural Networks and Fuzzy Logic applications to DSP, and Digital Systems.