An Automated Surveillance System Based on Multi-Processor and GPU Architecture

Mossaad Ben Ayed
AlGhat College of Science and Humanities
Al Majmaah University, KSA
and University of Sfax, Tunisia
mm.ayed@mu.edu.sa

Sabeur Elkosantini
College of Engineering
King Saud University
Kingdom of Saudi Arabia
selkosantini@ksu.edu.sa

Mohamed Abid
Sfax National School of Engineers
University of Sfax
Tunisia
mohamed.abid@enis.rnu.tn

Abstract–**Video surveillance systems are a powerful tool applied in various systems. Traditional systems based on human vision are to be avoided due to human errors. An automated surveillance system based on suspicious behavior presents a great challenge to developers. Such detection is a rather complex procedure and also a rather time-consuming one. An abnormal behavior could be identified by: actions, faces, route, etc. The definition of the characteristics of an abnormal behavior still present a big problem. This paper proposes a specific architecture for a surveillance system. The aim is to accelerate the system and obtain a reliable and accelerated suspicious behavior recognition. Finally, the experiment section illustrates the results with comparison of some of the most recent approaches.**

Keywords—Surveillance system; suspicious behaviors; multi-processor; GPU

I. INTRODUCTION

Surveillance systems are increasingly monitored by computers. The main goal of a surveillance system is identifying suspicious or undesirable behaviors such as theft and loitering with intent [1]. By definition, suspicious behavior is an anomalous behavior which is likely to threaten human life, health, property and freedom. Developers propose three essential steps, to recognize suspicious behavior: object detection, tracking, and behavior exploration. The first challenge is to define models to recognize a suspicious behavior. An anomalous behavior is not a fixed action but it is a complex behavior combining a series of actions and simple behaviors. Thus, it is difficult to be recognized accurately. The object's actual detection presents a previous step which must be considered in order to recognize the potentially suspicious people. A comparison among the main background subtraction methods is used to detect objects. Then the tracking step is essential to define trajectory or behavior kind/type. Several algorithms are used in literature but the results are not always satisfactory [2]. This paper is focused on a specific case: detect the attempt of theft or scam in the case of ATM security surveillance. This detection is performed by the exploration of tracking and squatting action.

The second challenge is the design of a real-time surveillance system. It is known that video/image processing requires a specific architecture to obtain real-time results. In this paper, we present different techniques used to accelerate the execution speed and we propose an attempt based on DECOC classifier to ensure the real-time execution. Existing surveillance systems suffer from several issues:

1. Most traditional methods in surveillance are based on manual/visual detection [4].

2. Most of today's surveillance is not used to prevent an incident but only used to identify what has already happened [5].

3. Most of surveillance system suffers from no real-time detection of suspicious behavior. This problem is due of the complexity of the algorithms [5].

4. Surveillance system violates the privacy of citizens. For example, in USA, many groups are against the use of surveillance system in public areas [6].

In light of this brief introduction on suspicious behavior based on surveillance system, this paper purports to contribute the following tentative proposals:

1. it proposes an embedded intelligent camera for real-time execution. The intelligent camera ensures the privacy of citizens because all the treatment will be done in camera.

2. It applies the proposed design with respect to the ATM system.

II. RELATED WORK

A. Suspicious Behavior Recognition Algorithms

Video surveillance systems pass by three phases in literature. The first phase uses analog CCTVs and the automation is little exploited (1960-1980). The second phase was based on computer vision using digital CCTVs (1980-2000). From 2000, the third phase is based on semi-automated video-surveillance systems [7]. As mentioned in the previous section, each suspicious behavior recognition is composed essentially by three steps: object detection, tracking, and behavior exploration.

1)   Object Detection

There are many algorithms related to object detection, but they still suffer from complexity due to different specific situation. One of the most used methods for object detection is the subtraction of the background [8, 9]. Other works based in the last method are improved by formulated technique [10]. Multi-layer background subtraction represents another method based on color and texture [11]. Second works are based on segmentation algorithms [12]. As conclusion, the object detection is commonly done by using the background subtraction method [13].

2)   Tracking

Tracking methods is widely described in previous works. But these works suffer from low accuracy. Tracking object system is used in many fields as: crowded environment [14], traffic situation [15] and maritime surveillance [16]. In the field of surveillance, the essential goal of the tracking object is to analyze or to extract the human behavior: trajectory, gesture, event [13, 17-18].

3)   Behavior Recognition

All suspicious behavior recognition methods can be classified into two categories: single-layered and hierarchical approaches [3]. The first is suitable for gesture recognition and the second is adopted for complex activities (Figure 1). Single-layered approaches can be classified into two types depending on how they model human activities: space–time approaches [19-22] and sequential approaches [23-25]. Space–time approaches consider the input video as a set of frames at a particular point of time. These approaches recognize behavior by extracting trajectories or local interest points. Whereas sequential approaches interpret an input video as a sequence of observations and recognize it by exemplar based methodologies or model-based methodologies. Hierarchical approaches are classified based on the recognition methodologies they use: statistical approaches [26], syntactic approaches [27], and description-based approaches [28, 29]. But all the previous works cannot be applied in real-time due to the enormous amount of computation required [30].
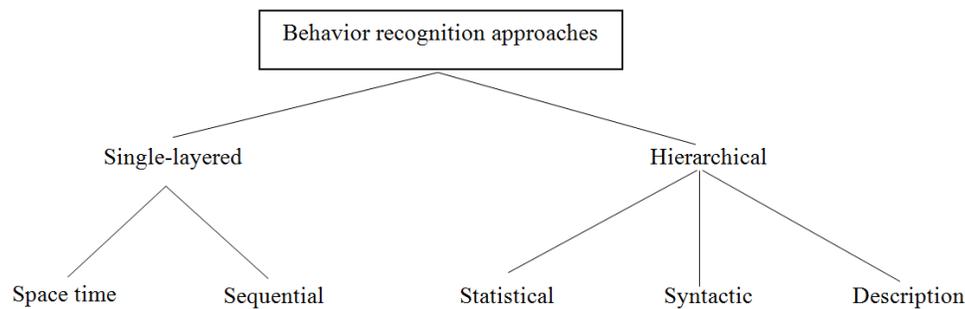

Fig. 1.   Different behavior recognition approaches.

B. Real-time detection based on hardware acceleration

One of the most important goals of security surveillance is to collect and disseminate real-time information and provide situational awareness to operators and security analysts [5]. There are many research works that propose an advanced architecture in the field of video-surveillance. In [41] a co-design strategy is adopted with FPGA to ensure automated video surveillance in the case of the object detection. Other works are focused in embedded cameras for tracking systems [42, 43]. In [41, 44, 45], advanced designs are proposed to accelerate the detection of human motion. Surveillance systems are facing different challenges like the low accuracy of tracking due to the low quality videos, and the real-time detection. The present paper proposes an accelerated architecture for suspicious behavior. This attempt uses Arena tool as a higher level of description for modeling and simulation. Based on obtained results, an automated and accelerated architecture is proposed to ensure real-time recognition.

III.   PROPOSED SURVEILLANCE SYSTEM: MODELING AND SIMULATION

Embedded system is used in industries, surveillance, smart cities, intelligent systems, etc. There are several environments for modeling and simulation depending on level description and system's field. Arena environment provides the description of an event-discrete system at conceptual level. Based on Suspicious behavior's recognition steps, Arena's model is composed by three processes: object detection, tracking, and behavior exploration. Two models could be adopted for design, (Figures 2, 3): (1) mono-processor, (2) multi-processor.

A. Model 1: Mono-Processor Architecture

Entity: Name: Frames Arrivals

                 Arrival time: Const (0.1) (s)

Process: Name: Object Detection

                 Resources: Processor

                 Delay type: TRIA (0.02,0.04,0.05) (s)

Process: Name: Tracking

                 Resources: Processor

                 Delay type: TRIA (0.05, 0.08, 0.1) (s)

Process: Name: Behavior Exploration

                 Resources: Processor

                 Delay type: TRIA (0.4, 0.5, 0.9) (s)

B.   Model 2: Multi-Processor Architecture

Entity: Name: Frames Arrivals

                Arrival time: Const (0.1) (s)

Process: Name: Object Detection

                Resources: Processor1

                Delay type: TRIA (0.02,0.04,0.05) (s)

Process: Name: Tracking

                Resources: Processor2

                Delay type: TRIA (0.05, 0.08, 0.1) (s)

Process: Name: Behavior Exploration

                Resources: Processor3

                Delay type: TRIA (0.4, 0.5, 0.9) (s)

The simulation is run for 1 hour. The simulation results of model 1 are shown in Figure 3a. They show that an architecture based on single processor is insufficient for surveillance system. The system should have big buffers between components to ensure the safety of data and that the bandwidth is quickest than the time treatment. The real-time execution capability is not proved for this model. The simulation results of model 2, are shown in Figure 3b, and are better than those of model 1. The multi-processor architecture proposed for surveillance system improves speed and requires a big buffer between Processor2 and Processor3. The design still suffers from no real-time execution.

Based on the simulation results, a specific architecture is described in the next section

IV.   EXPERIMENT RESULTS

This section presents the implementation of the surveillance system using a hybrid architecture based on multi-processor and GPU. This solution is well justified by simulation results presented in the previous section. The aim of this section is to describe a hardware accelerator based on the Gabor filter. The proposed architecture is composed by Microblaze soft-core processors interfaced with shared memory using the AXI4 bus [46]. All processor elements, on-chip memories, and the AXI4 bus are clocked at 100 MHz. Off-chip memory is clocked at 200 MHz and the AXI-lite bus is clocked at 50 MHz.
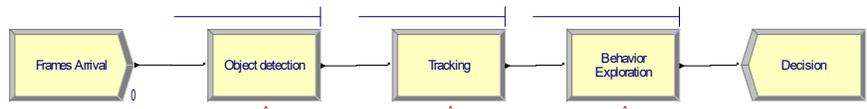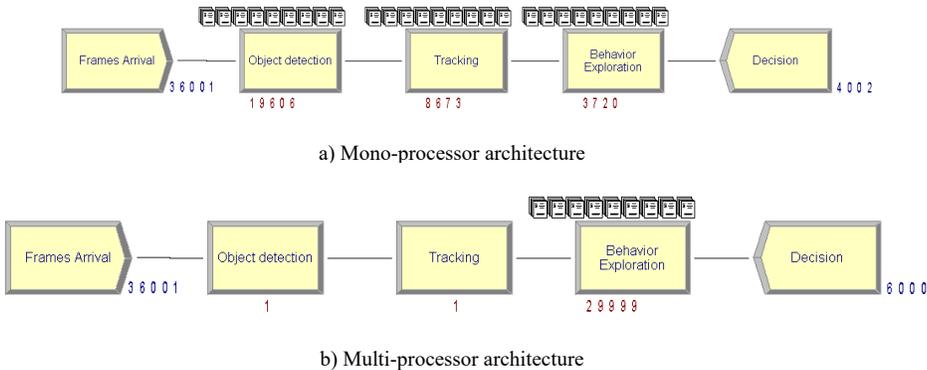

Fig. 2.   Surveillance system model.

a) Mono-processor architecture

b) Multi-processor architecture

Fig. 3.   Simulation results.

The surveillance system's architecture is divided into:

- Hardware components: Behavior exploration.

- Software applications: Object detection and tracking.

The frame is received by a professional camera with frame size of 640x840 pixels. The proposed design is composed by SRAM memory, control unit, 2 processor elements (cores) and hardware accelerators based on behavior exploration, see Figure 4. The proposed architecture belongs to SIMD/MPSOC field [47]. The control unit represents an essential component in design. The last aims to handle all processes:

- Arbitrating the access of processors unit to/from memory

- Handling and commanding all the processor units

At first, the frame is captured with the camera and saved into global memory. The algorithm of surveillance system is divided into software modules (object detection and tracking) and hardware components (behavior exploration). In Table I, the proposed design is compared with others and it shows the best execution time for a surveillance system. Our proposition's speed-up factor is about 20 times the ones in [40], and [37], about 2 times the one in [30], 10 times the one in [39] and 4 times the one in [38].
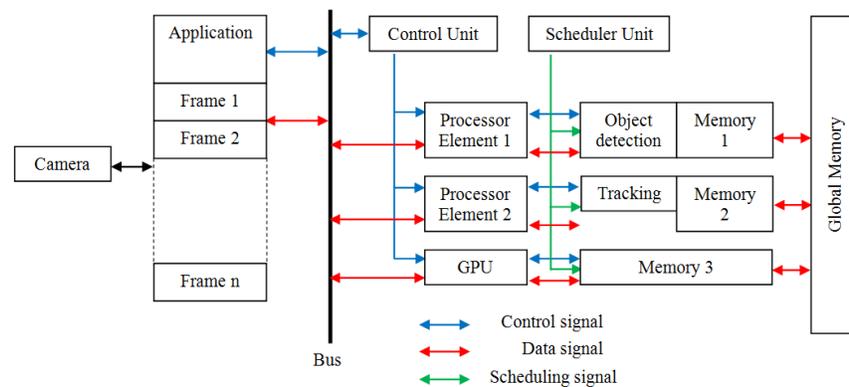
Fig. 4.　　Multi-processor and GPU architecture for surveillance system.

From the experiment's results mentioned in Table I, the proposed algorithm has the best result in terms of reliability and speed. This proves that the proposed architecture based on combination between Multi-processor and GPU ensures the real-time execution.

TABLE I.　　A COMPARISON BETWEEN DIFFERENT SUSPICIOUS BEHAVIOR DETECTION METHODS BASED ON EXECUTION TIME

Method	Execution time (s)
[37]	0.2>
[38]	0.041
[39]	0.1
[40]	0.2>
[30]	0.02
Proposed method	**0.01**

V.　CONCLUSION

This paper resumes the different automated surveillance systems in the literature. The goal is to obtain a reliable detection of suspicious behavior with respect to the real-time constraint for an ATM system. This successful attempt proposes a hybrid architecture based on multi-processor and GPU to accelerate the treatment time of a frame. This architecture ensures the no loss of data between processes. As a main contribution, a whole hardware design based on multi-processor architecture is described. The proposed design combines with 2-cores and GPU. Compared with other works cited, our design achieves the best execution time.

ACKNOWLEDGMENT

REFERENCES

[1]　E. R. Davies, Computer and Machine Vision, Academic Press, 2012

[2]　R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, J. Almazan, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls", Expert systems with Applications, Vol. 42, No. 21, pp 7991-8005, 2015

[3]　J. K. Aggarwal, M. S. Ryoo, "Human activity analysis: a review", ACM computing surveys, Vol. 43, No. 16, 2011

[4]　R. Shimonski, "Digital reconnaissance and surveillance", Chapter 1 in Cyber Reconnaissance, Surveillance and Defense, Elsevier, 2015

[5]　L. Deligiannidis, H. R. Arabnia, Emerging Trends in Image Processing, Computer Vision and Pattern Recognition, Morgan Kauffman 2015

[6]　Electronic Frontier Foundation, "NSA Spying", Available at https://www.eff.org/nsa-spying

[7]　T. D. Raty, "Survey on contemporary remote surveillance systems for public safety", IEEE Transactions on Systems, Man and Cybernetics Part C, Vol. 40, No. 5, pp. 493–515, 2010

[8]　L. D. Stefano, C. S. Regazzoni, D. Schonfeld, "Advanced video-based surveillance", EURASIP Journal on Image and Video Processing (JIVP), Vol. 2001, No. 1, p. 857084, 2011

[9]　S. Brutzer, B. Hoferlin, G. Heidemann, "Evaluation of background subtraction techniques for video surveillance", IEEE Conference on Computer Vision and Pattern Eecognition, pp. 1937–1944, 2011

[10]　L. Maddalena, A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications", IEEE Transactions on Image Processing, Vol 17, No. 7, pp. 1168–1177, 2008

[11]　J. M. Odobez, J. Yao, "Multi-layer background subtraction based on color and texture", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2007

[12]　F. E. Baf, T. Bouwmans, B. Vachon, "Background modeling using mixture of gaussians for foreground detection - a survey", Recent Patents on Computer Science, Vol. 1, No.3, pp. 219–237, 2008

[13]　R. Arroyo, J. J. Yebes, L. M. Bergasa, G. Daza, J. Almazn, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls", Expert Systems with Applications, Vol. 42, No. 21, pp. 7991–8005, 2015

[14]　D. Chau, M. Thonnat, F. Bremond, E. Corvee, "Online parameter tuning for object tracking algorithms", Image and Vision Computing, Vol 32, No. 4, pp. 287–302, 2014

[15]　S. Alvarez, D. Llorca, M. Sotelo, "Hierarchical camera auto-calibration for traffic surveillance systems", Expert Systems With Applications , Vol. 41, No. 4, pp. 1532–1542, 2014

[16]　Z. L. Szpak, J. R. Tapamo, "Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set", Expert Systems With Applications, Vol. 38, No. 6, pp. 6669–6680, 2011

[17]　M. Cristani, R. Raghavendra, A. Del Bue, V. Murino, "Human behavior analysis in video surveillance: A social signal processing perspective", Neurocomputing, Vol. 100, pp. 86–97, 2012

[18]　W. Hu, T. Tan, L. Wang, S. Maybank, "A survey on visual surveillance of object motion and behaviors", IEEE Transactions on Systems, Man and Cybernetics Part C, Vol 34, No. 3, pp. 334–352, 2004

[19]　C. Rao, M. Shah, "View-invariance in action recognition", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. II-316-II-322, 2001

[20] S. Savarese, A. Delpozo, J. Niebles, L. Fei-Fei, "Spatial-temporal correlations for unsupervised action classification", IEEE Workshop on Motion and Video Computing, pp. 1-8, 2008

[21] M. D. Rodriguez, J. Ahmed, M. Shah, "Action MACH: a spatiotemporal maximum average correlation height filter for action recognition", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008

[22] M. S. Ryoo, J. K Aggarwal, "Spatio-temporal relationship match: video structure comparison for recognition of complex human activities", IEEE 12th International Conference on Computer Vision, pp. 1593-1600, 2009

[23] H. Jiang, M. Drew, Z. Li, "Successive convex matching for action detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1646-1653, 2006

[24] A. Veeraraghavan, R. Chellappa, A. Roy-Chowdhury, "The function space of an activity", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 959-968, 2006.

[25] P. Natarajan, R. Nevatia, "Coupled hidden semi-markov models for activity recognition", IEEE Workshop on Motion and Video Computing, pp. 10, 2007

[26] D. Damen, D. Hogg, "Recognizing linked events: searching the space of feasible explanations", IEEE Conference on Computer Vision and Pattern Recognition, pp. 927-934, 2009

[27] S. W. Joo, R. Chellappa, "Attribute grammar-based event recognition and anomaly detection", IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 107, 2006

[28] M. S. Ryoo, J. K. Aggarwal, "Semantic representation and recognition of continued and recursive human activities", International Journal of Computer Vision, Vol. 82, No. 1, pp. 1-24, 2009

[29] M. S. Ryoo, J. K. Aggarwal, "Recognition of composite human activities through context-free grammar based representation", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1709-1718, 2006

[30] C. Mu, J. Xie, W. Yan, T. Liu, "A fast recognition algorithm for suspicious behavior in high definition videos", Multimedia Systems, Vol. 22, No. 3, pp 275-285, 2016

[31] S. Shaikh, A. Maiti, N. Chaki, "A new image binarization method using iterative partitioning", Machine Vision and Applications, Vol. 24, No. 2, pp 337-350, 2013

[32] J. Wang, L. Wu, Y. Liu, "Nios II Processor-Based Fingerprint Identification System", Nios II Embedded Processor Design Contest— Outstanding Designs 2007

[33] A. Beristain, M. Grana, "A stable skeletonization for tabletop gesture recognition", International Conference on Computational Science and Its Applications, pp. 610-621, 2010

[34] M. Ben Ayed, F. Bouchhima, M. Abid, "A novel application of the classifier DECOC based on fingerprint identification", Workshop on Database and Expert Systems Applications, pp. 288-292, 2010

[35] C. Schuldt, I. Laptev, B. Caputo, "Recognizing human actions: a local SVM approach", 17th International Conference on Pattern Recognition, Vol. 3, pp. 32-36, 2004

[36] M. Marszalek, I. Laptev, C. Schmid, "Actions in context", IEEE Conference on Computer Vision and Pattern Recognition, pp. 2929-2936, 2009

[37] K. Schindler, L. J. Van Gool, "Action snippets: how many frames does human action recognition require?", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008

[38] A. Gilbert, J. Illingworth, R. Bowden, "Fast realistic multi-action recognition using mined dense spatio-temporal features", IEEE 12th International Conference on Computer Vision, pp. 925-931, 2009

[39] A. Yao, J. Gall, L. Van Gool, "A hough transform-based voting framework for action recognition", IEEE Conference on Computer Vision and Pattern Recognition, pp. 2061-2068, 2010

[40] S. Sadanand, J. J. Corso, "Action Bank: a high-level representation of activity in video", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1234-1241, 2012

[41] T. N. Hau, W. I. Robert, N. R. Ryan, P. B. Randy, "Real-time video surveillance on an embedded, programmable platform", Microprocessors and Microsystems, Vol. 37, No. 6-7, pp. 562–571, 2013

[42] M. Casares, S. Velipasalar, "Resource-efficient salient foreground detection for embedded smart cameras by tracking feedback", IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 369–375, 2010

[43] J. Schlessman, C. Y. Chen, B. Ozer, K. Fujino, K. Itoh, W. Wolf, "Hardware/software co-design of an FPGA-based embedded tracking system", Conference on Computer Vision and Pattern Recognition Workshop, pp. 123–130, 2006

[44] H. Meng, C. Freeman, N. Pears, C. Bailey, "Real-time human action recognition on an embedded, reconfigurable video processing architecture", Journal of Real-Time Image Processing, Vol. 3, No. 3, pp.163–176, 2008

[45] [45] B.M.A Amer, and S.A.R Al-Attas, "Smart surveillance using PDA, Word Academy of Science", Engineering and Technology, Vol. 66, pp.251–255, 2010

[46] S. Saadi, M. Touiza, F. Kharfi, A. Guessoum, "Dyadic wavelet for image coding implementation on a Xilinx MicroBlaze processor: Application to neutron radiography", Applied Radiation and Isotopes, Vol. 82, pp.200-210, 2013

[47] D. Watson, A. Ahmadinia, "Memory customisations for image processing applications targeting MPSoCs", Integration, the VLSI Journal, Vol. 51, pp. 72-80, 2015