Ain Shams Engineering Journal xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Electrical Engineering

Event triggered intelligent video recording system using MS-SSIM for smart home security

Haitham Abbas Khalaf *, A.S. Tolba, M.Z. Rashid

Computer Science Department, Faculty of Computers and Information, Mansoura University, Egypt

ARTICLE INFO

Article history:
Received 18 June 2016
Revised 8 September 2016
Accepted 11 October 2016
Available online xxxx

Keywords:
Smart homes
Multi-modal collaborative sensing
Intelligent video recording
Event-triggered recording
Motion detection
Structural similarity index

ABSTRACT

This paper presents an intelligent system for event-triggered video recording for smart home applications. Video recording is triggered through a collaborative sensing strategy. PIR motion detectors are used for both directing the master wireless IP-camera for recording in a specific direction in the entrance hall or initiating other wireless IP-cameras for recording inside the rooms. An activated wireless camera starts video recording only during a targeted motion interval. Motion detection for initiation of the recording process is based on an enhanced Multi-Scale Structural Similarity detection technique. RFID tags are used in all rooms to identify persons entering these rooms. When the moving object shifts to another location at home, the local PIR sends a signal to the Gateway which initiates another video camera. Sensors collaborate for identification of the area to be monitored and the events which are to be recorded. The proposed system helps cover all smart home areas, save the required storage space and speeds-up video event analysis.

1. Introduction

Cost minimization and target tracking facilitation are the major benefits of optimal sensor localization in smart homes. The Smart Home Management System (SHMS) needs to know the location of the different sensor nodes within the home in order to speed up the processing of data and to minimize the cost of system operation. In smart homes, fixed location or mobile sensor stations could be used.

Smart homes should have intelligent systems which could observe the events which occur 24 h/7 days. Recording and analysis of such events put a huge burden on the computational platforms and need huge amounts of storage space. To reduce both the computational and storage costs and speed-up event analysis, intelligent surveillance systems are badly needed. Fig. 1 shows the layout of an event-triggered video recording system which forms a key component of SHMS.

Here, we describe an intelligent event-triggered video recording system which couples a set of passive infrared detectors (PIRs), wireless IP cameras, RFID tags with an Arduino based gateway (Fig. 2). PIRs enable the detection of moving objects, in order to direct the suitable camera towards the moving target. A Multi-Scale Structural Similarity (MS-SSIM) algorithm is then used for motion detection in its view window. Power saving is also controlled by a PIR sensor. The system is turned-on upon motion detection using a PIR.

A gateway (Fig. 3) acts as an anchor device which controls the activation of the wireless webcams adaptively according to the activities of smart home inhabitants. For example, the first webcam (WC) located at the entrance hall starts recording the activity of an entering inhabitant. Once the inhabitant disappears from the FOV of the WC, the gateway (Anchor device-Arduino board with wireless connectivity) initiates a second WC based on a PIR based Motion Detection Module (MDM) signal received from the new inhabitant location which lies outside the coverage area of the fist WC (Fig. 4).

Motion detectors in smart home could be classified into types: Wireless cameras and Passive Infra-red (PIR) detectors. The PIR motion detection module is shown in Fig. 4. The module includes the binary mode PIR sensor, a wireless communication module and a control board based on ATMEL 8, microcontroller. The module is powered by a chargeable battery. The MDMs are allocated at important entries in the smart home like the doors of home, rooms, kitchen and bath room.

Fig. 2 shows the MDMs distribution in an experimental home. MDMs are placed in all places where inhabitants frequently move. PIR data is wirelessly transmitted from MDM to the Arduino based

Peer review under responsibility of Ain Shams University.

* Corresponding author.
E-mail addresses: co.alhaitham@gmail.com (H. Abbas Khalaf), ast@astolba.com (A.S. Tolba), magdi_z2005@yahoo.com (M.Z. Rashid).

http://dx.doi.org/10.1016/j.asej.2016.10.001
2090-4479/© 2016 Ain Shams University. Production and hosting by Elsevier B.V.

Please cite this article in press as: Abbas Khalaf H et al. Event triggered intelligent video recording system using MS-SSIM for smart home security. Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.10.001

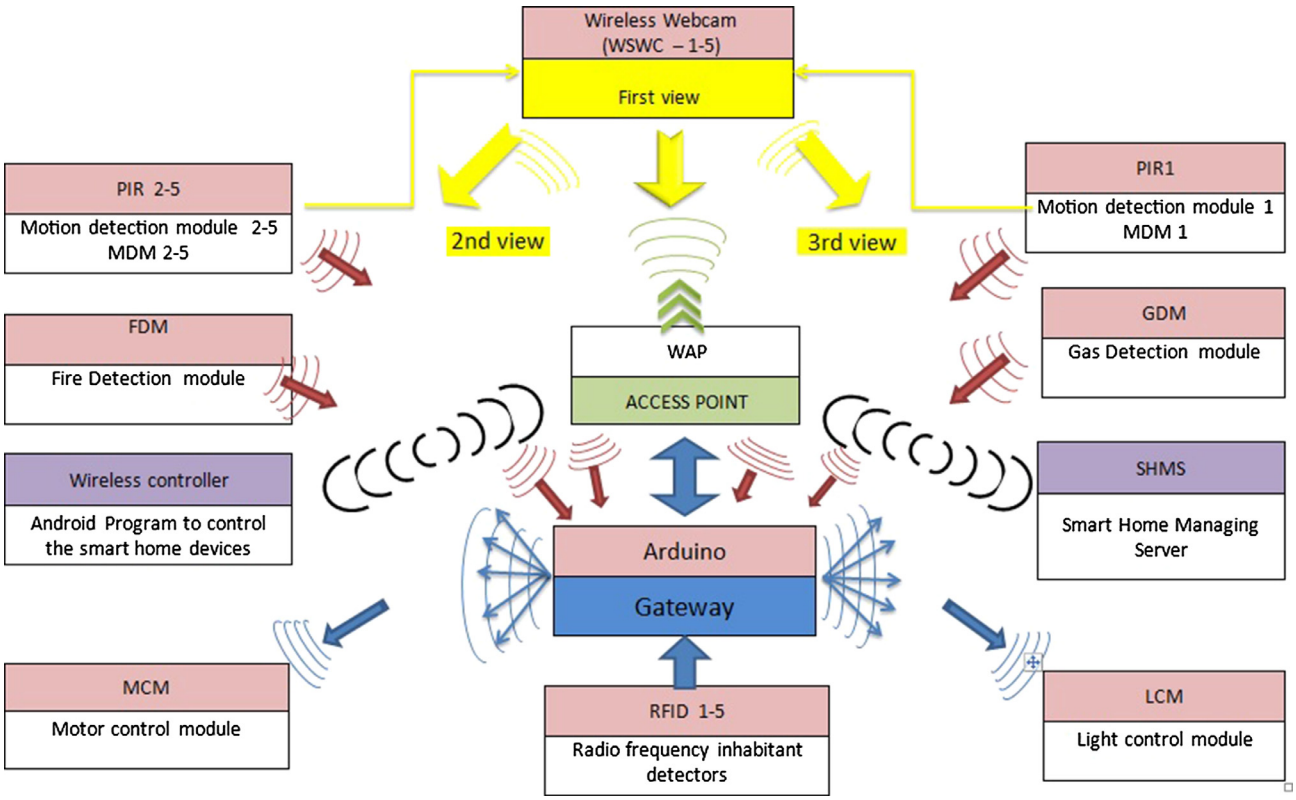

Figure 1. System layout.

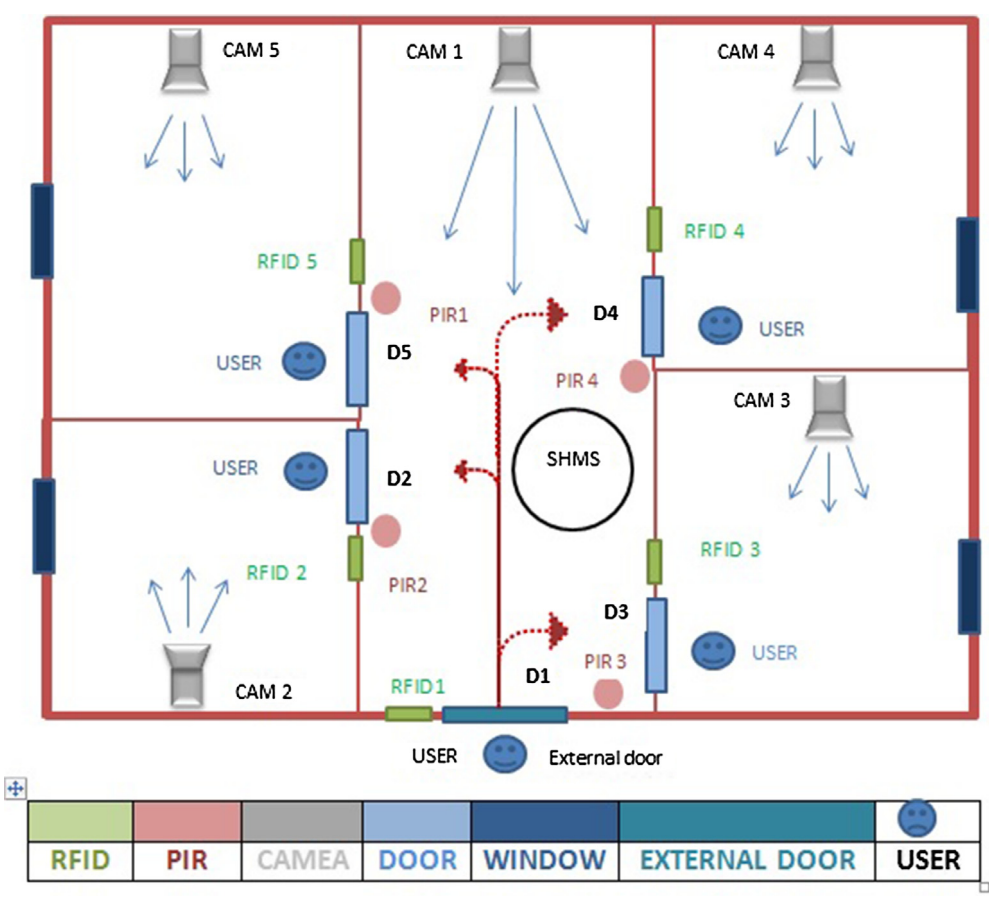

Figure 2. Sensor distribution within the smart home.

Please cite this article in press as: Abbas Khalaf H et al. Event triggered intelligent video recording system using MS-SSIM for smart home security. Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.10.001

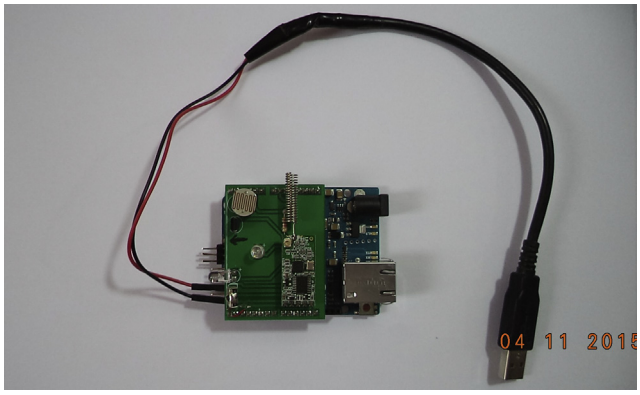

Figure 3. Gateway (Arduino Board).

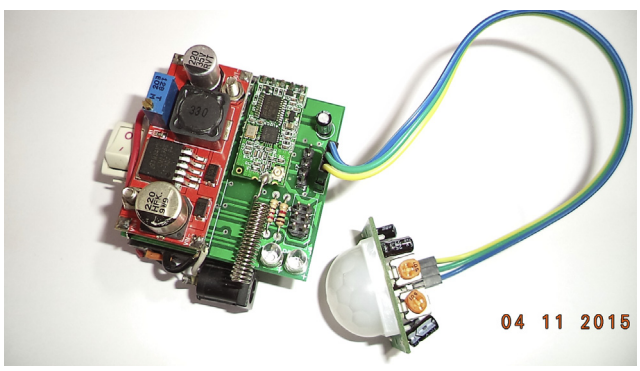

Figure 4. PIR based Motion Detection Module (MDM).

gateway through a wireless access point. A Smart Home Server (SHS) is connected wirelessly with the Arduino Gateway. The sensory data received from sensors are processed for decision making. Detected motion initiates the wireless camera for video event detection to save storage and processing cost.

2. Related work

Table 1 summarizes the most widely used motion detection methods and highlights the limitations and advantages of each method.

Table 1
Comparison between motion detection methods.

Reference	Motion detection method	Description, advantages and limitations
[1–4]	Background subtraction (BS)	Description: each frame in a video sequence is compared with a reference background image Limitations: using a static reference image is not always accurate Light or illumination changes and small structural changes highly affect detection results Sensitivity to local illumination changes such as shadows and highlights Sensitivity to global illumination changes Noise effect needs filtering Background interruption needs morphological filtering Very sensitive to the changes in the external environment and has poor anti-interference ability Advantages: simple processing Provide complete object information in the case of known background Moderate accuracy and processing time Using dynamic template matching for reference image extraction results better detection accuracy
[5,6,19]	Optical flow (OF)	Description: motion of corresponding pixels in consecutive frames is calculated Limitations: high computation overhead High sensitivity to noise, poor anti-noise performance, make it not suitable for real-time demanding occasions Requires additional hardware to support the performance Advantages: provide direction information Has moderate accuracy
[3,7]	Sum of absolute differences (SAD)	Description: Calculates the sum of absolute differences of gray/color levels of corresponding pixels in consecutive frames Limitations: sensitivity to light and structural changes Less detection accuracy Advantages: Less processing Simple and easy to implement processing
[3,7,8]	Frame difference	Description: Frames at times t and t – 1 are compared Advantages: simple and easy to implement processing Less detection accuracy Low to moderate processing High accuracy
[3,7]	Double differences	Description: frames at times t and t – 1 and frames at times t – 1 and t – 2 are compared Limitations: Sensitivity to light and structural changes Advantages: moderate detection accuracy
[9–12]	Combined methods	Description: the temporal differencing method, optical flow method and double background filtering (DBF) method and morphological processing methods are combined to achieve better performance Advantages: high accuracy Limitations: high computation overhead
This paper	MSSIM	Advantages: adaptive Less sensitive to structural, illumination or contrast change More accurate [18] Limitations: Slow at larger number of scales

3. A cooperative sensor activation (CSA) algorithm

Multiple sensors (Wireless cams and PIRs) cooperate through the central Smart Home Gateway (SHG) to ensure complete coverage of the monitored Smart Home Area (SHA):

1. The main camera at home entrance is activated by system start or by a PIR detecting a motion and the camera starts recording upon motion detection. Moving object is tracked and video is recorded as long as a motion is detected.
2. Once the moving object disappears from the viewing field of the main camera, the next camera is activated based on the active PIR's location.
3. Tracking continues and repeats.

4. SSIM AND MS-SSIM based motion detection

The most widely used metrics for measuring image quality are the Mean-Square Error (MSE) and the Peak-Signal-to-Noise Ratio (PSNR). Structural Similarity Index (SSI) is a new powerful quality metric which is based on the characteristics of the Human Visual System (HVS) in contrary to MSE and PSNR. Fig. 5 shows the diagram of SSMS which is based on modeling of image luminance, contrast and structure [13,14].

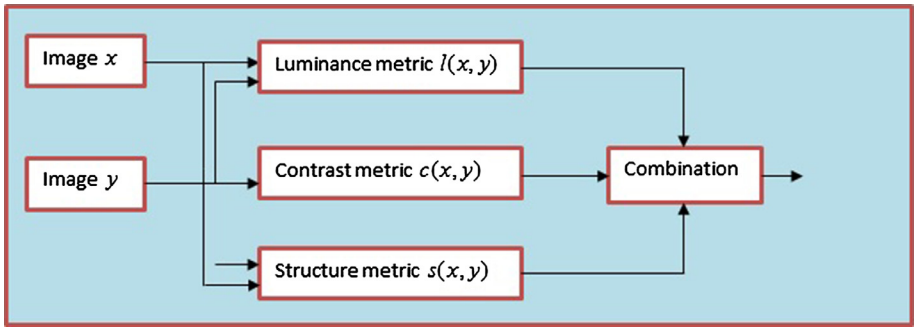

Figure 5. Structure Similarity Measurement System (SSMS) [15,16].

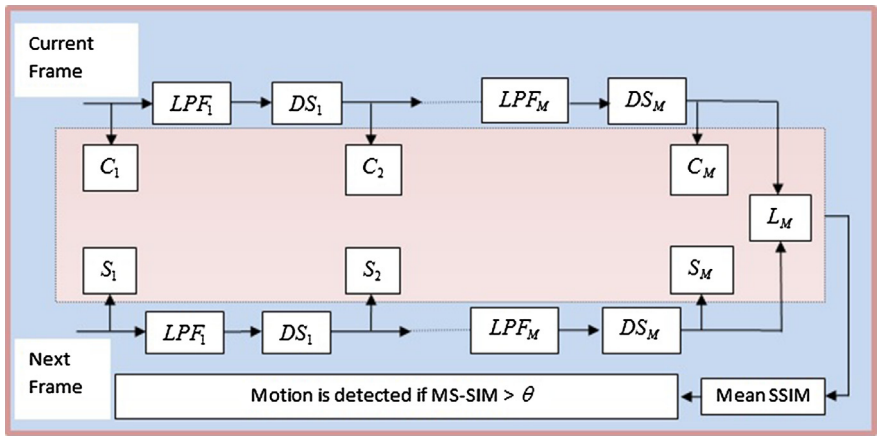

Figure 6. Architecture of the MS-SIM based motion detection system [16].

The SSI is defined in [13] as:

$$SSI(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{1}$$

where μ_x, μ_y,σ_x, and σ_y are the means and standard deviations of both the original and reference images respectively and C1 and C2 are constants. The three models considered in building the similarity index between the two images **x** and **y** are given by [13]:

$$\text{Luminance}: \ l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{2}$$

$$\text{Contrast}: \ c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{3}$$

$$\text{Structure}: \ s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{4}$$

where μ_x, σ_x^2, and σ_{xy} the mean of **x**, the variance of **x**, and the covariance of **x** and **y** respectively, while C1, C2, and C3 are constants given by $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, and $C_3 = C_2/2$. L is the dynamic range for the sample data, i.e. L = 255 for 8 bit gray level image and K1 << 1 and K2 << 1 are two scalar constants. Given the above measures the structural similarity can be computed in [13] as

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{5}$$

where α, β, and γ define the weight given to each model. Fig. 6 shows the architecture of a motion detection system that is based on the SSI.

The MS-SSIM quality metric, computes quality metrics at various scales, and combines them using according to the following equation [13]:

$$MSSSIM(\mathbf{x},\mathbf{y}) = [l_M(\mathbf{x},\mathbf{y})]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(\mathbf{x},\mathbf{y})]^{\beta_j} \cdot [s_j(\mathbf{x},\mathbf{y})]^{\gamma_j} \tag{6}$$

where M corresponds to the lowest resolution (i.e. the times of down samplings performed to reduce the image resolution), while j = 1 corresponds the original resolution of the image. The architecture of the motion detection system is shown in Fig. 3. Since the performance of the motion detection systems relies heavily on the distance between the vision system and the acquired scene, resolution of the analyzed video has a significant impact on motion detection results. The interaction between defect size and image resolution is also an important factor. Therefore, using the MS-SSIM metric renders itself a good adaptive measure for motion detection. Fig. 6 presents the architecture of a novel system for motion detection based on the Multi-scale Structural Similarity Index [15], where frames are progressively low pass filtered (LPF) and downscaled (DS).

5. Adaptive thresholding for triggering motion recording

Estimation of the appropriate threshold level for the SSI test is very critical to the success of the presented motion detection system. The system is configured to work for a specific environment by adaptively calculating the motion detection threshold based on the difference between two successive frames. During the configuration phase, the average similarity index of N = 100 pairs of two-successive frames is calculated together with the standard deviation. Then the similarity threshold is calculated according to the following equation:

$$\theta = \mu - 3\sigma \tag{7}$$

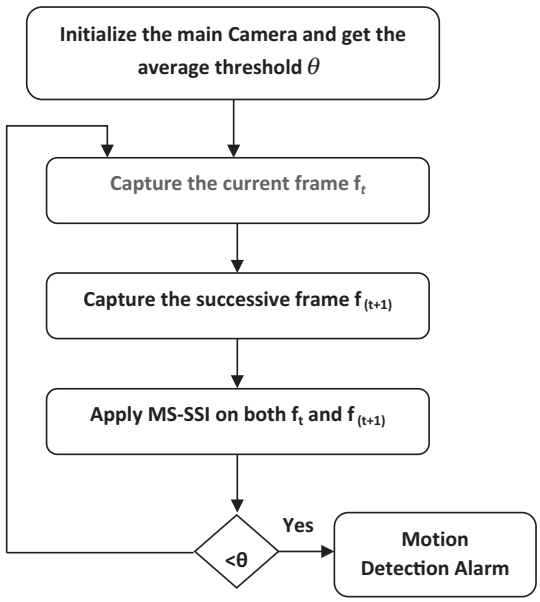

Figure 7. The MS-SSIM Based Motion Detection system.

$$\mu = \frac{\sum_{i=1}^{N} MSSIM_i}{N} \tag{8}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (MSSIM_i - \mu)^2}{N-1}} \tag{9}$$

During the implementation phase, if the similarity of two successive video frames is lower than θ, motion is detected and recording starts until the similarity exceeds the threshold.

6. Enhanced multi-scale structural similarity index based motion detection algorithm

In [17], MS-SIM has been used for motion detection in videos and its performance has been evaluated. Evaluation results showed that the MS-SIM based method outperformed the well-known motion detection techniques. Motion detection accuracy ranges between 0.985 and 0.995 in most experiments. The major advantages of the presented approach in [17] are: the higher motion detection accuracy and the fast processing speed. The problem

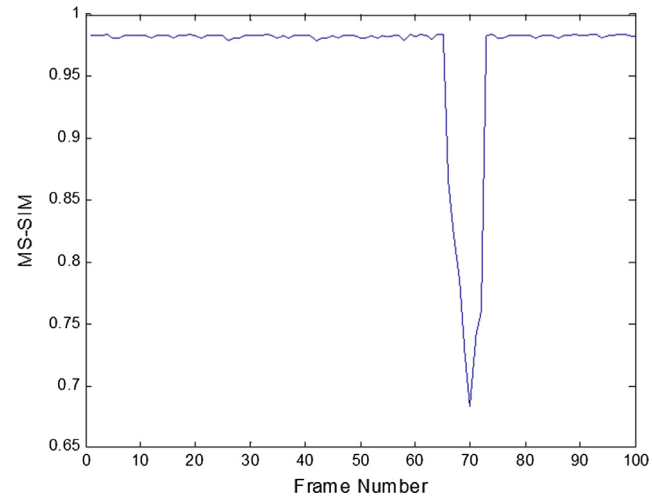

Figure 9. MS-SSIM index variation in a Motion Case.

which faces this method is the selection of the suitable threshold limit. In this paper, Section 5 introduced an adaptive method for threshold selection. In [17], a fixed threshold level has been used for the MS-SSIM based motion detection, which is very critical to the success of the presented detection system. The adaptive threshold selection presented in Section 5 solves this problem.

Define the size [NR × NC pixels] of each frame in the video sequence acquired by the camera, where NR is the vertical resolution and NC is horizontal resolution. Apply the following steps (Fig. 7):

1. Identify the adaptive threshold θ of the SSI's according to Section 5 in a still-stand acquired video sequence of 100 frames.
2. Calculate the MS-SSIM index of each two successive frames.
3. Repeat.
   If
   MS-SSIM > θ start video recording
   else
4. Stop video recording.

7. Performance evaluation

The performance of the MS-SMIM based motion detection algorithm has been evaluated by one of the authors in [17] by

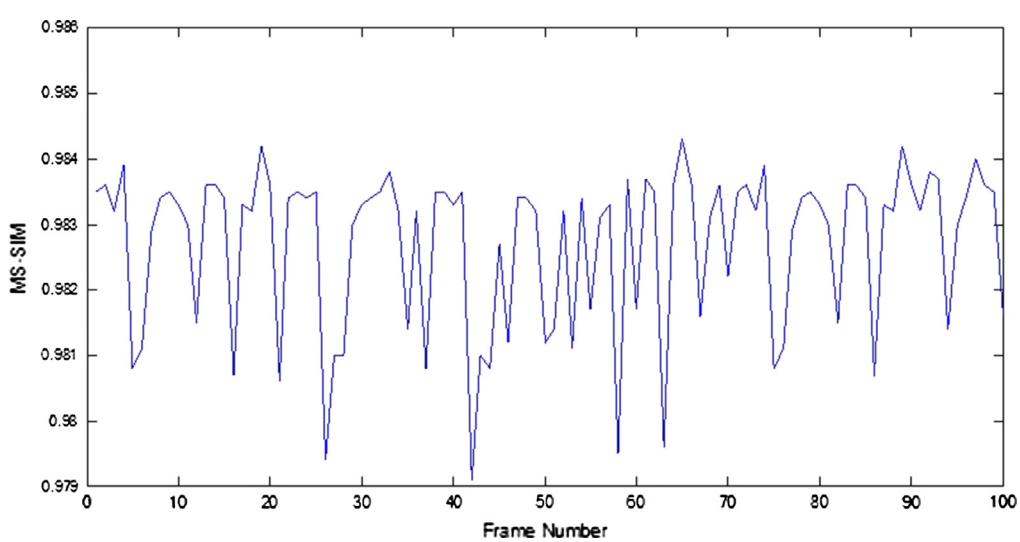

Figure 8. MS-SSIM index variation in a motion free video segment.

Please cite this article in press as: Abbas Khalaf H et al. Event triggered intelligent video recording system using MS-SSIM for smart home security. Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.10.001

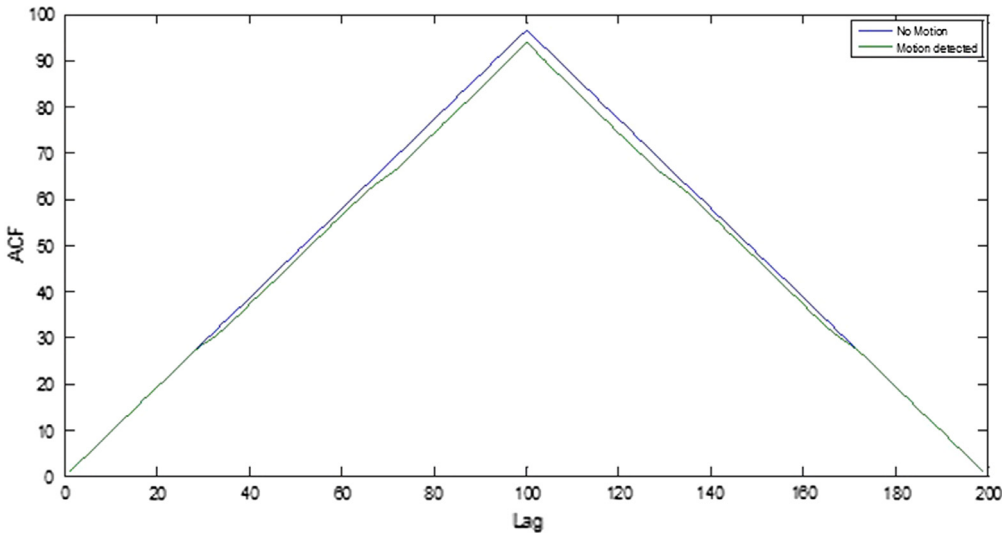

Figure 10. Autocorrelation of MS-SSIM index values for both motion and motion free cases.

Table 2
Comparison between fixed and adaptive threshold results.

Fixed			Adaptive	
θ	Motion segment size	Number of recorded segments	θ	Size
0.95	0.0	0	0.984	3.2 MB
0.96	0.0	0		
0.97	1.39 MB	2		
0.98	2.42 MB	1		
0.99	2.57 MB	1		

comparison with the Gaussian Mixture Model (GMM) in three main aspects: the memory requirements, computation time complexity and the accuracy. It has been shown that the proposed method which is based on the MS-SSIM requires less than 5 floating point operations for processing each pixel in a video frame, which implies less storage than the GMM-based method.

It has also been shown that the MS-SSIM is faster than the first stage of the GMM-based method, which clearly emphasizes the efficiency of the proposed method. The detection performance has been shown to excel that of GMM. The MS-SIM method provides very high specificity, accuracy and precision in the detection with a higher sensitivity and lower false rates in comparison with the GMM [17].

8. Results and discussion

Figs. 8 and 9 shows the variation of MS-SSIM index over time in the case of a motion free and motion cases in the smart home. To start event triggered recording based on the MS-SSIM index, a threshold value of MS-SSIM is adaptively calculated using Eq. (7) of Section 5. Fig. 5 shows small variations in the values of MS-SSIM index in a set of motion free video segment. The variation of the index results from slight change from one frame to the next as a result of light oscillation or non-significant shadows from outside.

Fig. 9 shows the case of large decrease in the MS-SSIM index values as a result of motion captured by the video camera.

The autocorrelation (ACF) of a recorded sequence of MS-SSIM index values is compared for motion and motion free cases. The Autocorrelation function $R(\tau)$ for N samples of the MS-SSIM index is calculated according to the following formula [18]:

$$R(\tau) = \frac{1}{N-\tau} \sum_{i=0}^{N-\tau-1} (x(i)x(i+\tau)) \tag{10}$$

Fig. 8 shows that the MS-SSIM index values are highly correlated in the case of a motion free video sequence and will show abrupt drop in the case of motion. Based on the data set of Fig. 10, the average MS-SSIM = 0.9826 and standard deviation = 0.0013 resulted in a threshold limit of MS-SSIM = 0.9786 according to Eq. (7). Table 2 shows the effect of thresholding approach on the system performance. It could be seen that a fixed threshold needs trial and error to specify a reasonable value and is mostly is not suitable for.

9. Conclusion

An intelligent system for automated video recording in smart homes is presented. We believe that the proposed sensor - collaboration strategy implemented in this paper together with motion detection activated recording presents many interesting opportunities for low-cost home security monitoring. The MS-SSIM resulted in a highly accurate event detection system. The event triggered video recording saves huge amounts of storage space and simplifies further video analysis. Future work will ensure the security of the system against hacking and intrusion.

References

[1] Widyawan Muhammad Ihsan Zul. Adaptive motion detection algorithm using frame differences and dynamic template matching method. In: The 9th international conference on ubiquitous robots and ambient intelligence (URAI 2012), November 26–28, 2012. Daejeon, Korea: Daejeon Convention Center (DCC); 2012.
[2] Spagnolo P, D'Orazio T, Leo M, Distante A. Moving object segmentation by background subtraction and temporal analysis. Image Vis Comput 2006;24:411–23.
[3] Mishra Sumita, Mishra Prabhat, Chaudhary Naresh K, Asthana Pallavi. A novel comprehensive method for real time video motion detection surveillance. Int J Scient Eng Res 2011;2(4).
[4] Tang Zhen, Miao Zhenjiang. Fast background subtraction using improved GMM and graph cut. In: Congress on image and signal processing, 2008, CISP'08. p. 181–5.
[5] Allili M, Auclair-Fortier M-F, Poulin P, Ziou D. A computational algebraic topology approach for optical flow. In: ICPR'02 proceedings of the 16th international conference on pattern recognition (ICPR'02) volume 1, Washington DC, USA.
[6] Jung Ho Gi, Suhr Jae Kyu, Bae Kwanghyuk, Kim Jaihie. Free parking space detection using optical flow-based Euclidean 3D reconstruction. In: Proceedings of the IAPR conference on machine vision applications (IAPR MVA 2007), Tokyo, Japan. p. 16–8.

[7] Yu Zhen, Chen Yanping. A real-time motion detection algorithm for traffic monitoring systems based on consecutive temporal difference. In: Proceedings of the 7th Asian control conference, Hong Kong, China, August 27–29.
[8] Singla Nishu. Motion detection based on frame difference method. Int J Inform Comput Technol 2014;4(15):1559–65. ISSN 0974-2239.
[9] Kenchannavar HH, Patkar Gaurang S, Kulkarni UP, Math MM. Simulink model for frame difference and background subtraction comparison in visual sensor network. In: 2010 The 3rd international conference on machine vision (ICMV 2010), Hongkong China.
[10] Murali S, Girisha R. Segmentation of motion objects from surveillance video sequences using temporal differencing combined with multiple correlation. In: 2009 Sixth IEEE international conference on advanced video and signal based surveillance, Genova, Italy. p. 472–7.
[11] Yong Ching Yee, Sudirman Rubita, Chew Kim Mey. Motion detection and analysis with four different detectors. In: 2011 Third international conference on computational intelligence, modelling and simulation, Langkawi. p. 46–50.
[12] Lu Nan, Wang Jihong, Wu QH, Yang Li. An improved motion detection method for real-time surveillance. Int J Comput Sci 2008;1(6).
[13] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.
[14] Shang X. Structural similarity based image quality assessment: pooling strategies and applications to image.
[15] Wang Z, Bovik AC. Modern image quality assessment. Morgan and Claypool Publishers; 2006.
[16] Tolba AS, Raafat Hazem M. Multiscale image quality measures for defect detection in thin films. Int J Adv Manuf Technol 2015.
[17] Abdel-Salam Nasr M, Al Rahmawy Mohammed F, Tolba AS. Multi-scale structural similarity index for motion detection. J King Saud Univ – Compute Inform Sci. available online at: <www.ksu.edu.sa> <www.sciencedirect.com>.
[18] Orfanidis SJ. Optimum signal processing. An introduction. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall; 1996.
[19] Chauhan Abhishek Kumar, Krishan Prashant. Moving object tracking using gaussian mixture model and optical flow. Int J Adv Res Comput Sci Software Eng 2013.

Please cite this article in press as: Abbas Khalaf H et al. Event triggered intelligent video recording system using MS-SSIM for smart home security. Ain Shams Eng J (2016), http://dx.doi.org/10.1016/j.asej.2016.10.001