Journal of King Saud University – Computer and Information Sciences (2017) 29, 399–409

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

Multi-scale structural similarity index for motion detection

M. Abdel-Salam Nasr [a], Mohammed F. AlRahmawy [b,*], A.S. Tolba [b]

[a] Computer Science Department, Faculty of Computers and Information Sciences, IDA, Ministry of Trade and Industry, Egypt
[b] Computer Science Department, Faculty of Computers and Information Sciences, Mansoura University, Egypt

Received 8 October 2015; revised 9 February 2016; accepted 25 February 2016
Available online 2 April 2016

KEYWORDS

Motion detection;
Multi-scale structural similarity index;
Dynamic template matching

Abstract The most recent approach for measuring the image quality is the structural similarity index (SSI). This paper presents a novel algorithm based on the multi-scale structural similarity index for motion detection (MS-SSIM) in videos. The MS-SSIM approach is based on modeling of image luminance, contrast and structure at multiple scales. The MS-SSIM has resulted in much better performance than the single scale SSI approach but at the cost of relatively lower processing speed. The major advantages of the presented algorithm are both: the higher detection accuracy and the quasi real-time processing speed.

1. Introduction

Fighting terrorism is becoming imperative and is driving researchers toward designing robust video surveillance systems. Event detection and analysis is one of the major applications of video surveillance. A first step in video event analysis is motion detection. A motion detection algorithm should avoid releasing false alarms by considering image luminance, contrast and structure at multiple scales. Accurate motion detection algorithms are the key components of video surveillance systems. It can be applied as well in some interesting applications such as the automatic light control and automatic door opening in smart homes and work areas resulting in efficient energy use.

Intrusion detection in secure areas is a common application of motion detection. A very important application of motion detection is to limit video recording storage requirements to times of important events by starting the recording only at the start of an important event. This on demand recording helps avoiding continuous video recording of static scenes and speeds up video event analysis. Motion detection helps focusing on processing and storage of relevant events.

In this study, a new algorithm is presented which showed promising results compared with MSE (Winkler, 2005; Susstrunk and Winkler, 2004; Wang and Bovik, 2002) and Dynamic Template Matching DTM (Martinez-Martin and del Pobil, 2012a, b; Widyawan et al., 2012) based algorithms. The algorithm captures the video of a scene and detects motion by comparing successive frames in the video sequence based on

* Corresponding author.
E-mail addresses: mohtrue.net@gmail.com (M. Abdel-Salam Nasr), mrahmawy@mans.edu.eg (M.F. AlRahmawy), ast@astolba.com (A.S. Tolba).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2016.02.004
1319-1578

the modeling of image luminance, contrast and structure at multiple scales.

The paper is organized as follows, in Section 2 some of the related work is presented; then, an overview of the basic and the proposed structural similarity index mechanisms for motion detection is presented. In Section 4, the performance of the proposed mechanism is evaluated using a set of experiments on both offline and online videos. Then, in Section 5, the efficiency of the proposed mechanism is proved by comparing it with the well-known GMM method of motion detection both analytically, by analyzing their memory requirements complexity and execution complexity, and experimentally by comparing their efficiency through a set of performance metrics.

2. Related work

Motion detection starts from a given frame as a reference and subsequent frame is compared with it; then the subsequent frame becomes a reference frame and the process is repeated with next frames. Mishra et al. (2011) discussed three commonly used methods to detect a motion: background subtraction, optical flow and temporal differences. Background subtraction was used by Spagnolo et al. (2006), Tang and Miao (2008), Li and Cao (2010) and it depends on a comparison of an image with a static reference image. The optical flow method specifies how much each pixel of the image moves between adjacent frames, this method may require additional hardware to support the performance and monitoring of systems (Allili et al., 2002; Jung et al., 2007). The temporal difference method relies on comparing consecutive frames by analyzing all frame pixels (Yu and Chen, 2009), e.g. by applying the concept of Sum of Absolute Difference (SAD), where SAD is used to determine whether there is a movement within an image pair (Kenchannavar et al., 2010).

There is a lot of research work based on the above methods and combinations of them. Kenchannavar et al. (2010) described an algorithm combining background subtraction and frame differences for motion detection. Zheng et al. (2009), Murali and Girisha (2009), Fang et al. (2009), used frame differences coupled with an adaptive threshold setting and statistical correlation to analyze the temporal differences in some of the image frames.

Other methods for motion detection exist in the literature. For example, Yong et al. (2011) studied four methods for motion detection: frame differences, background subtraction, pixelate filter, and blob counter. Li and Cao (2010) used Support Vector Machines (SVM) for motion detection and segmentation of moving objects. Yokoyama et al. (2009) applied the concept of vectors to movement detection by comparing multiple frames and marking the points of difference among the frames. An advantage of this method is that it yields information about the direction of object movement. Kameda and Minoh (1996) used the double difference technique for motion detection. Double differences are conducted by comparing two successive frames at times t and $t-1$ and then performing a second comparison between the frames at times $t-1$ and $t-2$. In contrast, Collins and et al. (2000) reported a video surveillance and monitoring system which depends on comparisons between the image t with the image of $t-1$, and the image t with the image of $t-2$.

From the above, we can see that there are many research works done on the motion detection which uses different methods and have achieved different results.

3. Structural similarity index mechanisms for motion detection

The SSI measurement system (Wang and Bovik, 2002) is based on modeling of image luminance, contrast and structure. The MS-SSIM is an extension of the SSI system that achieves better accuracy than the single scale SSI approach but at the cost of relatively lower processing speed. However, the computations required for the MS-SSIM does not require such large computational time required by other efficient Statistical Learning Algorithms (Avcıbas et al., 2002) since it requires complex calculations. Our system applies the MS-SSIM for motion detection in videos; either on online videos directly captured from camera, or on recorded video stored in a file. Any video is sequenced into frames and successive frames are compared with each other and if a difference is detected, an alarm is released.

The proposed approach relies on a new algorithm which helps local adaptation of the multi-scale structural similarity measure for motion detection in videos. In the next sections we present the basic SSI measurement system as it is the base of the used algorithm, the MS-SSIM; next we present MS-SSIM itself. Then, we present our proposed algorithm for motion detection using the MS-SSIM.

3.1. SSI measurement systems

The Structural Similarity Index (SSI) measurement, shown in Fig. 1, is based on modeling image luminance, contrast and structure.

Mathematically, the SSI is defined by Wang and Bovik (2006) as:

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

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of both the original and reference images respectively, C_1 and C_2 are constants. The three models considered in building the similarity index between the two images x and y are given by Lavielle (1999), Wang et al. (2003), and Wang et al. (2004):

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

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

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

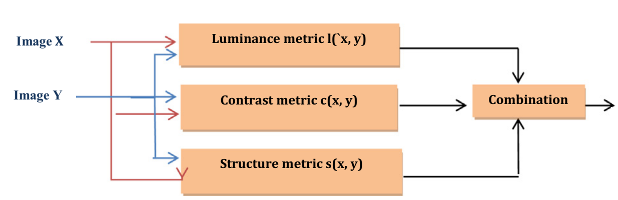

Figure 1 Structure similarity measurement system (Tolba and Raafat, 2015).