PASSIVE APPROACH FOR VIDEO FORGERY DETECTION AND LOCALIZATION

Cheng-Shain Lin and Jyh-Jong Tsay

Department of Computer Science and Information Engineering
National Chung Cheng University
Chiayi 621, Taiwan
{lchh95p, tsay}@cs.ccu.edu.tw

ABSTRACT

In this paper, we present a passive approach for effective detection and localization of forgery from video sequences. Our approach analyzes spatio-temporal slices from 3-D video volumes to detect and localize regions tampered by temporal copy-and-paste and texture synthesis. Experiment shows that the proposed approach outperforms previous approaches, and can effectively detect and localize tampered regions.

KEYWORDS

Passive forgery detection; region-level forgery; temporal copy-and-paste; texture synthesis; spatio-temporal slices

1 INTRODUCTION

Visual imagery has been widely used to provide essential evidence in many diverse areas, ranging from mainstream media, journalism, and scientific publication, to medical imaging, criminal investigations, and surveillance systems, to name a few. While we have historically had confidence with the integrity and authenticity of visual imagery, such trust has been gradually lost. With the rapid growth of digital devices and multimedia editing technology [1], [2], it has become easier than ever to produce and modify digital video with increasing sophistication. Doctored video are very difficult, if not impossible, to identify through visual examination. Therefore, digital video forensics, which aims to verify the trustworthiness of digital video, has become an important and exciting field of recent research.

Over the past few years, many passive approaches have been proposed, which can be roughly classified into four categories [3], pixel-based, format-based, camera-based, and geometric-based. Pixel-based approaches, which examine pixel level anomalies caused by tampering, such as correlations between frames arising from duplicate frames [4], [5], and inpainted regions [6]. Format based approaches exploit the unique properties of video compression, such as periodic properties [7], [8] and blocking artifacts [9] in MPEG-1 and MPEG-2 video. Camera-based approaches analyze the specific sensor artifacts caused by components in the imaging pipeline, such as sensor noise [10], [11] and interlaced scanning [12]. Geometric-based approaches inspect the geometric properties of objects and their positions relative to the camera [13].

This paper presents an effective and robust approach based on analysis of spatio-temporal slices extracted from 3-D video volumes. The experiment results show that the proposed approach outperforms previous approaches [6], and can effectively detect and localize areas tampered by temporal copy-and-paste and texture synthesis.

The remainder of this paper is organized as follows. Section 2 briefly overviews the problem, and sketches our main approach. Section 3 presents the details of the proposed approach. Section 4 presents the experimental results. Finally, section 5 gives conclusions.

2 OVERVIEW

In this section, we briefly overview the problem of tampering detection in video sequences, and sketch our proposed approach.

The spatio-temporal slice technique is widely used in analyzing the spatio and temporal relationships of video sequences [14], and is commonly used in various research areas, such as motion analysis and segmentation [15], human gait analysis [16], and video gradual transition detection [17]. Figure 1 shows a video sequence arranged as a 3D volume, with (x, y) and t representing image dimensions and temporal dimension, respectively. Fig. 2(a) and (b)

illustrates two *XT* spatio-temporal slices extracted at position y_1 and y_2 from Fig. 1, respectively.

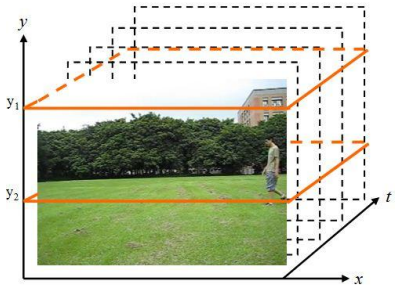


Fig. 1. Two spatio-temporal slices taken from a 3D video volume along the temporal dimension.

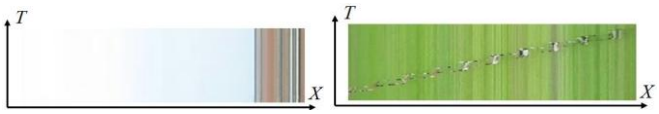


Fig. 2. Examples of spatio-temporal slices extracted from the 3D volume Fig. 1: (a) the *XT* spatio-temporal slice extracted at location y_1 contains several spatially uniform color-texture regions; (b) the *XT* spatio-temporal slice extracted at location y_2 contains a crisscrossing pattern.

Extraction and analysis of the spatio-temporal correlation have played an essential role in detection and localization of tampered areas for video sequences. Fig. 3 gives an illustration of temporal copy-and-paste tampering, which is performed to copy an undamaged region from the nearest frame f_{D-1}, and paste it to region Ω in frames f_D, …, f_{D+k} to form a tampered video. It has been noticed that, in order to maintain temporal coherence between successive frames to create a plausible tampered video, the tampered region is often replaced with similar areas of the closest frame [2], and consequently it will result in high coherence pixels in tampered region. In addition, the examplar-based texture synthesis technique [1] is commonly used to inpaint each tampered frame independently, and it will result in low coherence pixels in tampered region.

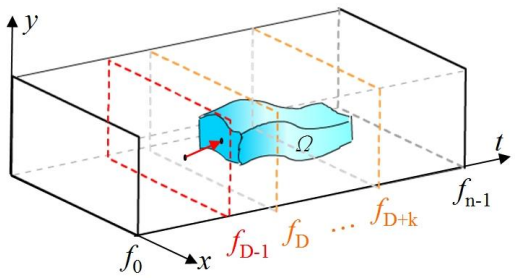


Fig. 3. An illustrated example: the tampered 3D hole (Ω) is coping the untampered area (white area) of the nearest frame and pastes it to the region Ω.

Spatio-temporal artifacts caused by spatio-temporal incoherence in tampered regions give crucial evidence for detection and localization of tampered regions. In this paper, we present an effective approach for detection and localization of tampered regions in video sequences manipulated by temporal copy-and-paste and texture synthesis. Our approach is based on analysis of spatio-temporal artifacts resulting from spatio-temporal incoherence in tampered regions. The proposed approach consists of two major steps: 1) spatio-temporal artifact analysis; and 2) detection result refinement. In the first step, spatio-temporal artifacts are extracted and analyzed. Regions with abnormally high similarity or inconsistency are detected as tampered regions. We integrate techniques from salient region detection and seed region growing segmentation methods to obtain a reliable map of the whole spatio-temporal slice artifacts (*WSTSA*). We finally perform detection result refinement which uses the *WSTSA* map to compare each spatio-temporal slice artifact for positioning of the actual tampered slices, and relocation of the artifacts. Consequently, the 2D refinement results can be converted into the general 3D video volume. The details of each step of the proposed approach are given in subsequent sections.

3 THE APPROACH

In this section, we present the details of the proposed approach for detecting video forgery. The proposed approach consists of two processes: spatio-temporal artifact analysis and refinement. The detail of the proposed approach is described below.

3.1 Spatio-temporal artifact analysis

Let a set of spatial-temporal slices XT_i, $i \in [0,...,M-1]$ is obtained by sampling a 3D video volume at different *y*-axis over time using the spatio-temporal slicing method (described in more details in section 2). The $XT_i(x, t)$ is the intensity value of a slice pixel (x, t), i.e. the pixel (x, i, t) in the 3D volume. To examine the correlation of spatio-temporal variations of 2D slices, we apply gradient based sobel filter to each XT_i and compute a matrix MXT_i of magnitudes which is defined as follows.

$$MXT_i(x,t) = \sqrt{(XT_i|_{(x,t)} * H)^2},\ i = \{0,...,M-1\}, \qquad (1)$$

where H is the sobel horizontal filter, and * is the convolution operator. It should be noted that, according to our observation, when the magnitude of $MXT_i(x,t)$ is very small, it has possibly suffered high coherence tampering; by contrast, when the magnitude of $MXT_i(x,t)$ is very high, it may have suffered low coherence tampering.

For each magnitude map MXT_i, we compute two different maps BM_{HT_i} and BM_{LT_i}, one for detecting unnaturally high coherence tampering and one for detecting anomalous low coherence tampering, by comparing the values in MXT_i with predefined threshold value, respectively. We compute the following two different binary maps BM_{HT_i} and BM_{LT_i} which are defined below.

$$\begin{cases} BM_{HT_i}(x,t) = 1, \text{ if } MXT_i(x,t) \le \Theta_1, \\ BM_{HT_i}(x,t) = 0, \text{ oherwise,} \end{cases} \quad (2)$$

$$\begin{cases} BM_{LT_i}(x,t) = 1, \text{ if } MXT_i(x,t) \ge \Theta_2, \\ BM_{LT_i}(x,t) = 0, \text{ oherwise,} \end{cases} \quad (3)$$

where Θ_1 and Θ_2 are the threshold values for detection of high coherence and extreme low coherence malicious tampering in each MXT_i. In this paper, Θ_1 and Θ_2 are empirically set to 3 and twice the mean value of MXT_i, respectively. Note that because the BM_{HT_i} map often contains some small regions, this paper thus uses a morphological erosion operator to remove small regions [19]. Fig. 4(b) shows that the morphological erosion process removes most of the small regions, and will be used to obtain the reliable spatio-temporal slice artifacts.

(a) (b)

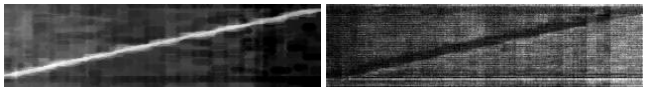
Fig. 4. (a) The original binary map of $BM_{HT_{100}}$, (b) the result of morphological erosion process.

To further expose the spatio-temporal artifacts for a video sequence, we compute the conspicuous map by the accumulation of all the values of BM_{HT_i} and BM_{LT_i}, respectively. The conspicuous map CSM_{HT} is defined as follows. Another conspicuous map CSM_{LT} is defined similarly.

$$CSM_{HT}(x,t) = \sum_{i=1}^{M} BM_{HT_i}(x,t). \quad (4)$$

Note that as in Fig. 5(a), when a video volume has been tampered by temporal copy-and-paste tampering, CSM_{HT} will have visually suspicious artifact with high intensity value. We next explain how to extract the spatio-temporal slice artifact.

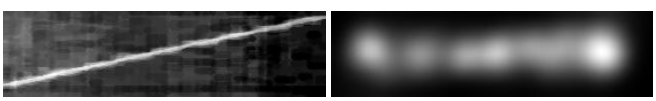
Figure 5 shows the conspicuous maps CSM_{HT} and CSM_{LT} for the tampered video sequence "person walking," which has been tampered by temporal copy-and-paste technique [2]. Obviously, as shown in Fig. 5(a), the conspicuous maps CSM_{HT} have a visually suspicious spatio-temporal slice artifact with a high intensity value. By contrast, the conspicuous maps CSM_{LT} has no significant suspicious region, as shown in Fig. 5(b). More specifically, we expect that the tampered region of video can be determined by the conspicuous map.



(a) (b)

Fig. 5. The conspicuous map reveals spatio-temporal artifacts: (a) the CSM_{HT} map accumulate from BM_{HT_i}, (b) the CSM_{LT} map accumulate from BM_{LT_i}.

To further reduce noise influence for the entire spatio-temporal slice artifact extraction, we adopt the simple saliency-guided region segmentation technique to extract the whole spatio-temporal slice artifact (*WSTSA*). The saliency-guided region segmentation technique contains the focus of attention area detection and seed region growing segmentation processing procedures. For focus of attention area detection, we use the human visual attention model, which is able to detect the significant region in order to detect the focus of attention area in the conspicuous map. In this study, the visual attention model, as proposed by Itti et al. [18], is employed for computing the saliency map of the conspicuous map CSM_{HT}, which can be used after thresholding to obtain the focus of attention area (FoA). To reduce the noise interference in the segmentation processing procedure, the seeded region growing method [19] is used to segment the CSM_{HT} map, and its seed point location can be set in the location of the maximum value of the CSM_{HT} map in FoA. Finally, the segmentation result of CSM_{HT} is also known as the whole spatio-temporal slice artifact map (*WSTSA*), as shown in Fig.6(d).



(a) (b)

(c) (d)

Fig. 6. The spatio-temporal artifact extraction: (a) the conspicuous map, (b) the saliency map of (a), (c) seed point (denoted by red color) is set at the FoA area, (d) segmentation result.

3.2 Detection result refinement

The main idea for result refinement is to use the *WSTSA* map to identify the most similar spatio-temporal artifact among the spatio-temporal slices, and then relocalize the location of the artifact, as shown in Fig. 7.

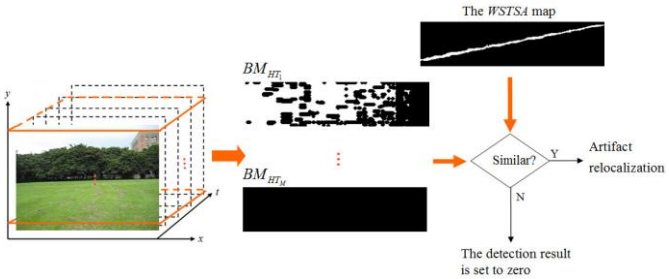


Fig. 7. The proposed detection result refinement approach.

To eliminate false detections, we first look for the actual tampered slices. We compute the similarities and differences of BM_{HT_i} and the *WSTSA* map. A slice is deemed a tampered slice if it has sufficiently high similarity and low difference. The similarity is computed as follows.

$$sim\_score = \sum_{x=1}^{N}\sum_{t=1}^{L} BM_{HT_i}(x,t) \bullet WSTSA(x,t), \quad (5)$$

where $\bullet$ denotes the logical AND operator. The threshold TH_2 for similarity comparison is defined as $w_1 \times (\sum_{x=1}^{N}\sum_{t=1}^{L} WSTSA(x,t))$, where w_1 is empirically set to 0.4.

The dissimilarity is defined as follows.

$$dsim\_score = \sum_{x=1}^{N}\sum_{t=1}^{L} BM_{HT_i}(x,t) \oplus WSTSA(x,t), \quad (6)$$

where $\oplus$ denotes the logical XOR operator. The threshold TH_3 for difference comparison is empirically set to be half of the values in *WSTSA*.

For each slice BM_{HT_i} which satisfies both similarity and difference criteria, we perform artifact re-localization as follows.

$$FDWSTSA_i(x,t) = BM_{HT_i}(x,t) \bullet WSTSA(x,t), \quad (7)$$

where $\bullet$ denotes the logical AND operator. If BM_{HT_i} does not satisfy both criteria, all $FDWSTSA_i(x,t)$ are set to 0. Consequently, the refined detection result is defined as follows.

$$FDWSTSA_i(x,t) = \begin{cases} 0, & \text{if } BM_{HT_i} \text{ fails to satisfy} \\ & \text{similarity and difference criteria,} \\ BM_{HT_i}(x,t) \bullet WSTSA(x,t), & \text{otherwise.} \end{cases} \quad (8)$$

Finally, we combine the resulting $FDWSTSA_i$ of 2D slices to form a resulting 3D volume.

4 EXPERIMENTAL RESULT

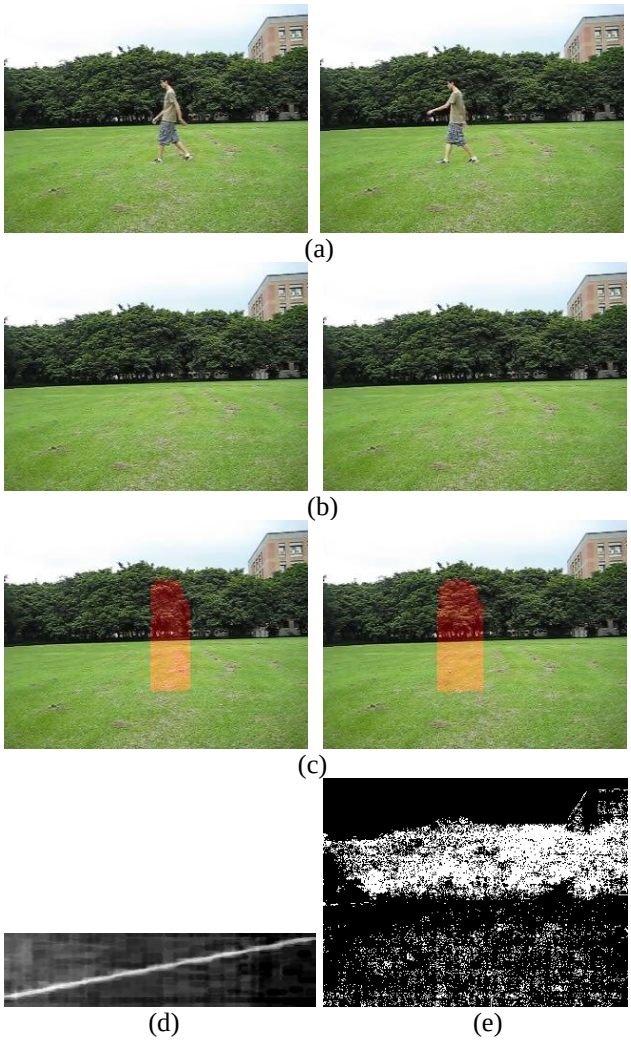
In the experimental study, we compare our approach with the ADI approach, which is proposed in [6]. We conduct experiment over video sequences manipulated by temporal copy-and-paste inpainting [2] and texture synthesis [1]. The temporal copy-and-paste inpainting [2] is mainly to fill in the unknown region left by the removal of large objects, while maintaining the temporal coherence between successive frames to generate the forged tampered video. The texture synthesis scheme, as proposed in [1], fills in a region from sample textures, which is one of the state-of-the-art image inpainting schemes. This paper uses it to simulate tampering processes; when the frame tampered regions are affected by other regions, it will affect the temporal correlation of successive frames; hence, it is used to simulate the low coherence tampering technique.

We present an experiment over test video "Person Walking" with 75 frames each of size 360×240. We carried out the experiment over two tampering methods: temporal copy-and-paste and texture synthesis. The experiment is run on a PC with an Intel Core i7-920 CPU 2.67GHz and 4G RAM, using the Matlab software development tool.

Fig. 8(a) gives some snapshots of the original video frames, and Fig. 8(b) gives the corresponding frames tampered by temporal copy-and-paste[2]. In this sequence, we removed the person in frames 5 to 71, and fill in the left region by temporal copy-and-paste. Fig. 8(c) shows the regions detected and localized by our approach. The result shows that our approach can effectively detect and localize the tampered regions. Fig. 8(d) is the corresponding conspicuous map computed to identify tampered regions in our approach. Fig. 8(e) from [6] gives the intensity value of the detection result by ADI approach. The result shows that although the ADI approach can detect tampered

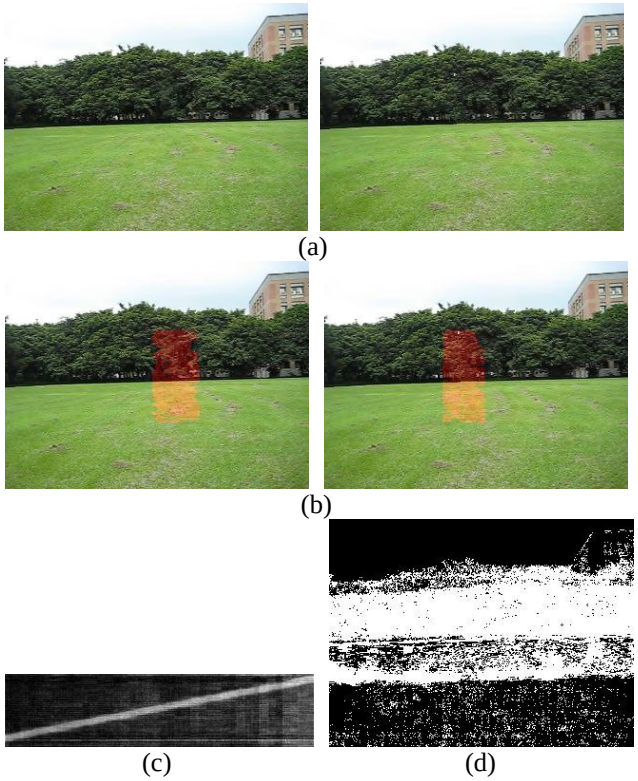
frames, it cannot effectively localize tampered regions. Note that the ADI approach is vulnerable to the impact of noise and generates large amount of false detection.

Figure 9 gives the result for the first sequence tampered by texture synthesis[1]. Fig. 9(a) gives the corresponding tampered frames. Note that texture synthesis generates abnormally low coherency between successive frames. Fig. 9(b) shows that the regions detected by our approach are very close to the tampered regions. Fig. 9(c) gives the corresponding conspicuous map. Fig. 9(d) shows that although the ADI approach can detect the tampered frames, it cannot effectively localize the tampered area in each frame.



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

Fig. 8. Detection results of the first sequence with temporal tampering. (a) The original sequence, from left to right: frames 33 and 39; (b) temporal tampering results from [2]; (c) the region detected by the proposed approach; (d) the conspicuous map; (e) the result of ADI approach [6].



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

Fig. 9. Detection results of the first sequence with texture synthesis. (a) Frames tampered by texture synthesis[1]; (b) the region detected by the proposed approach; (c) the conspicuous map; (d) the result of ADI approach [6].

5 CONCLUDING REMARKS

In this paper, we have presented an effective passive approach for detection and localization of forgery in video sequences. Our approach is based on analysis of spatio-temporal artifacts extracted from 2D spatio-temporal slices. We have developed an efficient method to produce a map, which strengthens spatio-temporal artifacts and relocalizes the tampered area in the refinement process to reduce false detection. We have conducted experiment to compare our approach with the ADI approach [6] over temporal copy-and-paste and texture synthesis tampering methods. The experiment shows that our approach can effectively detect and localize tampering areas in the test video presented in this paper, and improves the result of the ADI approach which suffers high false detection rates in localization of tampered regions. Notice that similar experimental results have been achieved for other video sequences which are presented in this paper due to space limitation. In the future, we will continue to improve our approach, and study how to apply the main idea developed in this paper to other problems, such as detection of the forgery in

which a moving object is replaced by another moving object.

6 REFERENCES

1. Criminisi, A., P´erez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. on Image Processing, vol. 13, no. 9, pp. 1200--1212 (2004).
2. Patwardhan, K.A., Sapiro, G., Bertalmio, M.: Video inpainting under constrained camera motion. IEEE Trans. on Image Processing, vol. 16, no. 2, pp. 545--553 (2007).
3. Farid, H.: A survey of image forgery detection. IEEE Signal Processing Magazine, vol. 2, no. 6, pp. 16--25 (2009).
4. Wang, W., Faird, H.: Exposing digital forgeries in video by detecting duplication. In: Proc. of ACM conf. on Multimedia and Security Workshop (2007).
5. Lin, G.S., Chang, J.F., Chuang, C.H.: Detecting frame duplication based on spatial and temporal analyses. In: Proc. of IEEE conf. on Computer Science & Education, pp. 1396--1399 (2011).
6. Zhang, J., Su, Y., Zhang, M.: Exposing digital video forgery by ghost shadow artifact. In: Proc. of ACM conf. on Multimedia in Forensics, Security and Intelligence, pp. 49--53 (2009).
7. Wang, W., Faird, H.: Exposing digital forgeries in video by detecting double MPEG compression. In: Proc. of ACM conf. on Multimedia and Security, pp. 37--47 (2006).
8. Wang, W., Faird, H.: Exposing digital forgeries in video by detecting double quantization. In: Proc. of ACM conf. on Multimedia and Security, pp. 39--48 (2009).
9. Sun, T.F., Wang, W., Jiang, X.H.: Exposing video forgeries by detecting MPEG double compression. In: Proc. of IEEE conf. on Acoustics, Speech, and Signal Processing, pp. 1389--1392 (2012).
10. Kobayashi, M., Okabe, T., Sato, Y.: Detecting forgery from static-scene video based on inconsistency in noise level functions: IEEE Trans. on Information Forensics and Security, vol. 5, no. 4, pp. 883--892 (2010).
11. Hsu, C.C., Hung, T.Y., Lin, C.W., Hsu, C.T.: Video forgery detection using correlation of noise residue. In: Proc. of IEEE Int. Conf. on Multimedia Signal Processing, pp. 170--174 (2007).
12. Wang, W., Farid, H., Exposing digital forgeries in interlaced and deinterlaced video. IEEE Trans. on Forensics Security, vol. 2, no. 3, pp. 438--449 (2007).
13. Conotter, V., O'Brien, J.F., Faird, H.: Exposing Digital Forgeries in Ballistic Motion. IEEE Trans. on Information Forensics and Security, vol. 7, no. 1, pp. 283--296 (2012).
14. Adelson, E.H., Bergen, J.R.: Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America, A: Optics and Image Science, vol. 2, no. 2, pp.284--299 (1985).
15. Ngo, C.W., Pong, T.C., Zhang, H.J.: Motion analysis and segmentation through spatio-temporal slices processing. IEEE Trans. on Image Processing, vol. 12, no.3, pp. 341--355 (2003).
16. Niyogi, S.A., Adelson, E.H.: Analyzing and recognizing walking figures in XYT. In: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 469--474 (1994).
17. Ngo, C.W., Pong, T.C., Chin, R.T.: Detection of gradual transitions through temporal slice analysis. In: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 1036--1041 (1999).
18. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254--1259 (1998).
19. Gonzalez, R.C., Woods, R.E., Digital Image Processing. 3rd Edition. New Jersey: Prentice Hall (2008).