We are IntechOpen, the world’s leading publisher of Open Access books. Built by scientists, for scientists.

- **6,600** Open access books available
- **177,000** International authors and editors
- **195M** Downloads
- **154** Countries delivered to
- **TOP 1%** of most cited scientists
- **12.2%** Contributors from top 500 universities

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected. For more information visit www.intechopen.com
Chapter

Change Point Detection-Based Video Analysis

Ashwin Yadav, Kamal Jain, Akshay Pandey, Joydeep Majumdar and Rohit Sahay

Abstract

Surveillance cameras and sensors generate a large amount of data wherein there is scope for intelligent analysis of the video feed being received. The area is well researched but there are various challenges due to camera movement, jitter and noise. Change detection-based analysis of images is a fundamental step in the processing of the video feed, the challenge being determination of the exact point of change, enabling reduction in the time and effort in overall processing. It is a well-researched area; however, methodologies determining the exact point of change have not been explored fully. This area forms the focus of our current work. Most of the work till date in the area lies within the domain of applied methods to a pair or sequence of images. Our work focuses on application of change detection to a set of time-ordered images to identify the exact pair of bi-temporal images or video frames about the change point. We propose a metric to detect changes in time-ordered video frames in the form of rank-ordered threshold values using segmentation algorithms, subsequently determining the exact point of change. The results are applicable to general time-ordered set of images.

Keywords: change point detection, time-ordered images, difference image, threshold values, segmentation algorithms

1. Introduction

Intelligent video surveillance involves automated extraction of information related to an object or scene of interest, including detection, localization, and tracking amongst other applications. One of the earliest comprehensive efforts in this regard was undertaken by Robert T Collins et al. [1] as part of a Defence Advanced Project Research Agency (DARPA) Video Surveillance and Monitoring (VSAM) project. The three fundamental methods tested for moving object detection were the background subtraction, optical flow and temporal differencing method. Due to the limitation of individual methods, hybrid schemes involving combination of individual methods were tested. Adaptive background subtraction was combined with the three-frame difference method in order to overcome the limitation of either method. The frame difference method it may be noted is a simple technique but, however, suffers from the limitation that the complete shape of the detected object cannot be extracted
precisely. The frame differencing method and in general temporal differencing make use of a static or dynamic threshold value in order to determine a change or no change scenario. This provides us a key to development of threshold as possible metric for our current work. Change detection (CD) is related to the fundamental task of object detection moving or static insofar as that it enables one to cull out relevant images or frames from a stack. Thus, the search space in scene analysis as part of the task for an image analyst gets reduced. This aspect is highlighted in the work by Huwer on adaptive CD for real-time surveillance applications [2]. CD enables one to detect viable changes, which can then be inputs for the subsequent object detection or tracking task. CD may be considered as an elementary stage in the video analytics framework entailing segmenting a video frame into the foreground and background. This may be considered a simple task but is an important precursor to further high-end processing. A most comprehensive recent review on a deep learning framework-based CD has been carried out by Murari Mandal et al. [3, 4]. Various applications of CD as part of video analysis including video synopsis generation, anomaly detection, traffic monitoring, action recognition, and visual surveillance have been covered as part of the study.

“CD is the process of identifying differences in the state of an object or phenomenon by observing it at different times.” Singh [5, 6]. This standard definition of the CD process though applied to the context of remote sensing images articulates the objective and purpose clearly insofar as even video surveillance is concerned. The objective is to detect the relevant change as part of the video surveillance in form of the object or activity (phenomenon) of interest. Considering the fact that today the quantum of data in form of the video feed to be analysed by the image analyst has increased vastly in recent times, there is scope for automation in the analysis process at various levels. Determination of the exact change point (CP) within a set of video frames or sequences will reduce the workload of the image analyst by filtering in only the relevant changes that have occurred during the period of interest. This in turn shall increase the overall efficiency of the video analysis workflow by rendering the necessary automation as a useful aid to the analyst. Limited work in the domain of applied CD exists with regard to the aspect of determination of the exact point of change. This is the objective of the current work wherein we make use of the threshold of the difference image sequence based on various segmentation algorithms as a metric for the determination of the possible CP in an image sequence or video feed. Malek Al Nawashi et al. [7] have made use of the simple temporal differencing approach along with a threshold function in order to determine the moving image as part of their work on abnormal human activity detection in an intelligent video surveillance system. Thus, there is a scope to apply the image difference approach in order to determine the point of change while subsequently overcoming its limitation in terms of the inability to detect the complete target shape [1].

CP detection has been studied in time series data analysis. In the context of remote sensing images as a sample case from an image processing perspective, Militino et al. [8] have carried out a very comprehensive survey recently (2020), of the various methods and tools available for CP detection. They infer that the methods applied to time series data may be applied in the context of time-ordered satellite images and image processing as well. We would like to extend this notion to the case of image processing as applied to video analytics in general. Amongst the techniques studied the nonparametric approach is a viable option given the fact that abrupt changes are likely to occur in a video sequence at any point of time, rendering it difficult for an underlying Bayesian or model-based approach to be followed. The nonparametric approach is
applicable to a wider variety of problems in CP detection since no assumption is made regarding any underlying model as surmised by Samaneh Aminikhanghahi et al. [9] in their comprehensive survey on CP detection methods for general time series data. The study points out that the inferences are applicable to the domain of image analysis as well. Nonparametric approach has also been analysed by Murari et al. [3, 4] too as part of their comprehensive survey on the DL-based CD methods as well.

One of the few studies on CP detection approach in a time-ordered set of images is that carried out by Manuel Bertoluzza et al. [10]. The objective of their work was to determine an accurate CD map between a selected pair of images amongst a time-ordered series of images by representing the changes along a temporal closed loop as binary sequences. In order to analyse the consistency of changes determined within a closed loop, the notion of a binary change variable was introduced. In our opinion, the use of the metric in order to compare the changes and finally achieve the desired accuracy is a novel idea. Though this step improves accuracy in existing methods of CD, the important question of determination of when a change has occurred or the CP still remains unanswered. The answer will enable efficient filtering of the video frames to a select few in form of image pairs about the respective CPs. The likely object or phenomenon of interest lies amongst these image pairs or frames. This can be a primary step yielding increased speed in processing within the overall intelligent video surveillance framework.

Based on the above discussion, the objective of our study is to determine a simple and robust method to determine CP within a set of segmented video frames forming part of a video surveillance feed. A change of variable or metric [10] in form of the threshold based on different image pairs is utilized to determine the point of change from amongst a set of images or frames. Rank ordering the changes based on the thresholds enables second or third CP detection as deemed fit by the image analyst. Nonparametric methods are more robust making no assumptions about the underlying model structure [9] and amongst these, Pettitt’s approach [11] is a simple and widely used technique. Our proposal for the change metric is similar to that of Pettitt’s.

The main contribution of the work is the following: 1) Determination of a suitable CD metric based on a comparative analysis of various segmentation methods to include Otsu, K Means based (denoted by ISODATA), minimum cross-entropy threshold (MCE) methods, 2) Application of the CD metric CP detection within the set of segmented video frames, and 3) Proposed framework to apply the CD metric-based CP concept to the intelligent video surveillance problem.

The chapter is organized as follows. Subsection 1.1 after introduction covers the aspects of the data set. Section 2 describes the basic CP detection algorithm based on the change metric concept. Section 3 discusses the results obtained based on a comparative analysis of respective CD metrics obtained from the four segmentation algorithms tested. Section 4 describes the proposed application framework of the results to the intelligent video surveillance framework.

1.1 Data set description

Most CD open source data sets are in form of image pairs as the objective is the application and testing of specific CD methods or algorithms to the same. In order to achieve the objective of the current work, there is a need for a time-ordered image data set. For this purpose, Google Earth-based time-ordered satellite image data sets of specific locations sourced from open source data [12] have been used and customized
for testing purposes. The satellite image data set has viability for automation in terms of information extraction by the image analyst, which is currently being done manually. Hence, the choice of this data set for developing results as part of the study has been undertaken. However, it is worth mentioning that the results obtained subsequently can be well applied to a general image processing scenario including video analysis. Google Earth images are a valid source of satellite imagery used for research purposes as evinced in work such as Ur̈ka Kanjir’s et al.’s survey [13].

The sample data set is as shown in Figures 1-3. Out of the time-ordered data set of 19 images, the relevant point of change is that between the fourth and fifth image (refer red arrow in Figure 1) when the object of interest or change has first appeared. The testing has been carried out on 10 such sets with the object appearing at an instance within the data set, which denotes the point of change. The spatial resolution of the data set is as per the standard Google Earth platform (> 5 cm) with each image corresponding to an area of 12 x 12 km on ground. The average temporal resolution of the 10 data sets of images was 10–15 years calculated between the first and last set of images.

CDNET2014 [14] is another standard open source data set for testing various CD algorithms based on static images and video sequences. We make use of this data set to demonstrate a more general application of the algorithm and analyse results on a test case along with those obtained for the above cases (Figures 1-3). The data set sample pertains to the intermittent object motion category and is like a parking lot with a man entering the scene at a certain point (frame number 57). Figures 4 and 5 show the sample data set that actually consists of 2500 frames forming part of a video feed in which testing is carried out on a selected number of frames (e.g. 80). The objective is to detect the point of change which is at the point of entry of the

![Figure 1.](image)

Sample data set sequence 1.
individual. As can be observed, the changes are extremely minor and difficult to
detect between respective frames as it is of a video recording. Application of various
segmentation algorithms such as Otsu, MCE, ISODATA and analysis therein to the
Google Earth and CDNET2014 data set shall enable the selection of a suitable method
accordingly.

2. Concept: Change point detection

2.1 Background

CP detection in time series is a well-researched area with a comprehensive survey
on various methods carried out by Aminikhanghahi et al. [9]. The application areas
include medical condition monitoring, climate change monitoring, speech recognition
and image analysis. CP detection in image analysis is the least researched area, and our
endeavour in the current work is to apply the useful lessons learned in the case of the
time series approach to that of image or video analysis. CP detection in time series is
much simpler compared to the case of image or video analysis considering that the
numeric values to be compared are easily extracted from the data itself. CP detection
in case of image or video analysis requires the determination of a suitable change
metric to be applied in a similar framework of time series in order to apply the
benefits of the same in this case. Trend and CP detection in remote sensing has been
well studied and classified by Militino et al. [7]. The nonparametric methods are
robust and applicable to a larger variety of problems compared with parametric
methods since changes in phenomenon or objects may be arbitrary not following any pattern or model. Amongst nonparametric methods, Pettitt’s method [11] is a well-established and applied method. We take a cue from this approach wherein the random variables forming part of the test hypothesis are substituted by the respective threshold values of the difference image sequences in order to determine the CP as explained below.

A suitable change variable or metric for the determination of the maximum CP in a time-ordered image set is the threshold values obtained from image segmentation of the different pairs of images. Subject to a minimum or no change scenario between images there will be minimum or no variation amongst the respective threshold values in the set. This premise has a rationale that any change in the sequence of images shall result in a variation in pixel values. This variation can be directly captured in form of a variation in the threshold values of the segmented image as per different algorithms applied. Otsu Binary segmentation algorithm [15] is a standard segmentation algorithm along with Li’s information theoretic MCE threshold method [16] and Coleman’s K means clustering image segmentation algorithm [17]. The threshold

Figure 3.
Sample data set sequence 3.
Figure 4. CDNET2014 data set result (MCE).

Figure 5. CDNET2014 data set result (Otsu).
values determined by these algorithms along with a mean threshold method are proposed to be used as the change variable or metric for the determination of CP in the time-ordered image sequence.

The methodology is thus based on thresholding (via application of the respective segmentation algorithms) of the binary image difference sequences constituting the image set. The point of maximum change is determined by the maximum value from amongst the threshold sequences of the binary image difference sequence. The algorithm is described in steps in the next section as illustrated in Figure 6.

2.2 Steps

Let us consider the set of time-ordered image sequences or video frames as $T = \{I_n | n \in [1, N]\}$ where $N$ is the total number of the images being processed. The objective is to select the pair of images that define precisely the maximum CP and further rank order the images in reducing relevance of CPs. This will assist the image analyst in sifting the images so as to determine the exact point of change while analysing the phenomenon or object of interest. This will enable timely and efficient analysis of the time-ordered image sequences or video frames. The steps are as follows:

1. Determine the image difference sequence (e.g. based on the symmetric difference \textit{absdiff} method in python) as $T_{\text{diff}} = \{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\}$.

2. Segment the image difference sequence based on methods such as [15–17] $S(T_{\text{diff}}) = S(\{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\})$ and buffer the respective threshold values as $T_{\text{th}} = \{Th_1, Th_2, Th_3, \ldots, Th_{n-1}\}$, $T_{\text{th}} = \{Th_n | n \in [1, N - 1]\}$.

3. On lines of Pettitt’s method [11] the CP in terms of the threshold is determined as $CP = \max |T_{\text{th}}|$.

4. Rank order the sequence of threshold values from maximum to minimum to determine CPs in decreasing order of relevance to aid the image analyst.
5. Based on the index of CP the corresponding image pair may be processed further to extract information as desired by the image analyst.

3. Results and analysis

3.1 Results

Methodology and steps described in section 2 have been applied to 10 data sets of the type described in Figures 1–3, and results obtained therein are displayed in Tables 1 and 2, respectively. Table 1 pertains to the category 1 evaluation wherein no margin for error is permitted and a valid detection is considered if as per ground truth, the CP is detected based on the maximum threshold value of the segmented difference image sequence. This is in keeping with the requirements or validity of the algorithm. It is also possible that due to pixel value variations owing to noise, and in certain cases the precise point of change is not captured corresponding to the maximum threshold value but the second highest threshold value or subsequent. Corresponding to this relaxation (valid detection considered up to the second highest threshold value), the results are re-valuated and presented in Table 2 as category 2. The standard Receiver Operator Characteristic (ROC) metrics of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are applicable for the current methodology as well with a slight modification. A correct detection in form of a TP corresponds to a TN as well since we are interested in the detection of only the correct image pair and not in the number of targets detected in a particular image as per standard applications. Similarly, in case, the correct image pair is not detected, that is, a FP occurs that corresponds to a FN as well. Recalling as per the standard

<table>
<thead>
<tr>
<th>Method</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>MCE</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>ISODATA</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>Mean</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 1. ROC metrics: category 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>MCE</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>ISODATA</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Mean</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 2. ROC metrics: category 2.
definition = (TP/(TP + FN)) represents the number of valid targets correctly detected. Precision as per the standard definition = (TP/(TP + FP)) gives the quality of the detection in terms of the correct number of target detected with a minimum of FPs. F1 Score as per the standard definition = (2*Precision*Recall)/(Precision+Recall) represents the degree of balance obtained in terms of both the precision and recall. In the present case for reasons aforesaid, that is, FP and TP being coincidental with FN and TN, respectively, the values of Recall, Precision and F1 score will all give the same values. Hence, for ease of assimilation of the reader and analysis therein we only mention Recall. Certain applications such as military target detection call for a high degree of recall compared with precision, that is, a minimum or no target miss scenario wherein one is ready to compromise to a certain extent on precision viz.-a-viz. recall.

The threshold plot corresponding to the sample image sequence (refer Figures 1–3) is presented in Figure 7. The threshold plot displays the variation in values corresponding to the respective threshold methods. As is visible, the point of change is correctly detected between the fourth and fifth image, compared with the ground truth (refer red arrow in Figure 1).

The CDNET2014 data set sample results are shown in Figures 4 and 5, respectively, with the corresponding plot displayed in Figure 8. From Figure 4, it is observed that using the cross-entropy method, MCE is able to detect the point of change accurately (refer Figure 4) with Otsu method (refer Figure 5).

3.2 Analysis of results

Based on the results, following are the relevant deductions:
1. From Tables 1 and 2, and plot in Figure 7, it is observed that the three methods Otsu, MCE and ISODATA perform well and are able to detect the CP accurately in case of the Google Earth data set.

2. For the category 1, the metric value of MCE has a slight edge compared with the other two methods, that is, Otsu and ISODATA as seen in Table 1. This is important considering that it is an information theoretic approach. As per the threshold plot in respect of MCE, it is observed that a greater capability to distinguish CPs exists.

3. The plot of Figure 8 along with Figures 4 and 5 provides another dimension to compare the methods based on the standard CDNET2014 data set [14]. It is observed that when there is a minute variation in changes between images in a video frame format, MCE is the only method that can distinguish the changes and accurately determine the relevant point of change. This is due to the fact that when a minor target enters a frame, the Otsu method tends to shift the threshold towards the foreground thereby suppressing relevant details [18]. Similarly, the ISODATA and mean methods also do not yield the correct results. The entry of the target (person) into the frame is incorrectly detected by the Otsu method as bit late in 76th frame (refer Figure 5) as compared with the actual frame in which the person enters, that is, 57th detected correctly by MCE (refer Figure 4). The CDNET2014 data set results to corroborate the findings given in Table 1 wherein the cross-entropy method provides the best performance.

4. The segmentation methods in order of performance are ranked as MCE followed by Otsu, ISODATA and lastly mean method. The cross-entropy has a slight edge over the Otsu, which has been observed in the Google Earth data set case (refer Table 1) while being validated on the CDNET2014 data set.
5. Irrespective of the CD method used for example, image difference-based approach or transformation-based approaches such as Principal component analysis (PCA), it is observed that the change metric in form of the threshold values of the segmentation method is a viable option for detecting point of change as validated based on the two data sets described above.

6. The results thus obtained can be well applied to a general image processing scenario including the application towards intelligent video surveillance.

4. Proposed framework: CP detection in video analysis

4.1 Case I: Static format

Based on the basic CD concept described in section 2 and results obtained in section 3, we describe two formats for implementation as part of the intelligent video analysis framework. The current description in this subsection pertains to static format case (refer Case I in Figure 9) wherein only a limited number of video frames or images are received and required to be analysed. In this scenario, the determination of the important CPs and in turn the filtering of probable objects or phenomenon are based on the basic CD framework described in Section 2. The steps remain the same as described in subsection B of the section. Figure 1 is a diagrammatic description of the concept, which is further modified for video surveillance case vide as in Figure 9. The modification is the addition of level I or level 2 processing element in form of a basic segmentation algorithm or an object detection algorithm. The level 1 processing scenario entails application of a segmentation algorithm as used for change metric determination, applied to the different images or image pairs about the CP. In case of searching for a specific category of target, a level 2 processing step in form of an object detection algorithm may be applied. A level 1 processing step applies the same segmentation algorithm (e.g. MCE) that has been used to determine the CD metric.

Figure 9.
Proposed CD framework for video analytics.
ensures full exploitation of the notion of a segmentation algorithm in terms of its capability to distinguish or partition a scene into the foreground and background [3, 4]. The foreground is likely to contain the phenomenon of interest or object of interest. By filtering the entire set of images or video frames received, to a likely pair of images the overall time period of processing will definitely reduce and effort too on the part of the image analyst. The steps as described vide in subsection B in Section 2 are applicable in the current case too and are as follows:

1. Determine the image difference sequence (e.g. based on the symmetric difference `absdiff` method in python) as 

$$T_{\text{diff}} = \{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\}.$$ 

2. Segment the image difference sequence based on methods such as [15–17] 

$$S(T_{\text{diff}}) = S(\{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\})$$ 

and buffer the respective threshold values as 

$$T_{\text{th}} = \{Th_1, Th_2, Th_3, \ldots, Th_{n-1}\}, \quad T_{\text{th}} = \{Th_n|n \in [1, N - 1]\}.$$ 

3. On lines of Pettitt’s method [11] the CP in terms of the threshold is determined as 

$$CP = \max \{T_{\text{th}}\}.$$ 

4. Rank order the sequence of threshold values from maximum to minimum in order to determine CPs in decreasing order of relevance in order to aid the image analyst.

5. Based on the index of CP, the corresponding image pair may be processed further to extract information as desired by the image analyst.

4.2 Case II: Fixed/moving calibration window format

This format is applicable when a continuous feed of video frames is being received for analysis in a fixed or moving camera scenario. The fixed window implies application of a calibration module over the first set of frames (refer red box and title First frame set). As part of the calibration module, the corresponding thresholds of the difference image sequences are determined. Once all calibration frames are received, the minimum and maximum thresholds corresponding to the segmented difference images are determined. The premise of employing a calibration module is to capture the background model in form of the thresholds of the successive difference images prior to the system being applied in a live scenario. Live scenario pertains to the actual phase of application wherein the information regarding the object or scene of interest is to be captured. Thus, in order to analyse the environment or background where the fixed or moving camera is employed, the calibration module enables capturing the background information or no change detected scenario. Once the thresholds of successive difference images are captured as part of the calibration module, subsequent threshold of difference images lying outside the range of those of the calibration module is indicative of a probable CD scenario. The yellow rhombus indicates this decision in Figure 9. The issues of false triggering are likely to be reduced as minor variations in the scene, which are to constitute the background captured in the calibration module prior to the application of the live phase. The steps for fixed calibration window are as follows:
1. Determine the image difference sequence for the say first \( n \) image difference sequence forming part of the calibration frame set as \( T_{\text{diff}} = \{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\} \).

2. Segment the image difference sequence based on methods such as \([15–17]\) and buffer the respective threshold values as \( T_{\text{th}} = \{Th_1, Th_2, Th_3, \ldots, Th_{n-1}\} \), \( T_{\text{th}} \equiv \{Th_n | n \in \{1, N - 1\}\} \).

3. Bracket the max and minimum thresholds as \( CP = \left[ \max [T_{\text{th}}], \min [T_{\text{th}}] \right] \).

4. From image sequence number \( n + 1 \) and beyond post-application of the calibration module described in above steps, the live phase commences wherein each image difference pair starting with \( I_{n+1} - I_n \) is tested for being a valid CP by segmenting the same to yield thresholds starting with \( Th_{n+1} \). Thus, the step is Check if \( \min [T_{\text{th}}] < Th_{n+1} < \max [T_{\text{th}}] \), that is, whether \( Th_{n+1} \) lies in \( CP \). If no, then the image pair constitutes a valid \( CP \). This step is represented by the yellow rhombus in Figure 9.

5. If the current difference image constitutes a \( CP \), then apply the level 1 or level 2 processing for further analysis or else repeat step 5.

6. Based on the application of the level 1 or level 2 processing present results with regard to the probable target or scene of interest duly processed as an aid to the image analyst.

The moving window concept is similar to the static case with the difference that the corresponding maximum and minimum threshold values vary as per the shifting window or set of frames over which calibration is carried out. In this case, the problem of dynamically changing scenarios such as vehicles starting and stopping abruptly is addressed. In such cases, the background needs to be dynamically updated for which adaptive algorithms have been proposed \([1]\). However, the CD metric is a powerful concept which in the current scenario is representative of the background static or dynamic as captured in form of the calibration module. In case of an envisaged scenario wherein the dynamic variation in background continues for a longer period, the moving window calibration module is applied to overcome these problems. Here, the threshold ranges detected over a fixed calibration frame within the static format are varied to change over sequences of frames being captured. The moving window calibration frames are depicted \( \text{via} \) the dashed lines in Figure 9. As the video frames are received, the set of thresholds corresponding to the calibration module are captured over the latest set of video frames in a pre-decided interval (corresponding to the anticipated degree of dynamism in background). Thus, the range of threshold values of the calibration module will be shifted over the next set of say \( n \) video frames thereby capturing the latest background in order to detect corresponding changes in subsequent frames. The steps for moving calibration window are as follows:

1. Determine the image difference sequence for the say first \( n \) image difference sequence forming part of the calibration frame set as \( T_{\text{diff}} = \{I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n\} \).
2. Segment the image difference sequence based on methods such as [15–17] $S(T_{\text{diff}}) = S(I_1 - I_2, I_2 - I_3, \ldots, I_{n-1} - I_n)$ and buffer the respective threshold values as $T_{th} = \{Th_1, Th_2, Th_3, \ldots, Th_{n-1}\}$, $T_{th} = \{Th_{n\leftarrow 1, N-1}\}$.

3. Bracket the max and minimum thresholds as $CP = [\max T_{th}, \min T_{th}]$.

4. From image sequence number $n + 1$ and beyond post-application of the calibration module described in above steps the live phase commences wherein each image difference pair starting with $I_{n+1} - I_n$ is tested for being a valid CP by segmenting the same to yield thresholds starting with $Th_{n+1}$. Thus, the step is Check if $\min T_{th} < Th_{n+1} < \max T_{th}$, that is, whether $Th_{n+1}$ lies in CP. If no then the image pair constitutes a valid CP.

5. If the current difference image constitutes a CP, then apply the level 1 or level 2 processing for further analysis or else repeat step 5.

6. Based on the application of the level 1 or level 2 processing present results with regard to the probable target or scene of interest duly processed as an aid to the image analyst.

7. The steps from 1 to 6 are repeated by modifying the calibration frame set starting with step 1 as $T_{\text{diff}} = \{I_{t+1} - I_{t+2}, \ldots, I_{t+n-1} - I_{t+n}\}$. It may be noted that $t$ pre-decided number of frames after which recalibration is carried out in terms of the fresh set of frames. The setting of the value of $t$ corresponds to the degree of dynamism anticipated in terms of the changing background wherein erstwhile foreground elements are anticipated to merge with the background. Thus, the least value of $t$ set to 1 corresponds to a highly dynamic scenario wherein the foreground elements tend to merge with the background rapidly.

The limitation in the simple frame differencing method of being unable to recover a complete shape of detected target [1] is overcome in our proposed framework by application of a Level 1 or Level 2 processing step post-detection of the CP as shown in Figure 9. Thus, once the point of change is detected, further application of say a level 2 processing will enable determination of the complete shape of the intended target.

4.3 Implementation issues

The CP detection-based methodology proposed for video analysis as described above in subsections A and B, respectively, is a simple adaptation of the CP-based approach. The advantage of the approach as adapted for video analysis is that it is simple and independent of the time-ordered set of video frames being received. Both offline (refer subsection A) and online (refer subsection B) options of implementation exist and it is a nonparametric approach, not making any assumptions regarding the underlying model. The change metric is a single value derived in a simple manner independent of any probabilistic methodology. Thus, the approach being nonparametric is applicable to a large number of scenarios since no assumptions are made regarding any specific scenario. The methodology is unsupervised not requiring any training data as in case of many deep learning or machine learning-based approaches.
Thus, the speed of implementation will be inherently higher in our case. The challenges in application of the method proposed are that initially it will require certain amount of testing and fine tuning in conjunction with an image analyst (for checking the performance of the algorithm). Factors such as the number of calibration frames, that is, window size for determination of the CD metric, will require certain fine tuning and innovation during implementation stage. The basic CP framework as described in Sections 2 and 3 was executed in Python code and the adaptation for the video analysis framework as described in the current section may follow suite. The architecture described in Figure 9 is simple and flexible and may hence be modified suitably as per results obtained during implementation stage.

4.4 Comparison with the state of the art (SOA) in intelligent video surveillance

The current focus of the SOA in the field of video surveillance is primarily on specific application scenarios as described in the comprehensive review by Guruh Fajar Shidik et al. [19]. Intelligent video surveillance includes anomaly detection, object detection, target tracking, etc., as few of the applications, which could apply a CD algorithm component as an important precursor step. It is worth noting that the CP detection concept as described in Sections 2 and 3 covering the application to the video analysis framework has not been well researched. Hence, a valid comparison with an equivalent method in context of video analytics does not exist. The closest semblance to the proposed method based on the CD concept is that of a discriminative framework for anomaly detection proposed by Allison Del Giorno et al. [20]. The proposed method endeavours to overcome the limitation in the existing anomaly detection methods namely the requirement of training data and dependence on temporal ordering of the video data. Their method is based on a nonparametric technique inspired by the density ratio estimation for CP detection. The approach is novel and similar to our proposed method in terms of the nonparametric approach wherein no assumptions are made about the underlying model. Further, the method proposed by Allison et al. does not require training data and is unsupervised similar to our case as well. They endeavour to use a metric- or score-based approach in order to determine anomaly points in a video sequence independent of the ordering of the video frames. However, the method does require an input of the features to aid in distinguishing the anomalies. It may be noted that the proposed methodology in our case is much simpler wherein no such feature set description is required to determine the CP and a single metric in form of the threshold of the image difference pairs is sufficient. This metric-based approach in our case makes the method simple and fast. Moreover, the CP concept is robust and adaptable to an anomaly detection framework. Thus, our method is simpler than the approach proposed by Allison Del Giorno et al. [20], which ultimately utilizes a probabilistic approach to determine the metric used to determine the anomaly points. The proposed CP-based video analysis methodology may be considered as a primary step in the intelligent video analysis framework prior to application of subsequent steps and a potential field for research. This analysis is to the best of the knowledge of the authors, the most relevant possible comparison with the SOA. A thorough review of the existing CD methods in other areas such as time series analysis and remote sensing has already been covered as part of literature review in Section 1. Thus, Section 1 and the current subsection comprehensively cover all the aspects of the proposed method and its typesetting viz.-a-viz. other areas of research.
5. Conclusion

To the best of the knowledge of the authors, this is the only study on CP detection in respect of image processing in particular as applicable to video surveillance as well. Important results have been obtained with the best method being determined as the cross-entropy MCE, followed by Otsu and ISODATA thereafter. The image difference-based CD metric method is by no means limited only to time-ordered set of images as represented in Figures 1–3. The method has been applied to a selected CDNET2014 data set as well as displayed in Figures 4 and 5. It may be noted that the sequence of images taken from the CDNET2014 data set are originally part of a video sequence, and hence, the results demonstrated in Section 3 (Refer Figures 4 and 5) are well suited to be applied to a video surveillance scenario. Thus, formulating the method in sliding window format will enable application to video surveillance scenario including suspicious activity detection scenarios. The block diagram for the proposed application of the CD concept is displayed in Figure 9 and proposed methodology has been described in detail in Section 4. The scope of applications possible is by no means limited to these two cases. In summary, the CD metric methodology in form of the threshold value needs to be exploited in an innovative manner. Further alternate change variable metrics may be a good area for further research. The objective of the current work has been to answer the important question of Where the change lies? or when has it occurred in a time-ordered set of images? This is important in order to act as a precursor for pin-pointed analysis of the images about the detected point of change as proposed in Section 4.

The level 2 processing in Figure 9 may also be in an Object Based CD (OBCD) framework [21]. Alternate options for processing the images detected about the CP may be considered part of future research scope.

Author details

Ashwin Yadav1, Kamal Jain1, Akshay Pandey1, Joydeep Majumdar2 and Rohit Sahay3

1 Department of Civil Engineering, Indian Institute of Technology, Roorkee, India

2 Department of Mechanical Engineering, Indian Institute of Technology, Indore, India

3 Department of Computer Science Engineering, Indian Institute of Technology, Kharagpur, India

*Address all correspondence to: ashwiny77@gmail.com
References


