Limitations with Measuring Performance of Techniques for Abnormality Localization in Surveillance Video and How to Overcome Them?

Manoj Kumar Sharma
Indian Institute of Technology
Kharagpur
manojsharma.net@gmail.com

Sayan Sarcar
Kochi University of Technology
mailtosayan@gmail.com

Debdoot Sheet
Indian Institute of Technology
Kharagpur
debdoot@ee.iitkgp.ernet.in

Prabir Kumar Biswas
Indian Institute of Technology
Kharagpur
pkb@ece.iitkgp.ernet.in

ABSTRACT

Now a days video surveillance is becoming more popular due to global security concerns and with the increasing need for effective monitoring of public places. The key goal of video surveillance is to detect suspicious or abnormal behavior. Various efforts have been made to detect an abnormality in the video. Further to these advancements, there is a need for better techniques for evaluation of abnormality localization in video surveillance. Existing technique mainly uses forty percent overlap rule with ground-truth data, and does not considers the extra predicted region into the computation. Existing metrics have been found to be inaccurate when more than one region is present within the frame which may or may not be correctly localized or marked as abnormal. This work attempts to bridge these limitations in existing metrics. In this paper, we investigate three existing metrics and discuss their benefits and limitations for evaluating localization of abnormality in video. We further extend the existing work by introducing penalty functions and substantiate the validity of proposed metrics with a sufficient number of instances. The presented metric are validated on data (35 different situations) for which the overlap has been computed analytically.

CCS Concepts

•**Computing methodologies** → **Scene anomaly detection;** *Visual inspection;* •**Applied computing** → **Surveillance mechanisms;**

Keywords

Video surveillance; Abnormality localization; Evaluation technique

ICVGIP, December 18-22, 2016, Guwahati, India

DOI: http://dx.doi.org/10.1145/3009977.3010044

1. INTRODUCTION

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior [4]. It is used in a wide variety of applications such as fraud detection, military surveillance of enemy activities, etc. [1, 5]. Anomaly detection in the data also indicates critical and actionable information [4]. As an example, an anomalous MRI image may indicate the presence of malignant tumors. Over time, several anomaly detection techniques have been developed across various research communities [14, 10]. In these applications, anomalies are recognized as a deviation from the normal model. Recently, there is a growing need of exploring anomaly detection in video surveillance domain [7, 13, 2, 3]. Traditionally, in a video surveillance system, monitoring, and analysis of the data captured by cameras were performed by a human operator. Since manual monitoring lacks continuous attention, which leads to missing of actions during anomalous events which occur rarely, it is difficult to take action at the time of occurrence. Additionally, closed-circuit television (CCTV) cameras in the existing system are used to monitor and store huge data without human intervention. The stored data are useful only for video forensic, i.e., to know the cause of annoying events, and prevention can be taken in the future. It does not help in stopping or taking proper actions at the time of occurrence. Hence, there is a need for automated video surveillance system for detection and recognizing abnormal events.

Developing such an automated video-based abnormality detection system would be highly useful in reducing the amount of data to be processed manually by directing attention to a specific portion of video surveillance data [15]. However, detecting and localizing abnormal behavior automatically in video surveillance for individual, group, or crowds are very challenging [10]. The effects of noisy data and the choice of representation of features significantly influences the performance of the system for detecting and localizing abnormality[15]. Extraction of suitable features is not only difficult but also time-consuming and requires expert domain knowledge [6].

Evaluation of video-based abnormality detection systems is both vital and highly challenging in computer vision applications. The goal of existing work in abnormality detection

in video surveillance is to correctly localize the abnormality. However, localization of abnormality can be evaluated quantitatively if suitable metrics are available. Generally, evaluation is performed for two aspects: frame-level and pixel-level anomaly detection [7, 9, 8]. However, these metrics have their own limitations which we will highlight in this paper and propose the possible solution.

The rest of the paper is organized as follows. State-of-art evaluation approaches are described in Section 2. Section 3 presents the proposed solution for evaluating localization of abnormality followed by analysis of proposed metric. Finally, Section 4 presents the discussion and Section 5 conclude the paper.

2. EVALUATION APPROACHES

Anomaly detection can be performed at frame-level and pixel-level. It is compared to the ground-truth to determine the number of true and false-positive frames. Here, the *presence* and the *absence* of anomalous events are considered as *positive* and *negative*, respectively [7, 8].

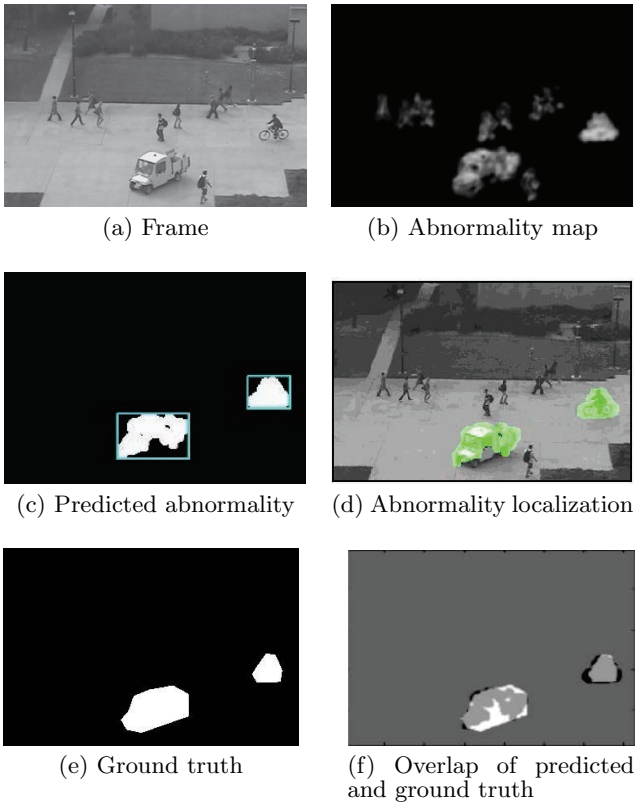
- *Frame-level :* A frame is considered as abnormal if it contains at least one abnormal pixel.
- *Pixel-level :* A frame is considered as true positive if at least 40% of the pixels in ground-truth are detected correctly.

Say, the abnormality detection system takes an input frame (Fig. 1(a)) and generates the anomaly map (Fig. 1(b)). After processing this map it creates a predicted regions which belongs to abnormality (Fig. 1(c)). This is further marked on original video frame for visualization by an operator who is monitoring the system (Fig. 1(d)). The original ground-truth for abnormality is shown in Fig. 1(e). Finally, Fig. 1(f) presents both ground-truth and the predicted regions.

In frame-level analysis, it indicates whether a frame contains abnormality or not, here localization of an abnormality is not of importance. However, there exists a deceitful chance of it being a lucky guess as it is not necessary that detected regions always overlap with the actual location of the abnormality. Therefore, a frame which is considered as true-positive may become anomalous due to *lucky* co-occurrences of erroneous detection [5, 7].

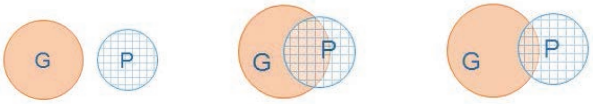
To illustrate this further, say G represents ground-truth region and P represents the predicted region within frame (F). In frame-level analysis, a frame is considered as abnormal if P is present in F i.e $P \subseteq F$ as shown in *Case-1* (Fig. 2(a), $G \cap P = \varnothing$), *Case-2* and *Case-3* (Fig. 2(b) and 2(c), $G \cap P \neq \varnothing$). In *Case-1*, although frame is considered as abnormal, however, it is not correctly localized. On the other hand, in *Case-2* and *Case-3* predicted data is overlapped with ground-truth.

In case of pixel-level accuracy, this problem is somewhat eliminated. Here, identifying object location is also important. A pixel-level abnormality localization for each frame (F) is measured by comparing the predicted region (P) with respect to the ground-truth region (G). If at least 40% of the truly abnormal location are detected correctly, then the frame is considered to be anomalous (true-positive). Hence, 40% of the ground-truth data is considered as threshold value, denoted as K. Further, in Fig. 2, *Case-2* is only considered to be abnormal as $\frac{G \cap P}{G} \geq 0.4$ (see Eqn.1), whereas



(a) Frame (b) Abnormality map (c) Predicted abnormality (d) Abnormality localization (e) Ground truth (f) Overlap of predicted and ground truth

Figure 1: Abnormality detection and localization



(a) Case-1 (b) Case-2 (c) Case-3

Figure 2: Set representation for ground-truth and predicted regions

in *Case-1* and *Case-3*, $\frac{G \cap P}{G} < 0.4$. On the other hand, a frame is considered false-positive if ground-truth indicates it to be normal but one or more if its pixels are detected as abnormal [7, 2].

Once the true-positive and false-positive counts are known, we compute true-positive rate (TPR) and false-positive rate (FPR) and finally using this we compute ROC and equal error rate represented as EER [5, 7, 11]. Further, The definition of TPR, FPR, ROC and EER are as follow.

- *TPR : True positive rate* (also called *sensitivity*, or *recall*) measures the proportion of positive frame that are correctly identified.

$$TPR = \frac{\text{Number of true-positive frame}}{\text{Number of positive frame}}$$

- *FPR : False positive rate* (also known as the false alarm ratio) is the proportion of all negatives that still

yield positive test outcomes.

$$FPR = \frac{\text{Number of false-positive frame}}{\text{Number of negative frame}}$$

- *ROC : Receiver operating characteristic* curve is created by plotting the *true positive rate* (TPR) against the *false positive rate* (FPR) at various threshold settings.
- *EER : Equal Error Rate* also know as the error crossover point (ratio of misclassified frames at which $FPR = 1 - TPR$). In other words, if we plot both *False acceptance rate* ($FAR = FPR$) and *False reject rate* ($FRR = 1 - TPR$) curves versus the *Sensitivity* setting then the point where the two curves cross is the *ERR*.

This paper extends the initial work of Mahadevan et al. [7, 8] to present a generalized framework for measuring the localization capability of abnormality in the video. The new metric focuses on measuring the accuracy at pixel-level localization of abnormality.

Challenges and limitations

We discuss three existing metrics used in the current state-of-art approaches, namely *Overlap-based method*, *Jaccard-index* and *Dice-coefficient*, followed by their limitations in performance evaluation.

Assuming G and P represent the ground-truth and predicted data, respectively, the metrics are defined as follows.

1. Overlap-based method
   The existing pixel-level anomaly detection metric [7, 8] is represented as overlap-based method.

$$\text{Overlap-coefficient} = \frac{G \cap P}{G} \tag{1}$$

2. Jaccard-index

   The Jaccard-index, also known as the Jaccard similarity coefficient, measures similarity between sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$\text{Jaccard index} = \frac{G \cap P}{G \cup P} \tag{2}$$

3. Dice-coefficient

   Dice-coefficient, also known as Sorensen-Dice index, is used for comparing the similarity of two samples. It can be measured as follows.

$$\text{Dice coefficient} = \frac{2|G \cap P|}{|G| + |P|} \tag{3}$$

   Where $|G|$ and $|P|$ are the number of object in samples G and P, respectively.

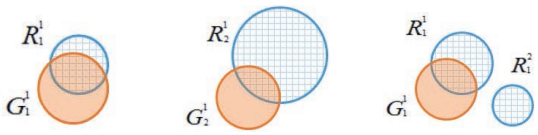
All these existing metrics support values ranging from 0 to 1. Currently, the Dice similarity index is more popular and has values higher compared to Jaccard-index. The existing overlap-based approach (Eqn.1) focuses mainly on the intersection of predicted and ground-truth data and does not provide any information about the non-overlapping portions of data or mismatch (see Fig. 3). This problem is somewhat eliminated in Jaccard-index and Dice-coefficient. However, Jaccard and Dice have their own limitations [12]. In the following, we highlight some of the issues for measuring the performance in proper localization of the abnormality detection system through several scenarios as follows.

(a) Abnormalities are bicycle and skater

(b) Localization of abnormality

Figure 3: Frame containing hit, miss and false alert (best viewed in color)

- Predicting extra region is not preferred: Suppose we have two scenarios, *Case-1* (as in Fig. 4(a)) and *Case-2* (Fig.4(b)). Here, *Case-2* covers the original object, but with extra surrounding region other than the object of interest. In other words, let object of interest be denoted by G, predicted region by P, R_i^j and G_i^j represent j^{th} region in predicted and ground-truth data belonging to *Case-i*. Here, $P_1 = \{R_1^1\}$, $P_2 = \{R_2^1\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1\}$, $G_1 = G_2$, $G_1 \cap P_1 = G_2 \cap P_2$ and $G_1 \cup P_1 < G_2 \cup P_2$. In such case, preference should be given to *Case-1* over *Case-2* (represented as *Case-1* $\succ$ *Case-2*).

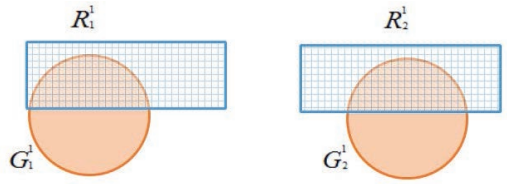


(a) (b) (c)

Figure 4: Challenges and limitations: Example-1

Existing Overlap-based metric (Eqn.1) fails to distinguish between two aforementioned scenarios. However, Jaccard-index and Dice-coefficient (Eqn.2 and 3, respectively) will be able to handle such situations.

- Predicted region should be close to the ground-truth region or object of interest: Consider the two cases, namely *Case-1* and *Case-2* (Fig. 5(a) and Fig. 5(b)). Let $P_1 = \{R_1^1\}$, $P_2 = \{R_2^1\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1\}$, $G_1 = G_2$ and $G_1 \cap P_1 = G_2 \cap P_2$, $G_1 \cup P_1 = G_2 \cup P_2$ i.e is $P_1 = P_2$. However, in *Case-1* predicted regions are distributed such that it appears to be shifted from the region of interest, although they have the same area. In such case, *Case-2* $\succ$ *Case-1*. Existing metric (Eqn.1, 2 and 3) fails to distinguish the scenario.



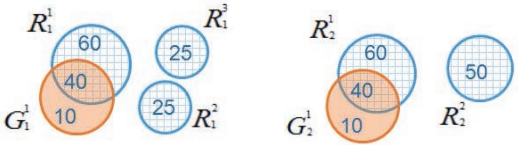
(a) Case-1 (b) Case-2

Figure 5: Challenges and limitations: Example-2

- False alert should be minimum: We explain this point with two scenarios.

  Scenario-1: Consider the two cases, namely *Case-1* and *Case-2* (figure 4c and 4b). Let $P_1 = \{R_1^1, R_1^2\}$, $P_2 = \{R_2^1\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1\}$ and $G_1 \cap P_1 = G_2 \cap P_2$, $G_1 \cup P_1 = G_2 \cup P_2$. In such case, *Case-2* $\succ$ *Case-1*.

  Scenario-2: We consider two cases as *Case-1* and *Case-2*. Let $P_1 = \{R_1^1, R_1^2, R_1^3\}$, $P_2 = \{R_2^1, R_2^2\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1\}$, assuming $G_1 = G_2$, $R_1^1 = R_2^1$, $G_1 \cap P_1 = G_2 \cap P_2$ and $G_1 \cup P_1 = G_2 \cup P_2$ (i.e $R_1^2 + R_1^3 = R_2^2$). In such case, existing metric cannot distinguish them. However, *Case-2* $\succ$ *Case-1* as shown in Fig. 6.

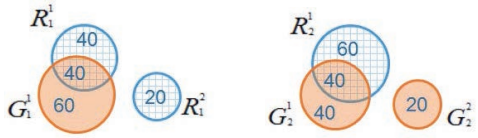


(a) Case-1 (b) Case-2

Figure 6: Challenges and limitations: Example-3

- High penalty should be imposed if abnormality is present but not detected: This situation can be occurred through two scenarios.

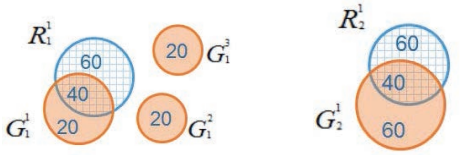
  Scenario-1: This refers the comparison between undetected and false detected region of interests. Let $P_1 = \{R_1^1, R_1^2\}$, $P_2 = \{R_2^1\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1, G_2^2\}$, $G_1 = G_2$ and $G_1 \cap P_1 = G_2 \cap P_2$, $G_1 \cup P_1 = G_2 \cup P_2$. In such case, *Case-1* $\succ$ *Case-2* as shown in Fig. 7.



(a) Case-1 (b) Case-2

Figure 7: Challenges and limitations: Example-4

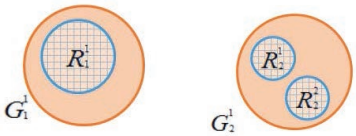
Scenario-2: It is to describe the comparison between detected and undetected region of interests. It is preferred to detect every object of interest. On the other hand, more number of undetected object degrades the performance of detection. A penalty should be imposed when object is not detected properly as shown in Fig. 8. Let $P_1 = \{R_1^1\}$, $P_2 = \{R_2^1\}$, $G_1 = \{G_1^1, G_1^2, G_1^3\}$, $G_2 = \{G_2^1\}$ and $G_1 \cap P_1 = G_2 \cap P_2$, $G_1 \cup P_1 = G_2 \cup P_2$. In such case, *Case-2* $\succ$ *Case-1*.



(a) Case-1 (b) Case-2

Figure 8: Challenges and limitations: Example-5

- Single predicted region is better compared to multiple sub-regions representing the same object: Let $P_1 = \{R_1^1\}$, $P_2 = \{R_2^1, ..., R_2^n\}$, $G_1 = \{G_1^1\}$, $G_2 = \{G_2^1\}$ and $G_1 \cap P_1 = G_2 \cap P_2$, $G_1 \cup P_1 = G_2 \cup P_2$. In such case, *Case-1* $\succ$ *Case-2* as in Fig. 9.



(a) Case-1 (b) Case-2

Figure 9: Challenges and limitations: Example-6

3. PROPOSED APPROACH

We propose new metric, which addresses the issues discussed in previous sections. We start with describing different penalty function used in our approach followed by a mechanism to compute true-positive and false-positive. This can be used to compute true-positive rate, false-positive rate and finally equal error rate.

The proposed solution (Eqn. 4) consists of two sub parts namely dice-coefficient (DC) and penalty function ($\mathcal{F}$).

$$vs\text{-}coef = \mathrm{DC} - \mathcal{F} \tag{4}$$

Further, $\mathcal{F}$ is divided into two penalty functions, namely $\mathcal{F}(\phi)$ and $\mathcal{F}(\theta)$. Different weights are assigned (α and β) to them (as shown in Eqn.5). Here, α and β are x and y percentage of dice-coefficient.

$$\mathcal{F} = \alpha \cdot \mathcal{F}(\theta) + \beta \cdot \mathcal{F}(\phi) \tag{5}$$

We describe the computation of $\mathcal{F}(\phi)$ and $\mathcal{F}(\theta)$ as follows. Here, $\mathcal{F}(\phi)$ represents correction due to mis-classifications and presence of extra cluster. It can be calculated using Eqn.6 and 7. Further, ρ controls the cost based on different scenarios. For example, high penalty is imposed when abnormality is presents but not detected ($n_g > n_p$) and medium penalty is imposed in case of false alert ($n_g < n_p$).

$$\mathcal{F}(\phi) = \frac{\rho}{I + \delta}\left[\frac{2\max(n_g, n_p, \psi)}{n_g + n_p} - 1\right] \tag{6}$$

$$\rho = \begin{cases} 0, \text{ if } n_g = 0 || n_p = 0 || I = 0 || n_p = n_g = I \\ 1, \text{ if } n_g \leq n_p \\ 2, \text{ if } n_g > n_p \end{cases} \quad (7)$$

In Eqn.6, I represents the number of cluster formed by intersecting G and P. Further, δ is constant to handle the division by zero issue (say, $\delta = 0.01$), n_g is the number of cluster in ground truth and n_p is the number of cluster in predicted data. In addition, ψ represent the number of connected component on ground truth and predicted data. The detail description of terminologies used in the proposed approach are listed in Table: 1.

We further demonstrate the computation of $\mathcal{F}(\phi)$ with an example shown in Fig. 10, containing two ground-truth region and three prediction regions. Here, $n_g = 2$, $n_p = 3$, $I = 3$, $\psi = 2$ and $\mathcal{F}(\phi) = \frac{1}{3+0.01}\left[\frac{2\max(2,3,2)}{2+3} - 1\right] = 0.066$.

(a) n_g=2 & n_p=3 (b) $I = 3$ (c) $\psi = 2$

Figure 10: Computation of $\mathcal{F}(\phi)$ an example

Table 1: List of symbols used

Symbols	Descriptions
G	Number of pixel in ground-truth data
P	Number of pixel in predicted data
$\mathcal{F}$	Penalty function
$\mathcal{F}(\phi)$	Penalty function due to mis-classifications and presence of extra cluster
$\mathcal{F}(\theta)$	Penalty function due to misalignment of predicted region
α	0.75 times of dice-coefficient (weights).
β	0.50 times of dice-coefficient (weights).
ρ	Cost function
I	Number of cluster formed by $G \cap P$
δ	0.01 (constant)
γ	0.8 (constant)
n_g	Number of cluster present in ground-truth
n_p	Number of cluster present in predicted data
ψ	Number of connected component formed by $G \cap P$
C_g	Center-of-mass for ground-truth data
C_p	Center-of-mass for predicted data
C_n	Combined center-of-mass for ground-truth and predicted data
AB	Width of bounding box around ground-truth data
BC	Height of bounding box around ground-truth data
K	Threshold value

$\mathcal{F}(\theta)$ represents the correction for misalignment of the predicted region with respect to ground-truth regions (as highlighted in Fig. 5) such that the predicted region should be close to the object of interest. It can be computed using Eqn. 8.

$$\mathcal{F}(\theta) = \frac{1}{I+\eta}\sum_{i=1}^{I}\frac{4\cdot(|C_g - C_p| - |C_g - C_n|)}{\chi_i} \quad (8)$$

$$\eta = \sum_{i=1}^{I}\frac{|AB - BC|}{|AB + BC|} \quad (9)$$

$$\chi_i = |AB + BC| \quad (10)$$

Here, C_g, C_p and C_n represent the center-of-masses for ground-truth data, predicted data and combined center-of-mass for ground-truth and predicted data, respectively. These metrics are computed for each object in ground-truth intersecting with predicted data. Further, position of bounding box around ground-truth data is represented by points A, B, C and D as shown in Fig. 11.

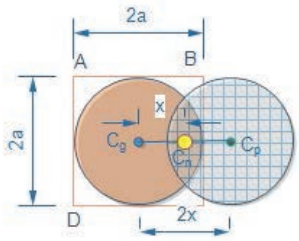


Figure 11: Computation of penalty function $\mathcal{F}(\theta)$

We combine two penalty functions $\mathcal{F}(\theta)$ and $\mathcal{F}(\phi)$ and represent them in Eqn. 11. Then the value of *vs-coef* is non-linearly stretched using Eqn.12. Finally we use a threshold value, represented as K, to cut the stretched version of *vs-coef* (represented as τ) to decide whether it belongs to true-positive or not (shown in Eqn.13).

$$\tau = L_{out} + \frac{(H_{out} - L_{out}) \times (\text{VS-coef} - L_{in})}{H_{in} - L_{in}} \times \gamma \quad (12)$$

Here, L_{in}, H_{in}, L_{out}, H_{out} are lower and upper limit of input and output values, γ denotes the gamma correction coefficient.

$$TP = \begin{cases} 1, \text{ if } \tau > K \\ 0, \text{ otherwise} \end{cases} \quad (13)$$

A frame is considered as true-positive (TP) if its value is greater than the predefined threshold (say, K). Otherwise, it is considered as false positive.

$$\textit{vs-coef} = \frac{2|G \cap P|}{|G| + |P|} - \frac{\beta \cdot \rho}{I + \delta}\left[\frac{2\max(n_g, n_p, \psi)}{n_g + n_p} - 1\right] - \frac{\alpha}{I+\eta}\sum_{i=1}^{I}\frac{4\cdot(|C_g - C_p| - |C_g - C_n|)}{\chi_i} \quad (11)$$

Analysis of Proposed Metric

The proposed evaluation metrics are applicable for wide variety of situations, including cases where the number of predicted regions and number of region in ground-truth are same and where they are not. In this concern, we present three case studies namely *Study-1*, *Study-2* and *Study-3*.

To validate the efficacy of the proposed metrics in *Study-1*, we discuss on two situations having perfect overlap and minimal overlap between predicted and region of interest.

- *Perfect overlap :* In this scenario, we assume that the predicted and ground-truth data overlap perfectly to each other. The simplest case occurs when number of predicted region are same as number of ground truth ($n_g = n_p$). To illustrate it further, we take simple case where centroid of predicted and ground-truth data overlap with the centroid of combined predicted and ground-truth as shown in Fig. 12. In this case, $C_g = C_p = C_n$. Say, Δ represents the difference between centroid of ground-truth and predicted region ($|C_g - C_p|$) and centroid of combined predicted and ground-truth with centroid of ground-truth ($|C_g - C_n|$). Hence, $\Delta = |C_g - C_p| - |C_g - C_n|$ which becomes zero. Here, $I = 1$, $\eta = 0$, $\chi_i = 4a$, hence $\mathcal{F}(\theta) = 0$.

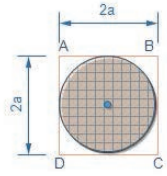


Figure 12: Perfect overlap example

- *Minimal overlap :* In this scenario, we assume that object of interest and predicted region touches each other. We illustrate it with a simple example shown in Fig. 13. In this case $\Delta = |C_g - C_p| - |C_g - C_n| = 2a - a = a$, $I = 1$, $\eta = 0$, $\chi_i = 4a$, hence $\mathcal{F}(\theta) = 1$. Therefore, more penalty is imposed in such a case.

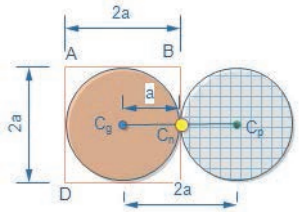
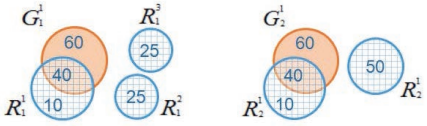


Figure 13: Minimal overlap example

In *Study 2*, we analyze the situation in presence of false alert. For this we consider two cases, *Case-1* and *Case-2*. In both the cases, suppose G and P contain 100 pixel. However, in *Case-1* there is one cluster belong to ground-truth ($n_g = 1$) and 3 belong to predicted data ($n_p = 3$) as shown in Fig. 14. In this case, $n_g < n_p$. In such situation, a medium penalty is imposed (see Eqn.7). Further, we have $\psi_1 = 3$, $I_1 = 1$.

On the other hand, *Case-2* shows 2 predicted regions ($n_g = 1$ and $n_p = 2$) and $\psi = 2$. However, both of them



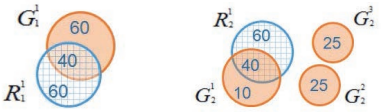
(a) Case-1 (b) Case-2

Figure 14: Computation of $\mathcal{F}(\phi)$ an example

contain $G = 100$, $P = 100$, $G \cap P = 40$, $G \cup P = 160$ and $I = 1$. According to Fig. 14, we should give preference to *Case-2* compare to *Case-1*.

We compute the cost function, (ρ) following the Equation 7. Here $n_g < n_p$, hence $\rho = 1$. Penalty function $\mathcal{F}(\phi_1) = 0.495$ and $\mathcal{F}(\phi_2) = 0.330$. Hence, less penalty is impose in *Case2* compare to *Case1* therefore we get *Case-2* $\succ$ *Case-1*. Here, the penalty function, $\mathcal{F}(\phi)$, is used to break the tie between different cases when dice-coefficient is same for the two cases.

In *Study 3*, we analyze the situation when abnormality is present but not detected (as shown in Fig. 15). In this case $n_g > n_p$ hence $\rho = 2$ for *Case-2* and $\rho = 0$ for *Case-1*. $G \cap P$ are same for both the cases similarly $G \cup P$ are also same. Further, $\mathcal{F}(\phi_1) = 0$ and $\mathcal{F}(\phi_2) = 0.99$ and hence more penalty is impose in *Case-2* compare to *Case-1* which lead to $Case1 \succ Case2$.



(a) Case-1 (b) Case-2

Figure 15: Computation of $\mathcal{F}(\phi)$ an example

Further, we summarize the computation of penalty function $\mathcal{F}(\phi)$, with various scenarios and present them in Table: 2 and 3.

Analysis of different performance evaluation metrics on the UCSD Ped-dataset is shown in Fig. 16. Here, top row represents the performance of different metrics, middle row displays the frame containing localization of abnormality and bottom row indicates the predicted and ground-truth data. Quantitative analysis of $\mathcal{F}(\phi)$ for each case is shown in Table. 2, where case (a), (c), (d) and (e) are diagrammatically mapped with the *Image* having ($S = 1, F = 8$), ($S = 1, F = 1$), ($S = 4, F = 2$) and ($S = 4, F = 3$), respectively. From the Fig. 16, we can analyze that in case of overlap based approach, the values are almost same for all the cases. This is because existing overlap based approach does not consider extra predicted region into the computation.

Further, Jaccard and Dice coefficients can produce different values for the cases. However, if we analyze cases (d) and (e) compared to case (c), it appears that Dice coefficient value is higher in spite of having false alert. This type of false alert is taken care in the proposed approach (as indicated in *vs-coef* and τ). Here, case (c) performs better compare to cases (d) and (e) containing one and two false alert(s), respectively. Source code of the paper can be ac-

Table 2: Quantitative analysis of $\mathcal{F}(\phi)$ part-1

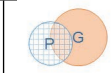
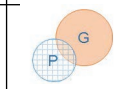
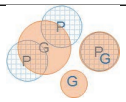
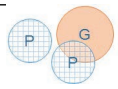
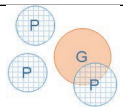
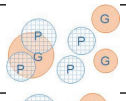
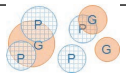
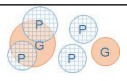
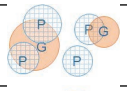
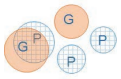
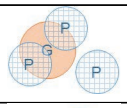
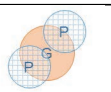
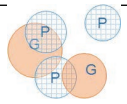
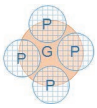
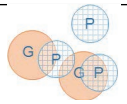
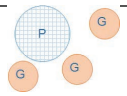
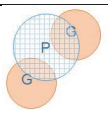
Set Representation								
S	Set	F	Image	n_g	n_p	I	ψ	$\mathcal{F}(\phi)$
1	$n_g = n_p = I$	1		1	1	1	1	0
		2		1	1	1	1	0
		3		2	2	2	2	0
		4		1	1	1	1	0
		5		3	3	3	3	0
		6		1	1	1	1	0
		7		3	3	3	3	0
		8		1	1	1	1	0
2	$I < n_g = n_p$	1		1	1	0	2	0
		2		2	2	1	3	0.495
3	$n_g = n_p < I$	1		2	2	3	1	0.166
4	$I \leq n_g < n_p$	1		0	1	0	1	0
		2		1	2	1	2	0.33
		3		1	3	1	3	0.495
		4		3	4	2	5	0.071
		5		3	4	3	4	0.047

Table 3: Quantitative analysis of $\mathcal{F}(\phi)$ part-2

Set Representation								
S	Set	F	Image	n_g	n_p	I	ψ	$\mathcal{F}(\phi)$
4	$I \leq n_g < n_p$	6		2	4	2	4	0.166
		7		2	4	3	3	0.111
		8		2	3	1	4	0.594
5	$n_g < I < n_p$	1		1	3	2	2	0.249
6	$n_g < n_p \leq I$	1		1	2	2	1	0.166
		2		2	3	3	2	0.066
		3		1	2	2	1	0.166
		4		1	4	4	1	0.15
		5		2	3	3	2	0.066
		6		1	2	2	1	0.166
		7		2	4	4	2	0.83
		8		2	3	3	2	0.066
7	$I \leq n_p < n_g$	1		2	1	1	2	0.066
		2		3	1	0	4	0
8	$n_p < I \leq n_g$	1		2	1	2	1	0.166

cessed from MATLAB central [1].

[1]https://in.mathworks.com/matlabcentral/fileexchange/60022-vscoeff-icvgip2016

4. DISCUSSIONS

This section sheds light on the effectiveness of the proposed metrics and its implications in order to develop more robust ways of measuring abnormality localization in the video surveillance system.

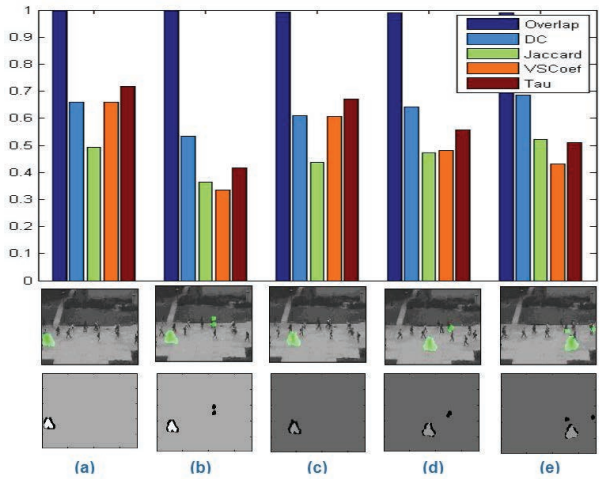


Figure 16: Analysis of different performance evaluation metrics (best viewed in color)

The proposed metrics precisely handle false alert in the video. This not only helps the system to identify abnormality more accurately, but also reduces the cognitive load of the operators as reduced number of false alert helps to find the abnormality quickly and accurately. In other context, when a large false alert is present, then it is likely that human operator who is monitoring the surveillance video will miss the actual abnormal data in the video. Solving this, a cost penalty is imposed in the proposed approach to measure the performance.

The new metrics ensure imposing a high penalty when an abnormality is present but not detected in the frame. This overcomes the shortcoming of existing metrics and enhances the chance of identifying such abnormality. For example, in a scene, we have car and bicycle as the region of interests. Assuming 80% area of the car is highlighted, but no region belongs to bicycle is marked. It would be much better if both car and bicycle are detected instead the accumulated region of interests of car and bicycle is less than 80%. Our current metrics address the problem and ensure every object contributes in the calculation.

The proposed metrics handle the position of predicted region surrounding the ground-truth. Suppose a predicted region has fixed overlap, but shifted more toward one side of the region of interest. It will be better if it is overlapped keeping the position of the predicted region at the center or close to center (surrounding the ground-truth). In such case, our proposed metric more accurately measures performance of object localization as it keeps track how near or far the centroid of the ground-truth and predicted region is located.

Unlike the overlap-based approach, the proposed metric is a modification over dice-coefficient, hence, it inherently takes care of under and over estimated region and imposes penalty accordingly.

5. CONCLUSIONS

Proper localization of abnormality in the video is important for video surveillance task. We analyzed and observed that existing metric to compute pixel-level accuracy is not sufficient. In this context, we explored three metric namely overlap-based approach, Jaccard-index, and dice-coefficient. We observed that dice-coefficient is performing better compared to overlap-based approach, however, it has its own limitations.

In this study, we highlighted several cases when existing metric are unsuitable for separating the data. We further proposed our own solution to overcome these issues. It was observed that compared with three existing metrics, newly proposed metric performs better and can be successfully applied to measure the performance of abnormality detection technique.

6. REFERENCES

[1] C. C. Aggarwal. *Outlier analysis*. Springer Science & Business Media, 2013.

[2] S. Biswas and R. V. Babu. Anomaly detection in compressed h.264/avc video. *Multimedia Tools and Applications*, 74(24):11099–11115, 2015.

[3] O. Boiman and M. Irani. Detecting irregularities in images and in video. *Int. J. Com. Vis.*, 74(1):17–31, 2007.

[4] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM comp. surv.*, 41(3):15, 2009.

[5] M. Javan-Roshtkhari. *Visual event description in videos*. PhD thesis, McGill University, 2014.

[6] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[7] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Trans. Patt. Anal., Mach. Intell*, 36(1):18–32, 2014.

[8] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In *Proc IEEE Conf. Comp. Vis., Patt. Recog. (CVPR), 2010*, pages 1975–1981, June 2010.

[9] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Proc IEEE Conf. Comp. Vis., Patt Recog. (CVPR), 2009.*, pages 935–942. IEEE, 2009.

[10] O. P. Popoola and K. Wang. Video-based abnormal human behavior recognition - a review. *IEEE Trans. Sys. Man., Cyber. Part C: Appl., Rev.*, 42(6):865–878, 2012.

[11] V. Reddy, C. Sanderson, and B. C. Lovell. Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture. In *Comp. Vis., Patt. Recog. Workshops*, pages 55–61, 2011.

[12] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442, 2004.

[13] M. J. Roshtkhari and M. D. Levine. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Com. Vis., Image Understanding*, 117(10):1436–1452, 2013.

[14] V. Saligrama, J. Konrad, and P.-M. Jodoin. Video anomaly identification. *IEEE Signal Process. Mag.*, 27(5):18–33, 2010.

[15] A. A. Sodemann, M. P. Ross, and B. J. Borghetti. A review of anomaly detection in automated surveillance. *IEEE Trans. Sys. Man., Cyber. Part C: Appl., Rev.*, 42(6):1257–1272, 2012.