J. Vis. Commun. Image R. 78 (2021) 103165

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Full length article

A comprehensive survey on computer vision based concepts, methodologies, analysis and applications for automatic gun/knife detection ☆

Rajib Debnath, Mrinal Kanti Bhowmik *

Department of Computer Science and Engineering, Tripura University, Suryamaninagar, Tripura, 7990022, India

ARTICLE INFO

Keywords:
Weapon detection
Survey
Security and surveillance
Multiple object detection
Visual and concealed

ABSTRACT

The ability to detect gun and gun held in hand or other body parts is a typical human skill. The same problem presents an imperative task for computer vision system. Automatic observer independent detection of hand held gun or gun held in the other body part, whether it is visible or concealed, provides enhance security in vulnerable places and initiates appropriate action there. Compare to the automatic object detection systems, automatic detection of gun has very few successful attempts. In the present scope of this paper, we present an extensive survey on automatic detection of gun and define a taxonomy for this particular detection system. We also describe the inherent difficulties related with this problem. In this survey of published papers, we examine different approaches used in state-of-the-art attempts and compare performances of these approaches. Finally, this paper concludes pointing to the possible research gaps in related fields.

1. Introduction

Detection of gun is a well-established, elementary and challenging problem in the field of computer vision. According to the records, number of crimes using guns are very concerning certain places in the world, especially in countries where use of guns is illegal or was illegal for a period. The last statistics reported by the United Nations Office on Drugs and Crime (UNODC) reveals that the number of crimes involving guns per 100,000 habitants are very high in many countries, e.g., 21.5 in Mexico, 4.7 in United States and 1.6 in Belgium [1]. In addition, several psychological studies demonstrated that the simple fact of having access to a gun increases drastically the probability of committing a violent behaviour.

In today's world use of Closed-Circuit Televisions (CCTVs) are increasing exponentially to combat crime. With the increasing demand of CCTVs, it makes challenge for a human operator to inspect and analyze the video feed from the remote camera and take any appropriate action thereon; his repeatedly burdens an unworkable amount of observance, also can be expensive and unproductive when several video streams are present. Different studies [2–4] suggested that the human operator suffers **video blindness** after 20 to 40 min of active monitoring and misses the screen activity as high as 95%, which drastically reduce the detection accuracy up to 83%. This manual observation problem and proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, obliged researchers to think for a vision based automatic gun detection system.

Over the past few decades, computer vision practitioners are attempting to build a vision system to detect a gun alone or gun present in other body parts. When an individual carries a gun or other weapon in hand or other body parts, it is a strong indicator of a possibly risky situation; this is because the gun is operative by hand only while committing any crime with it. By detection or identification, we refer to recognition and precise localization of gun in hand or other body parts of humans and classify different types of guns as well. It is assumed that the ideal crime gun image or shape similar to that gun is available to the vision system. The main goals of such a vision system are:

- To generate an alarm that able to alert surveillance human personnel or appropriate security person in real-time, resulting in immediate action thereon.
- Different types of guns can be classified (intra-object classification), which can provide vital information for the forensic team for analysis.
- Provide an analysis of Human Activity Recognition, Human Behaviour Recognition, Re-identification of Suspect, Recognition of Different Hand Position, etc..

In this paper, we survey the progress of a vision-based gun detection system. The block diagram of the computer vision-based system is shown in Fig. 1. Few exemplary gun images from publicly available IMFDB [5] is shown in Fig. 2(a). A set of firing scenarios showing the gun images of Fig. 2(a) are detected is shown in Fig. 2(b).

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: rajibdebnath.cse@gmail.com (R. Debnath), mrinalkantibhowmik@tripurauniv.ac.in, mkb.cse@gmail.com (M.K. Bhowmik).

https://doi.org/10.1016/j.jvcir.2021.103165
Received 20 October 2020; Received in revised form 19 March 2021; Accepted 15 May 2021
Available online 21 May 2021
1047-3203/

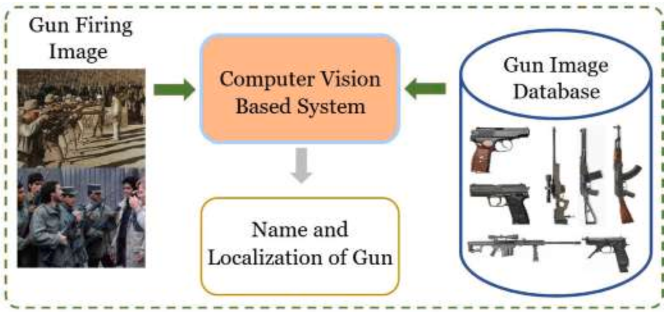

Fig. 1. A typical computer vision system for detection of weapon.

There are omnipresent sensor-based system (like metal detector [6, 7]) to monitor detection and recognition of weapon (gun/ knife) in different security zones. But these setups are mostly present in the entrance of the security zones, or some security personnel will manually scan for this. Those sensor-based systems can be seen with the naked eye, which aware intruder and that personnel will be careful about that. Moreover, it will only help security zones like inside of the airport, inside of the banks, inside of the religious places, etc. and those types of systems also cost highly. But in this digital era, CCTV systems are installed in most of the places because of its' affordable cost. In contrast, with the help of CCTV system, a Computer vision-based automated system can combat this type of crime, especially homicide, as well as different security areas like around airports, shopping mall premises, ceremonial parties, etc. Overall, computer vision-based approaches provide an inexpensive feasible alternative compared to sensor-based approaches.

In the present scope of this article, a comprehensive survey of the methods and results published over the last few decades in the context of detection of concealed and visual weapon has been specified. In a survey on weapon detection algorithms, Md. A.Slamani et al. [8] presents few methods of concealed weapon detection. In 2018, R. Mahajan et al. [9] published an article in a conference on various techniques for concealed weapon detection. However, [8] and [9] failed in represent a comprehensive survey and indicating specific challenges related to the weapon detection. On the contrary, we presents a comprehensive survey that explored specific challenges in detection of weapon and also reviewed published approaches in handling these challenges. The survey also indicate future scope of weapon detection problem. Furthermore, prime contribution of this comprehensive survey is a new taxonomy of computer vision based state-of-the-art methods in detecting weapons using different imaging modalities. Overall objectives of the survey work are:

(1) It presents a taxonomy based on the review work that able to describe published methods in a structured way for future reference.
(2) The review work revealed and analyze the challenges in the relevant field, which may lead future works.
(3) It also presents publications under each challenge and a brief comparison of the results published in these publications. The comparison enlightens the scope of future work in this field.
(4) A detailed description of the available datasets, along with their shortcomings, is represented in this work, which also paves the way for future research direction.

The rest of the paper is organized as follows. Section 2 discusses the challenges and benefits of the automatic detection of weapons. Section 3 describes each group of works of the proposed taxonomy of our survey. Section 4 presents an analysis of segmentation approaches adopted for weapon detection. The feature descriptors considered for the problem are analyzed in Section 5. Section 6 performances of the published research works under different challenges are presented. In the concluding section (Section 7), the vital takeaways from the survey are pointed and future scope related to the problem are discussed.

2. Challenges in weapon detection

Table 1 summarizes the possible challenges for a vision-based weapon detection system.

Table 1
Challenges in vision based automatic detection of weapon.

Category	Sub-category
Weapon Detection Environment	– High intra-class variations – Scene Complexity – Partial Occlusion
Digital Imaging	– Unusual Viewing Angle – Changing Illumination Condition – Image Blurring

Ideal gun/ knife images are often taken using different sensors (cameras), resulting in different intensities distributions of same image. Also, due to different imaging parameters, the length of weapon (in some unit of length, say cm, often refers to scaling effect) is mapped to different pixel resolutions for same weapon images.

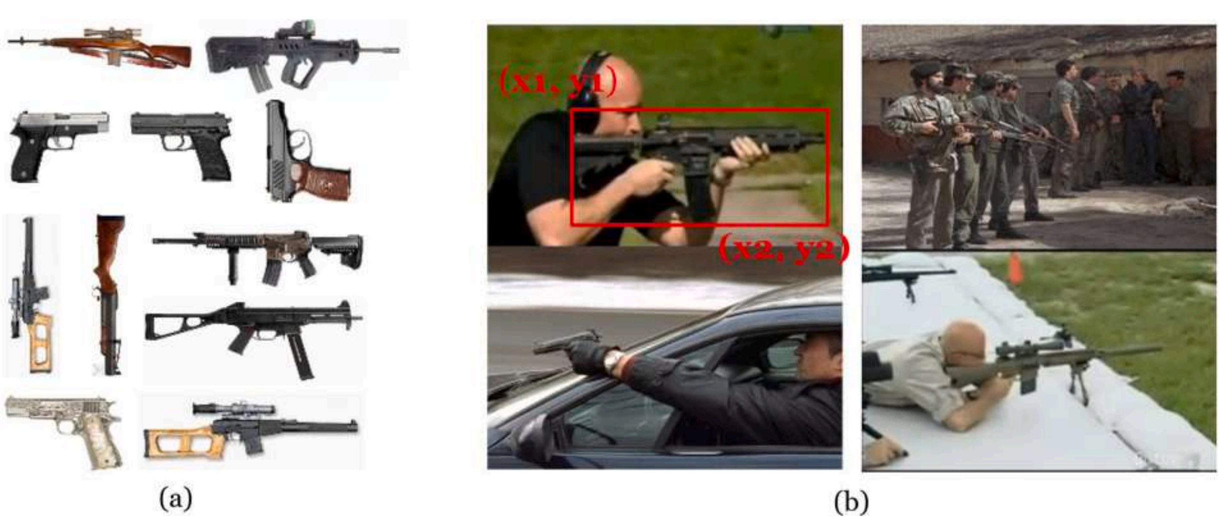

Fig. 2. IMFDB [5]: (a) Sample gun images are used for committing any crime, (b) sample firing images where gun are to be identified and localized. (x1,y1) and (x2,y2) are the spatial co-ordinates of a detected bounding box respectively. (x1,y1): upper-left co-ordinate and (x2,y2): bottom-right corners.

One of the major challenges of weapon detection, is high intra-class variations. This identification of major variation in shape and color for a wide variety of weapons demand high intra-class classification. Fig. 3 demonstrates a few examples of different varieties of shapes and colors of weapon.

Fig. 3. A few examples of different shape and color variations of weapon.

The weapon images are captures using different handheld cameras or using some mounted cameras. This often results in image blur due to camera shake and jitter (refer to Fig. 4(a)). Sometimes intruder's weapon may faces to the camera. Only the muzzle part is seen (refer to Fig. 4(b)), which is referred to as non-oblique viewing (fronto-parallel position of the camera with respect to the weapon) and sometimes weapon may appear perpendicular to the camera.

Fig. 4. (a) Blurred image due to camera shake and jitter; (b) Different viewing angle.

The challenge often extends when an object is moving with the gun. There might exist the problem of unusual viewing angle, this problem will lead to different rotation of weapon (refer to Fig. 5(a)) like panning and tilting of weapon, object carrying a gun can move toward camera and also can going away from the camera, this refers the problem of scaling of weapon (refer Fig. 5(b)).

Due to moving of object carrying a firearm, especially in indoor condition, there might exist illumination change. Because in the indoor condition, we rely on artificial lighting, which may not evenly be distributed in the whole area. This will lead to a slight change in the color of weapon (refer to Fig. 5(c)).

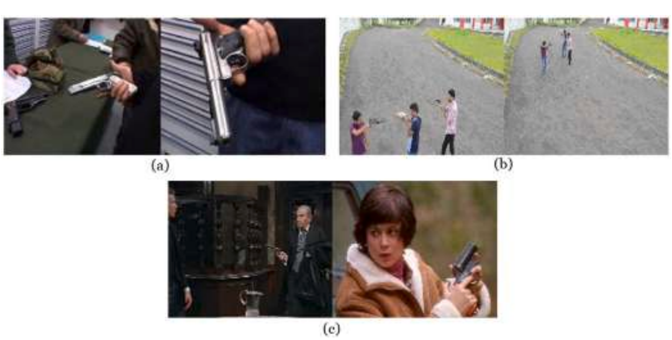

Fig. 5. (a) Different unusual viewing angle with rotation of weapon; (b) Scaling of weapon; (c) Illumination change for uneven distribution of light due to scene change.

Another challenges that make the problem more complex are partial or full occlusion of a gun (refer to Fig. 6(a)), deformation, loss of information due to transformation from the 3-D world into 2-D world. Partial or full occlusions occur because guns are mostly carried in either hand or holster. The problem also occurs when the color of gun looks a lot like a color of the background or alike the color of the dress code of that particular object (refer to Fig. 6(b)).

Fig. 6. (a) Shows the partial occlusion of weapon; (b) shows few examples where color of weapon is cluttered with background.

These reasons altogether pose significant challenge on top of typical object detection system studied in computer vision. The weapon detection problem bundles up various imaging modalities and various modalities of object detection problems like multiple object detection [10–12], detection of the multiple instances of the same object [13, 14], and multi-view object detection [15].

3. A taxonomy of computer vision approaches for handheld weapon detection

In 1996, ARPA, NIJ and Rome laboratory [16] initiated a joint program on the problem of concealed weapon detection. An automatic concealed weapon detection has potential application in security and surveillance. Afterwards, different sensors has been evaluated for visualization of the concealed weapon. Fusion of sensors based methods gained popularity regarding the visualization of the concealed weapon. Since 1998 automatic localization of concealed weapon have progressing with more involved approaches for fusion, recognition and localization. In contrast, detection of unconcealed weapon detection gained attention lately more specifically, after popularity of CCTV in security & surveillance. In 2009, Christos Grecos et al. [17] highlighted the facts regarding the challenges in monitoring CCTV footage for detection of suspicious objects. Christos Grecos et al. [17] also ensured a way out of these challenges by employing machine learning approaches for automatic detection of unconcealed weapon from CCTV footage. Since then, there are slightly more than 30 research publication directly related to the automatic unconcealed weapon detection system. In Table 2, we demonstrate a new taxonomy for automatic weapon detection. We consider both concealed and unconcealed weapon in the proposed taxonomy. It is important to refer that we consider the publications which are directly related to the automatic weapon detection using machine learning approaches.

From the pattern of the development over the last decade, we group the methods based on the used identification approach of the weapon. We group the methods into four different approaches: matching based, saliency map based, Multi-sensor based, classifier based and detector based.

The Table 2 also presents different areas of applications and corresponding categories of the problem. The area of applications are (CW) Concealed weapon detection, (UCW) Unconcealed weapon detection, (KD) Knife detection, (GD) Gun detection, (OW) Any Other weapon, (XW) Weapon detection using X-ray Imaging, (VW) Weapon detection using Visual Imaging, (IRW) Weapon detection using Infrared Imaging, (TW) Weapon detection using Terahertz Imaging. The categories of the weapon detection problem addresses in these publications are:

(DI) Complex Background: This relates to the backgrounds having higher number of false positives. (DII) High inter-class variability: This relates to the different shapes, types, colored weapons. For any category of weapon, these challenges are same. (DIII) Partial occlusion of weapon: Given a image with partially occluded weapon increase the difficulty in recognition of weapon. The challenges are briefly described in Section 2.

Performances of the approaches in handling these challenges is presented in Section 6. Next we elaborately describe each group and assess performances of each group of the taxonomy.

Table 2
A taxonomy of computer vision approaches for handheld weapon detection.

				Methods	Area of application	Category of challenges
Automatic weapon detection method	Matching based methods	Blob matching		[18]	UCW VW GD	DI DII
				[17]	CW GD TW	DII
				[19]	CW GD IRW VW	DI DII
				[20]	UCW GD VW	DII DIII
		Gradient based matching		[21]	CW XW GD	DI DII
				[22]	CW IRW VW GD	DII
		Grid based matching		[23]	CW XW GD	DI DII
				[24]	CW GD TW	
	Salience based methods			[25]	CW VW GD	DI DII
	Multi-sensor fusion methods			[26]	CW VW IRW GD	
				[27]	CW VW IRW GD	DII
				[28]	CW VW IRW OW	DII
				[29]	CW IRW GD	DII
				[30]	CW IRW VW GD	DII
				[31]	CW IRW VW TW OW	DII
				[32]	CW IRW VW GD	DII
				[22]	CW IRW VW GD	DII
				[33]	CW IRW VW GD	DII
				[34]	CW IRW VW GD	DII DIII
				[35]	CW IRW VW AG	
				[36]	CW IRW VW GD	DII DIII
				[37]	CW IRW VW GD	DII DIII
				[38]	CW IRW VW GD	
				[19]	CW GD IRW VW	DII
				[39]	UCW VW GD	DI DII DIII
	Classifier based method	Features from ROI	Traditional approach	[40]	CW GD TW	
				[19]	CW GD IRW VW	DII
				[41]	UCW OW XW	DII
				[42]	UCW VW GD	DI DII DIII
				[43]	UCW GD VW	DI DII DIII
			Deep learning based	[44]	UCW GD VW	DII DIII
				[45]	UCW GD VW	DIII
				[46]	UCW GD VW	
		Holistic feature based	Traditional approach	[47]	UCW VW GD	
			Deep learning based	[48]	CW XW OW	DIII
				[49]	UCW VW GD	DI DII
	Detector based method	Traditional approach		[47]	UCW VW GD	
				[2]	UCW KD VW	
				[50]	UCW KD VW	DII DIII
				[51]	CW OW TW	DIII
		Deep learning based		[52]	UCW VW GD	DI DII DIII
				[53]	UCW VW GD	
				[48]	CW XW OW	DIII
				[54]	UCW VW GD	DI DII
				[55]	UCW VW GD	DI DII
				[56]	UCW VW GD	
				[57]	UCW VW GD	DI DII
				[58]	UCW VW GD	DII
				[59]	UCW VW GD	DII
				[60]	UCW VW GD	DI DIII
				[61]	UCW VW GD	DI DII
				[62]	UCW VW GD	DII

3.1. *Template matching based methods*

3.1.1. *Blob matching based*

In blob based methods, the input image is clustered into certain segments, and the segment related to a gun have been extracted, called a blob. Discriminating features are extracted from that blob. The extracted feature vector is compared to the feature vectors of guns stored previously. The comparison is based on some matching procedures, and a matching score is calculated. If the matching score is higher, a predefined threshold, then the presence of the gun has confirmed.

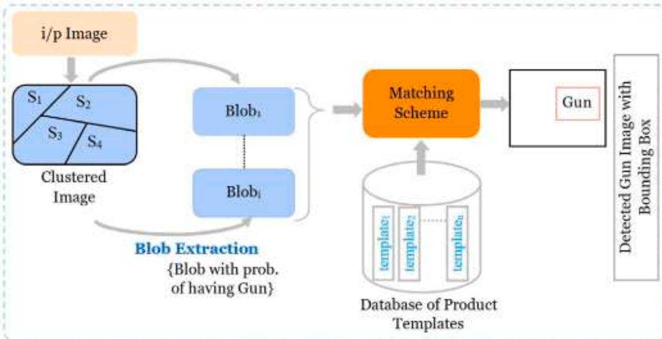

Fig. 7. Block diagram of a Blob Matching based method.

Fig. 7 has captured the basic framework of the Blob Matching Method. In 2015, G.K.Verma et al. [18] proposed blob matching based gun detection. They used SURF, a shape descriptor for matching purposes and at the primary step, they load SURF descriptor of stored guns. During detection, they used k-means clustering to the segment input image and extracted blob from the segmented image that is of an area greater than 1000 pixels. The sum of square distance (SSD) is used in [18] for matching the SIFT descriptor of the extracted blob to the stored ones. With more number of images, R.Kumar Tiwari and Gyanendra K. Verma improvised the previously proposed method with Harris plus FREAK descriptor in [20]. Harris plus FREAK descriptor implies FREAK descriptor of Harris corner of a gun. Other than this method described in [20] is very similar to the previous method. In [17], instead of searching the template in the entire image, they first employed a motion segmentation algorithm. Motion segmentation results in only the moving object. Afterward, the matching procedure follows, as described in the previous literature. They use only SIFT feature for the detection of weapons. Whereas in [19] fused image is used for the purpose of concealed weapon detection. Visual and corresponding thermal images are fused and on the fused image template matching is performed. They, after matching procedure, employed an SVM classifier to classify an input image as either an image with a weapon or image without a weapon.

3.1.2. Grid based methods

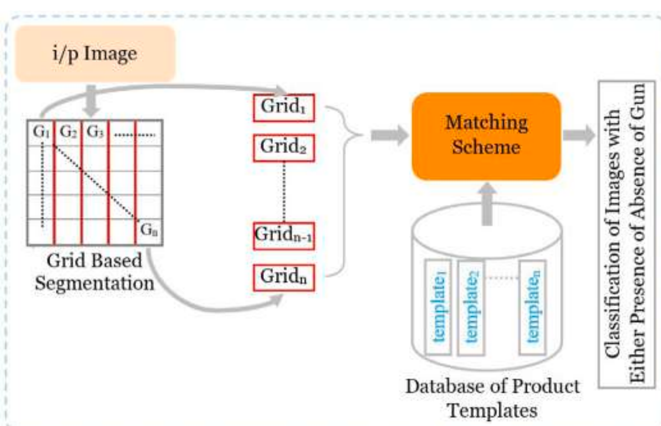

Fig. 8. Block diagram of a Grid based method.

In 2013, P.Vijayalakshmi et al. [21] proposed a grid based methods for gun detection. In the grid-based method, input image is divided into a number of overlapping grids. Afterward, features are calculated for each of the grid, as shown in Fig. 8. In [23], they concentrate on each type of features. To correctly identified a gun, they proposed a combined feature vector. The proposed feature vectors consist of both the shape features, texture features. Region Based shape features, as well as contour based shape features such as contour sequence moment and contour sequence central moment, are used. To represent the texture of gun Gray level co-occurrence based features are calculated. In addition to that, Gabor-energy based texture feature also used to capture texture information. Euclidean Distance is used for Matching purposes. The same approach has been used in [24] for the detection of weapons in CCTV videos.

3.1.3. Gradient based methods

Few matching based proposed methods for weapon detection used edge image for extraction of features. In the gradient based method, instead of dividing the input image into a number of blobs or grids, gradient information is extracted. Features are designed from the gradient information for further comparison. In [21], sobel edge detection is performed to obtain the edge image. Afterward, from the edge image SIFT features are extracted for the matching purpose. To reduce the computation time, they proposed to match a specified part of the gun instead of the whole gun image. The trigger guard portion has been used for the purpose. The trigger guard portion of the weapon is the common and unique part to be considered. SIFT feature is extracted from the trigger part of the gun and matched. [22] used the same procedure, but they matched the whole shape with the template instead of features. Hausdorff distance is employed to quantify the matching.

Observations: Template matching based methods are simple compared to the other category of the proposed taxonomy regarding implementation. Matching based methods with reduced time consumption also proposed by [17,19]. [17] and [19] performed motion segmentation to find the objects in the image. By detecting the objects, it just reduced the search space. Note that the procedure is applicable for videos, and therefore it is applicable in real life.

Disadvantage of the matching methods is the generation and maintenance of the template dataset. To make the template matching algorithm realistic, the template dataset should include different types of weapons in a different position. To generate such a dataset of templates is challenging in real life (see Fig. 9).

3.2. Saliency map based methods

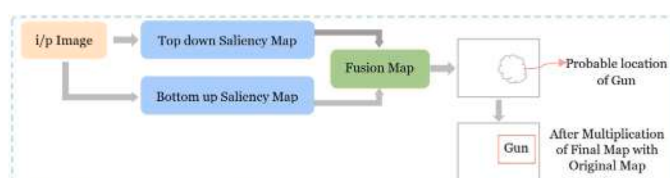

Fig. 9. Block diagram of Saliency Map based method.

Saliency refers to unique features (pixels, resolution, etc.) of the image in the context of visual processing. These unique features depict the visually alluring locations in an image. A saliency map is a topographical representation of them. These maps were first proposed by neuro-scientists Laurent Itti, Christof Koch, and Ernst Niebur in their study [63] on feature extraction in images. They give a detailed description which is given below.

"The purpose of the saliency map is to represent the conspicuity – or 'saliency' – at every location in the visual field by a scalar quantity and to guide the selection of attended locations, based on the spatial distribution of saliency. A combination of the feature maps provides bottom-up input to the saliency map, modeled as a dynamical neural network".

Saliency map based modeling influenced by bottom-up and top-down visual cues. The bottom-up visual attention is triggered by a stimulus, where saliency is captured as the distinction of image locations, regions, or objects in terms of low-level cues such as color, intensity, orientation, shape, T-conjunctions, X-conjunctions, etc. The task drives top-down visual attention. Top-down saliency models use prior knowledge, expectations, or rewards as high-level visual factors to identify the target of interest [Ali Mahdi et al.]. E.Ardizzone et al. [25] proposed a saliency map based method for firearm detection held by a person. In this process, they combine both the top-down and bottom-up saliency approach for firearm detection. Hence, E.Ardizzone et al. [25] used Firearm location w.r.t to the head of the person as the prior knowledge for top-down saliency generation. The prior knowledge obtained by calculating the posterior probability of firearm position conditioned by the face position, registered concerning the face center, and rescaled with respect to the face size. A model is generated by calculating the posterior probability of a number of training images. Afterward, the generated map is combined with the map obtained GVBS. GVBS is usually used to generate a bottom-up approach. The generated model is employed for firearm detection. They report good results in IMFDB dataset, mentioned in next section (Referred to the section:VI). The challenge with this architecture is its complexity, and furthermore, it required to generate different models as per the location of the face. For example, a model generated for a dataset having images, where the face is on the right side of the image will not work on the images where the face is right or center of the image.

Observation: Saliency features are considered very efficient as they are proposed based on the human visual system. [25] is the only method until now used a saliency-based approach for weapon detection. They used both top-down and bottom-up saliency map for weapon detection. This approach is computationally complex compared to the matching based method. The advantage of the method is it does not require to maintain and generate template dataset. The proposed method is able to detect any gun in the image or video. But the disadvantage of the method it cannot detect gun in any position. It can only identify the gun if any object in the shooting position holds it. Three models are generated for guns present on the left side of the image, present on the right side of the image, and center of the image.

3.3. Multi-sensor fusion based

Multi-sensor fusion-based methods used more than one sensor for weapon detection. The idea of using two sensors is more applicable for the detection of hidden weapons as a visual sensor is itself unable to detect the hidden weapon. The fusion of images taken from different sensors gives an intuition of the presence of weapons, as shown in Fig. 10.

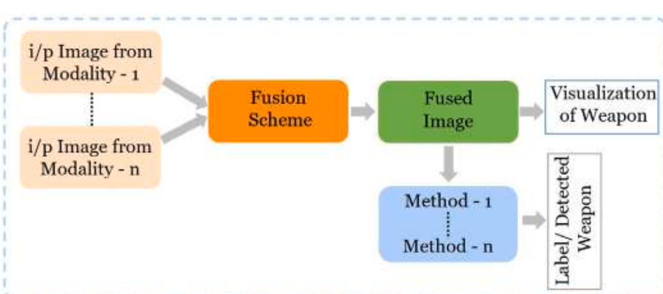

Fig. 10. Block diagram of Multi-sensor Fusion based method.

Fusion-based methods are used for visualization of a concealed weapon. For automatizing the detection procedure, any method described in the proposed taxonomy can be used. In Fig. 10, method 1 to method n referred to the methods described in the proposed taxonomy. In 1997, M. K. Uner et al. [35] proposed and proved the efficiency of a multi-sensor fusion based method for concealed weapon detection. An imaging method that capture temperature of the object will be useful for weapon detection, as weapons are made of metal, which has high emissivity. Therefore, fusing thermal imaging and visual imaging [35] provides a new insight into the detection of concealed weapon detection based on the fusion of different sensors. They used wavelet transform based fusion method for fusing infrared and visual images. The fusion of two images usually requires registration of images. To reduce the complexity of registration, Brut's first proposed a pyramid based fusion method. In these method, idea of fused image pyramid is used. For generating the fused pyramid, laplacian pyramid transform is applied on the source images obtained from different sensors. From these resultant pyramids on fused pyramid is generated using "maximum" feature selection rule. Finally, by taking an inverse pyramid transform a composite image is obtained and considered as a final resultant fused images. Afterward, different pyramid transforms and feature selection rules were proposed by various researchers. M. K. Uner et al. [35] used wavelet based transform over laplacian based transform as wavelet based transform has several advantages on laplacian based transform. For feature selection, they used Bruts's "maximum" feature selection method.

Since then, several literature used pyramid based fusion of multisensor images for concealed gun detection. L.C. Ramac et al. [33], proposed a method motivated from [25] and employed filtering to reduce artifacts before the fusion. [33,35] both showed fair accuracy in detection of concealed gun. The DWT is used in [22,26,28,29]but in different way. In [26], HSV color information of thermal images is used for fusing with visual image. They first combined the inverted infrared image with visual image by averaging afterward used DWT transform based fusion between combined image and HSV infrared image. They claimed to obtained better accuracy compared to the previous ones. In [29], they use DWT but for fusing infrared image and Passive millimeter wave (PMW) image. Likewise [33], they performed a two-way denoising filter for reducing blurring effect, as they aim to detect guns from a video sequence. Along with filtering registration of source, images are also performed using the maximization of mutual information (MMI) before fusion using DWT. Likewise [26], in [28] also exploits color information of visual and infrared image. They use V-channel of RGB image for fusion with infrared image and also enhance the contrast of the fused image. Afterward, they performed a couple of steps for extraction of the gun from the image such as contour detection and histogram equalization for enhancement of the obtained fused image. [22], they also used DWT but employed shape descriptor and shape matching for gun detection. After fusion, gun from the input image segmented by Gabor filter and for detection of edges from the segmented gun binarization and morphological closing performed. Stored gun shapes are searched and matched to the segmented gun of the input image. Hausdorff Distance measures are used for the purpose of shape matching.

Previously mentioned works used DWT as the primary fusion method for fusing multisensor images. A discrete wavelet frame (DWF) is a feature extraction method that uses DWT proposed by M.Unser et al. in [35], and they showed DWF performs better than DWT in extract texture features. DWF uses an over-complete wavelet decomposition (the discrete wavelet frame (DWF)) in which the output of the filter banks is not sub-sampled. Unlike other wavelet-based approaches, this should result in a texture description invariant with respect to translations of the input signal. Z. Xue et al. [37] first proposed a method that uses DWF transformation for fusing infrared image and visual image for weapon detection. Brut's algorithm is used for the selection of features from the transform. Z. Xue et al. [37] compared DWF with 15 different fusion algorithms and conclude DWF performs better than the others. Zhiyun Xue et al. [37] also employed DWF with Brut's algorithm for the fusion of visual and infrared image and exploit different color schemes to obtain better results. Z. Xue et al. [37], based on the comparison they have carried out, conclude that along with DWF, pixel-level maximum/minimum fusion also performs better. Pixel level maximum/minimum fusion is a straightforward fusion algorithm, where the maximum or minimum value of the source images pixel by pixel is used as the pixel of the fused image. Z. Zhang et al. [38]proposed an image fusion algorithm by combining the aspects of both pixel-level fusion and feature level fusion. They named the proposed method as a region based fusion algorithm. Where they used DWT to find the active regions and then used edge information of the region for fusion. The DWT transform is used to guide the fusion process carried out pixel-level fusion. Other fusion procedures other than DWT, DWF also used to fusing visual and IR images for the detection of a concealed weapon. In 2006, Alexander Toet [34] proposed Laplacian pyramid transform for fusion instead of DWT. To obtain better results, they used LMS color space i.e., LMS cone response space. LMS is related to the response of the three types of cones of the human eye. Afterward, principal component analysis, they rotate the axes in the LMS cone space to achieve maximal decor-relation between the data points. In LMS color space, primary component is achromatic channel and other two channels correspond to color opponent. [34] used pyramidal fusion scheme for fusion of achromatic channels of different sensor images. For better visualization, the mean and standard deviation of the fused luminance image is set to equal to those of the original image. The resultant fused image is transformed into RGB color space at the very last step. Whereas, Z. Liu et al. [30] used the Multiresolution mosaic technique(described in) for the fusion of IR and visual image. Registration is not required during this fusion; it is more like embedding on IR image on visual image. IR image is clustered first using the K-means algorithm to identify weapon in IR image. Then this probable IR gun image is embedded in the visual image for visualization of the concealed image. And Tuzhi Xu [27] proposed a Double density tree complex wavelet transform for the fusion of IR and visual image and the same for visualization. In the mentioned review works, one step of fusion is used, whether in E. M. Upadhyay et al. [36] three-step of fusion is presented, such as multi-focus fusion, multi-exposure fusion and after that both the fused images are fused to obtain the final fused image. 2D-CWT fusion is used for the multifocus fusion and blending function approach used for multi-exposure fusion. In [31], they used three source image for fusion, such as terahertz image, visual image, and infrared image. Terahertz image is used because it can penetrate various materials and also not harmful to humans.

Observation: Fusion-based methods are primarily employed for the detection of concealed weapon detection. In fusion-based methods, images from different modalities are fused for visualization or detection of concealed weapons. Detection of concealed weapons is not possible with only visual images. Therefore, the fusion of other modality image with the visual image is performed. Thermal images, X-ray images, terahertz images are able to visualize concealed weapons without fusion. Computationally fusion-based methods are complicated and time-consuming, but without fusion, detection of concealed weapons is a bit challenging task. After fusion, any method mentioned in the proposed taxonomy can be used to automatize the detection, which further increases the computational burden (see Fig. 11).

3.4. Classifier based methods

Here, classifier-based methods are not referring to classification methods used in computer vision depended on certain extracted features. Such classifiers are also used in previously discussed fusion-based methods. But in those methods, fusion of multi-sensor images are responsible for weapon detection.

Classifier based methods of the proposed taxonomy detailed the methods that used classifiers for weapon detection. Classifier based methods are also categorized into categories. One category noted the classifier based methods that depend on the features of the ROI where ROI is the gun area or object with gun area segmented from the input image. Another category describes the methods that classified input image based on the holistic features of the input image. In the first case, weapons are segmented from the input image, and afterward, extracted features from the segmented ROI are used for classification. The classification method classifies the input image, either as positive

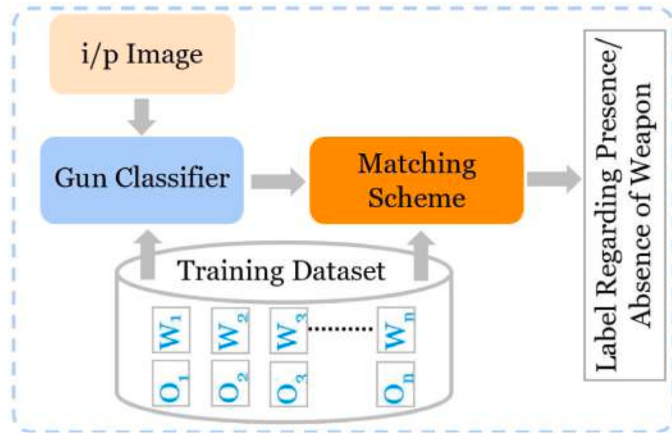

Fig. 11. Block diagram of a classifier based method; W_1, W_2, …., G_n are the Weapons and O_1, O_2, …., O_n are the other than weapon.

or negative. Positive with respect to the presence of the gun and negative for the absence of the gun. Therefore, segmentation of ROI is an important step to obtained discriminated features and correct classification. In 2008, A. J. Lingg [40] first proposed a classifier based method that extracted features from the segmented ROI, for that used active contour shape (ACS). ACS is a well-known method in shape-based segmentation method. ACS is based on energy minimization; energy minimizes when the initial contour finds the correct edge. The extracted ROI is refined by dilation and erosion. Traditional Fisher's linear discriminant employed for classification purposes. Shape dependent features are used, four moment based descriptors and two Fourier shape descriptors. Likewise [40], in [41] also perform segmentation of ROI. They compared region growing, Fuzzy connectedness, watershed level sets, and threshold based method. Among these algorithms, fuzzy connectedness performed better than the others. They extract Zernike descriptors and histogram shape index and Support Vector Machine (SVM) classifier for classification. Both [40] and [41] are used for detection of concealed weapons. [40] used terhertz image, whereas [41] used CT Scan image for implementing the proposed method. [19] used k-means segmentation for ROI segmentation. k-means algorithm performed on a fused image, as both infrared and visual image used here. SVM classification is used for the classification of input image. Unlike these [19,40,41,43,46] do not segment the gun directly rather they performed object detection algorithm, more specifically background subtraction algorithm. Background subtraction output a binary image, where moving objects are detected. Afterward, edge detection method employed for the detection of edges. Then based on the features of the edges classification has performed. [43] used simple frame differencing and [46] used VIBE [64] algorithm for background subtraction. CNN based classifier has been used in both [43,46].

In the previous para, we mentioned about the classification based methods where for classification features are extracted from the segmented ROI (i.e., gun area). As mentioned on the table, the classifier-based method can also be implemented by extracting holistic features from the input image. Holistic features based approaches become efficient in concerned classification(input image as the presence of a gun or absence of gun) after the introduction of Deep features. J. Lai et al. [49] first on 2017, introduce deep learning in the field of gun detection. They used googlenet, a deep learning network used to classify an input image, either positive (presence of a gun in the image) or negative (absence of a gun in the image). Afterwards on 2018, S.Akcay et al. [48] implement different deep network architecture for classification of images with weapon/ gun. They used a challenging dataset and analysis performances of different deep architecture. At first, they used pre-trained weights for the architecture and also trained the architecture to obtain weights for the architecture-specific to the detection of guns. Using quantitative comparison, they conclude that RESNET-50 performs better in comparison to the other networks. Likewise, [48], in [65] also a comparison of the deep network and traditional classifier network is carried out, and deep networks perform better than the conventional classifier. Holistic features are computed from the entire image, to reduce the size of the feature vector in [44], they first implement a deep detection method YOLO v2 for person detection. YOLO-v2 is a pre-trained architecture for person detection. After detection of the person from the image, deep architecture based classifiers are performed to classify input image as per the absence and presence of the gun with the reduced feature vector. The proposed approach performed well and logical because of the size of the gun. A gun or weapon in an image acquire a very small region of interest; therefore, holistic feature vector may dominate by the features of other objects greater than the gun/weapon. Proposed approach [44], able to overcome the challenges but for the shake of time and complexity.

Observation Now-a-days classification based methods are popular methods in computer vision. Classifier based methods are depending on the selection of features. For weapon detection, classifiers should be implemented in a different way. During weapon detection, the probable location of the weapon should be detected in the first step. Afterward, features from the weapon are compared with the features learned from training. Features from the whole image also can be used for classification but will not be efficient. After the invention of deep learning, classification using the features from the entire image are able to provide the correct classification. But for the small size of weapon make the deep learning-based classifier inefficient in the detection of weapon.

3.5. Detector based methods

Detector based method referring to the methods that not only classify but also localize the exact position of the gun in the input image. Referring to the Table 2, detector based methods are broadly classified as conventional methods and Convolution Neural Network-based methods. Conventional based methods are methods based on conventional image processing. Convolution Neural Network includes the methods based on the sliding window with CNN along with the deep architecture (see Fig. 12).

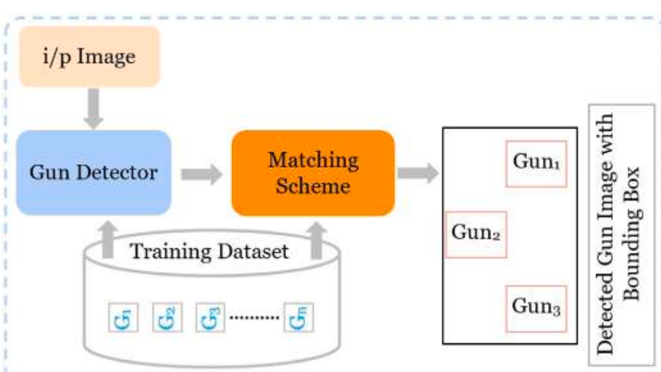

Fig. 12. Block diagram of a detector based method; G_1, G_2, …, G_n are the gun template.

Segmentation is a conventional method able to localize an object in an image. Based on the segmentation, [2,51] proposed gun detection method from the input image automatically. In [51], multi-label thresholding is performed for the segmentation of knives from the input image. Whereas in [2], active contour modeling (AAM) segment knife from the input image. Harris corner detector help to select initial seed for AAM model. This method is proposed for knife detection, and according to the proposal, it can only detect knives in the input image. [51] is also unable to detect any handheld gun from the input image, as the proposal is for the detection of concealed guns. The method is based on the edge of both the gun and the object under which the gun is contained.

The second category mentioned in the Table 2, specify the methods used for gun detection using a convolution neural network. Convolution neural networks are variants of neural networks with a large number of hidden layers. Specifically, they are based on deep features extracted in hidden layers. Convolution neural networks are proven very efficient in the computer vision field as they exploit a large number of deep features. CNN based classifiers are then evolved for localization of the objects in images through employing the sliding window method. These deep based detectors work on the entire image, and output image with bounding boxes bounded the object of interest. There are a number of pre-trained architectures present. [54] first proposed deep detector architectures, Faster RCNN, for gun detection. In addition, to reduce false positive, they exploit design issues in designing dataset for gun/weapon detection. [57] also exploits Faster RCNN, RFCN, and Single-shot detector deep architecture for weapon/gun detection. The detector architectures are primarily based on classifier models; the primary step of any deep based detector is classification. The performance of the detectors is based on the correct classification, i.e., the performance of the classifier. Considering this fact [57], compared the performance of these deep based detectors for different deep classifiers such as, Inception-ResNet-V2, ResNet-101, Inception-V2, ResNet-50, ResNet-101 and Inception-V2. The conclusion made by them is, FR-CNN with Inception-ResNet-V2 performed better for the detection of guns. They proposed a brightness modification based pre-processing to reduce the effect of brightness on steel weapons. [48] also compared deep based detection algorithms such as Faster CNN, R-FCN, YOLO-v2 and indicate YOLO-V2 is the best among the others. And [62] compared FRCNN, MobileNet, but before applying the detection algorithm, they subtract background using motion detection. They infer MobileNet performs better than the other two. [56,58,61] used deep based detection architecture such as YOLO, RCNN, and FRCNN for the detection of guns and observed good accuracy for each architecture.

Previously mentioned research works are able to attain good accuracy, but challenges related to the gun are not completely overcome. Challenges such as occlusion, different shapes of guns, lacking color texture information, and, more importantly, the small-sized guns are still present. These deep architectures described here are designed for the detection of certain objects such as persons, animals, vehicles, therefore to detect gun is a bit challenging for this architecture. The reason is gun has a different shape compared to the other, and the number of false positives is large in this case. Bottles in hand, pen in hand increase the number of false positives. To overcome these challenges, [52] proposed an orientation based deep architecture for the detection of guns. There are two-phase in this architecture. At the first phase, an orientation prediction module is trained to predict possible object orientation for each region proposal. Region proposals are generated through sliding window. Then ROI polling is employed where ROIs are cropped and wrapped with orientation information. In the second phase, deep based classifier has employed for prediction of the probability of a region proposal being a gun or rifle based on the deep features. They named the proposed architecture as Orientation Aware Object Detection (OAOD) architecture.

Observation: Detection based methods are performed well and have two-fold applications. In the classifier-based method, images are classified into either image with weapon or image without a weapon. In real life, classifier-based methods can be implemented in an automatic alarming system, whether the detection based method will provide the location of the weapon in the input image. Detection based methods are computationally complex and time-consuming. Detector based methods required time to complete the training procedure. These methods are very efficient, as shown in the result section.

Table 3
Segmentation based method.

Categories	Segmentation method	Approaches	Application
Clustering based methods	Color based K-means clustering	[18]	Gun segmentation from image
		[24]	Gun part segmentation
	K-means clustering	[19]	Gun extraction
		[20]	Visual Gun detection
		[30]	Weapon segmentation in thermal images.
Edge based segmentation	Canny edge segmentation	[26]	Post-processing after detection of gun
		[57]	Extract edge information
	Active contour based	[2]	Knife segmentation
		[40]	Handgun segmentation
	Sobel edge detection	[21]	Trigger guard segmentation.
Threshold segmentation	Otsu's Threshold	[41]	Gun Segmentation.
	Multilevel Thresholding	[51]	Concealed object segmentation.
Motion segmentation	Gaussian mixture model	[17]	Object with gun segmentation.
	Background segmentation	[53]	Object with gun segmentation.
Transform based Segmentation		[22]	Segmentation of weapon from Infrared image.

4. Segmentation methods based analysis

In this context of weapon detection, segmentation plays an important role. Success-full unsupervised segmentation of weapon from the input image can be consider for designing a automatic security monitoring or alarm-based detection system. In addition, in such system, where features are extracted from the ROI part of the image, segmentation of ROI plays an vital role. Correct segmentation has the probability to provide precise features for further classification or detection. For detection of any object (in this case weapon), sliding window based are employed for matching the features. Instead of sliding window, some of the work employed segmentation of the ROI. Therefore, this review on the segmentation method, able to provide direction to the researchers on segmentation of weapon from the input image. Table 3 describes segmentation method used till now in a simplified way.

4.1. Clustering based segmentation

Most commonly used clustering-based segmentation is k-means clustering in weapon detection. We can categorize k-means clustering in-terms of its application, such as color-based k-means clustering; another is gray level based clustering. As shown in Table 3, color-based k-means clustering is primarily used. The reason is weapons have specific color contrast. [18,24] used k-means clustering for the segmentation of weapons from the input visual image for feature extraction from ROI for further feature matching with the stored gun image. The same methodology is adopted in [19] but, [19]] used fused image. [18,19,24] empirically select the cluster number used for k-means clustering but did not mentioned. Whereas, [20], they mentioned 10 as cluster numbers for the segmentation of guns from the visual image.

Table 3 shows [30] used Fuzzy-k-means clustering on thermal images for detection weapons. They indicate that clusters with the highest center value correspond to the weapon. [30] uses Validity index such as partition index(SC), separation index(S), Xie and Beni's index(XB), and Dunn's index(DI) to find the optimal number of cluster.

4.2. Edge detection based segmentation

Edge detection is the primary step prior to the segmentation. The features of the detected edges is used for image segmentation. [26] used canny edge detection for extracting guns form the fused image. The fusion of visual and thermal image and enhancement of the fused image make visible the concealed gun. The canny edge detector is applied to the fused image for the extraction of boundary of guns.

Active Contour based segmentation is most commonly used for specific shape extraction. Active Contour is based on energy function and an initial seed. [2] used an active appearance model for knife segmentation. [2] deserve special attention as they proposed a simple seed selection for knife segmentation. [2] uses the Harris corner detector to find the tip of the knife. Whereas, in [19] a threshold based procedure is used for initial seed selection from concealed gun detection.

[21] used trigger guard segment for detection of gun. The presence of trigger guard segments in the input image specifies the presence of the gun. Sobel edge detection is used to find the trigger guard segment in the image.

4.3. Threshold based segmentation

Threshold Based segmentation is considered as the most simple segmentation method. The disadvantage of the threshold based segmentation is its dependency on the user-defined parameter. The resultant segmentation map solely depends on the selected threshold. The selected threshold usually depends on the used dataset. [41] compared threshold based segmentation to other segmentation approaches such as fuzzy-connectedness, region growing, and watershed level set. As a result of the comparison, they conclude that fuzzy-connectedness performs better other than thresholding based segmentation.

[51] in context with thresholding based segmentation proposed new multilevel thresholding for segmentation of concealed weapon along with the baggage. [51] used infrared images for the detection of concealed weapon detection. [51] found and noted that in infrared imaging, background pixels have the least intensity, whereas, concealed weapon contains pixels with higher intensity than background pixel but lesser intensity than pixels of the human body. By considering this fact, they employed multilevel thresholding for the detection of actual boundary of concealed weapons. Multilevel thresholding based on the fact that with a low threshold, the boundary of the human body shrinks inward continuously. Therefore, by multilevel threshold disjoint set of boundary pixels are extracted. From these disjoint sets, the concealed objects boundary is extracted by a simple set intersection. [51] deserves attention as [51] focused on accurate segmentation of concealed weapon and detection.

4.4. Motion segmentation

Motion segmentation is not employed for gun segmentation directly but indirectly helps weapon detection. Weapon detection in videos can employ motion segmentation. Motion segmentation detects moving objects and subtracts the background. Therefore, motion segmentation has two-fold advantages, such as it detects moving objects, and it reduces the search space for the detection of weapons. The underlying assumption is, there is a high probability that moving humans will carry weapons, or humans with weapons may become movable after few times. Based on this assumption, motion segmentation becomes the primary requirement for the detection of weapons. Weapon detection in video surveillance is the primary goal of this work. Inspired by this concept, [17] has employed the Gaussian Mixture Model(GMM) for the same. [43,66] used frame differencing for motion segmentation, as frame differencing is a simple and fast method for motion segmentation. In [53], frame differencing, Gaussian Mixture Model and visual background extractor (ViBe [64]). [53] also concludes that frame differencing is much faster and simple motion segmentation.

4.5. Transform based segmentation

[22] used Gabor filter for the segmentation of concealed weapons. Here weapons are hidden under the human body. They showed that after performing Gabor filtering with available six orientation, the hidden weapons are enhanced, and afterward, using binarization weapon is extracted. Note that the fused image has been considered in this work. The fusion of visual and thermal image with certain fusion technology showed up the hidden weapon (see Table 3).

5. Features based analysis

Feature extraction is the primary step in any computer vision algorithm. Based on the discriminative feature, a discriminative classifier or detector can be designed. The review focuses on the existing research on weapon detection. Concealed and unconcealed weapon detection of both types of weapons is considered here. Intuitively, a set of discriminative features are required for automatic detection of weapon which specifies the weapon in terms of image. The feature descriptors for the problem under consideration are broadly classified into five categories, such as key point-based features, gradient based features, region based features, texture feature, and shape-based feature. Table 4 represents the related work based on these categories, and a brief description presented as follows:

Table 4
Feature based method.

Categories	Feature Descriptors	Approaches
Key point based features	SIFT [67,68]	[17,21,69]
	SURF [70,71]	[18,69]
	Harris Corner Detector [72]	[2,20,24]
	Fast Retina keypoint (FREAK) [73]	[20,24]
Gradient Based features	HOG [74]	[42]
	Edge Histogram descriptor [75]	[43]
	Normalized Fourier descriptor [76,77]	[40]
	F-measure & gmeans	[63,78]
Region Based features	Haar-like features [79]	[47]
	Moment Based Descriptor [80]	[23,24,40]
	Zernika Moments descriptor [81,82]	[19,41]
Texture Based features	GLCM based descriptor [83]	[23,24]
	Gabor feature [84]	[23,24]
	Homogeneous texture Descriptor [85]	[43]
Geometry Based Features	Histogram shape index [86,87]	[19]
	Eccentricity descriptor [88]	[78]

5.1. Key point based features

Key point-based features (SIFT, SURF) are mostly used for matching based weapon detection from images/videos. The characteristics of the keypoint based feature make these features more applicable during matching. Rotation, scaling invariant key point-based features are mostly used for the considered application. Binary Keypoint based features are fast, less complicated compared to the other keypoint based features. Therefore, [20] used the FREAK feature for a training based weapon detection system. In [20], they used keypoint based features for classifying input image into either positive(presence of weapon) or negative(absence of weapon). They extract the interest region using color-based k-means clustering. [19] used keypoint descriptor to select initial seed for active appearance model. Based on that seed active appearance model segment knife from the image [19] automatically.

5.2. Gradient based methods

HOG is mostly used gradient based feature in computer vision as a shape descriptor. HOG focuses on the complete shape of the object. Unlike, edge features HOG not only considers edge magnitude but also takes account orientation information of edges. The advantage of HOG feature is, it is rotation, scale-invariant features. [42] shows promising results based on HOG features. Edge oriented histogram(EOH) features are a similar feature as HOG, but EOH only considers edge features, not the edge orientations. The advantage of EOH over HOG is of lesser complexity. HOG feature cannot handle light reflection; therefore, [43] uses EOH for knife detection. [40] used contour-based feature Normalized Fourier descriptor, here boundary information from Fourier transform is used. In order to remove dependence on position, orientation, and scale [40] proposed a new variant of Fourier descriptor, namely normalized Fourier descriptor for detection of weapon. [63,78] used canny edge-based feature vector f-measure and gmeans for the detection of weapons. [63,78] neither mentioned specific advantages nor draw any conclusion through the advantage of the results of using these features.

5.3. Region based features

Region based descriptors are the features that are extracted from a region specified by a rectangle. These features are related wavelet transforms of images. Therefore, these features are scale, rotation, and position-independent. [47] extract 2-rectangle features, 3-rectangle features and 4-rectangle features for weapon detection.

5.4. Boundary based features

Contour based moment features are performed well compared to area-based moment features used in [4,9,11,23,24,40] for weapon detection. This kind of feature is extracted from a binary image by considering a closed boundary. Based on the difference of each boundary point with the centroid normalized contour sequence moment and contour sequence central moment is calculated for shape representation. These features are translation, rotation, and scale-invariant, besides simple and require less computational time to calculate. Other features are calculated using these two terms. Zernike moments are also moment based feature but compared to other Zernike moments are powerful descriptor for their orthogonal nature. The disadvantage of Zernike moments is complexity.

5.5. Texture based feature

Texture based are a well-known feature used for shape detection. Here, the texture of the query image is compared with the texture of the stored template. Another way to implement texture-based features is to adopt texture features of the considered object by training procedure. The gray level Co-occurrence matrix is calculated from pixel pairs, which reflect orientation, homogeneity of texture of the image. Based on this matrix, certain features are calculated and used for recognition. Gabor features are among the top performers in face recognition and fingerprint matching. Gabor features extract local pieces of information which are then combined to recognize an object or region of interest. [23,24] used Gabor texture features along with the GLCM texture features. [43] makes the use of Gabor transforms, radon transform, and Fourier transform for the detection of weapons by using a homogeneous texture descriptor. The homogeneous texture descriptor describes the directionality, coarseness, and regularity of patterns in texture images.

5.6. Geometry based features

Histogram of shape index is a local surface shape measure based on the curvature of the object. It is constructed by segmenting the range of the shape index curvature measure into equal-sized histogram bins. The shape index represents the shape of a local surface by a single value angular measure. [41] make use of histogram shape index with zernika moments and achieve good classification result for concealed gun detection. Another geometry-based feature is eccentricity, the feature is calculated from a binary image. [63,78] used eccentricity on thresholded image for extraction of features afterward combining these features with canny edge feature detect weapon from the image. These features are scale, orientation, and position-independent and compared to the other lesser complex.

6. Comparison of published weapon detection methods

There exists more than 60 published papers for detection of weapon using different imaging modalities. We compared the performance of the existing approaches based on the published results of these approaches and described in this section briefly. Two constraints considered here for comparing the performances of the existing approaches: (a) Different approaches used several evaluation protocols for evaluating the performance of the proposed approaches. (b) Different approaches considered different challenges in weapon detection. The limited availability of benchmark public datasets is yet another bottleneck. We begin the comparison by presenting the details of public datasets.

6.1. Publicly available datasets

Table 5 lists the publicly available datasets indicating the year of publication, the number of images contains in each database along with the number of positive and negative images, the type of weapons present in the database, and also the type of database whether it is image database or video database along with image/ video types and pixel resolutions.

6.1.1. Internet Movie Firearm Dataset (IMFDB)

Internet Movies firearm database is a large image dataset of firearms. It is an online repository maintained by powered-wiki and publicly available at [5]. Approximately, it contains 4,50,000 images of the firearm; some are shown in Fig. 13. The database is composed of several thousand images taken from movie scenes or games. Few images also contain closeups of guns. IMFDB also contains images that had weapons that were occluded by darkness or rendered unseen able due to blurriness or scale. IMFDB dataset is an ideal dataset for guns as a variety of gun images, guns images with different orientations, different positioned guns are available in IMFDB. Images are middle-quality color spanning from 0.06 to 2 megapixels.

Fig. 13. Few samples of IMFDB [5].

6.1.2. Knives images database

The database contains a total of 12,899 images, with two classes of images: positive examples (PE) if the image features a knife (Fig. 14(a) and negative examples (NE) in all other cases (Fig. 14(b). The images were taken indoors or through car windows, since carrying knives in public is illegal in Poland. Out of 12,899 images, there are a total of 9340 NE, and 3559 PE is present in this database. All the images of the database are of resolution 100 × 100.

Fig. 14. Knives Image Dataset [43]: (a) Positive Sample (b) Negative Sample.

Table 5
Statistics of publicly available dataset.

Database	Publication year	No. of images/videos	Image format	Database type	Resolution	Weapon type
IMFDB [5]	2014	4,50,000*	.jpg	Image	Variable Size	Gun
Knives Images Database [43]	2015	12,899	.bmp	Image	100 × 100	Knife
Gun Movies Database [43]	2013	7 Videos	.mp4	Video	640 × 480	Pistol
Dataset of R. Olmos et. al [54]	2018	9261	.jpg	Image	640 × 480	Gun
Dataset of D. Ramerio et. al [44]	2019	17,684	.jpg	Image	224 × 224	Gun
ITU Firearm Dataset [52]	2019	10,973	.jpg	Image	480 × 800	Gun
CT based UBPS Dataset [41]	2010	236	.jpg	Image	60 × 60	Gun, Bottle
GD-Xray Dataset [89]	2018	11,267	.png	Image	Variable Size	NP

*Approx total images, NP = Not provided

6.1.3. Gun movies database

Gun Movies Dataset is a video dataset captured by CCTV cameras. CCTV cameras are mostly used for surveillance and security. Due to the non-availability of real-life gun-shooting videos, Grega et al. [43] generated this dataset by mimicking a gun-shooting condition. Therefore, this dataset consists of recordings of CCTV with an actor. There are 7 video recordings with 8.5 min recording. A total of 24000 frames are extracted from the 7 videos with frames where the firearm is not present. Few frames with weapons shown in Fig. 15. The videos are captured with a 640 × 480 resolution.

Fig. 15. Few samples of Gun Movies Database [43]. (a) Positive examples (b) Negative examples.

6.1.4. Dataset of R. Olmos et al.

R.Olmos et al. proposed both knife and handgun datasets. A total of 19,381 images are present in the knife dataset. They include i) cold steel weapon of diverse types, shapes, colors, sizes and made of different materials ii) knives located near and far from the camera, iii) knives occluded partially by the hand, iv) objects that can be handled in the same way as knives and v) images captured in indoor and outdoor scenarios. The images are downloaded from youtube videos or other sources of internet. Fig. 16 shows an example sample of the knife, where, in one image knife is closer to the camera and in other images, where the knife is far from the camera. It also shows, the dataset contains different knife with different shapes

Fig. 16. Sample images of Olmos et al. [54] knives dataset. (a) Positive examples (b) Negative examples.

They design two datasets: (a) One consists of 102 classes with a total of 9261 images. The pistol class has 200. The dataset is appropriate for classification task (b) the Second dataset contains 3000 images of guns with rich context. This dataset is suitable for detection purposes (c) The third dataset contains a total of 608 images, of which 304 are images of pistols. This dataset can be used for both classification and detection purposes. The dataset contains pistols in different position as shown in Fig. 17

Fig. 17. Sample images from the dataset of R. Olmos et al. [54].

6.1.5. David Ramerio et al.

The dataset images are obtained from the web such as "Google", "Instagram", and "YouTube and can be categorized into two straightforward classes. One class composed of images of people holding gun whether another class consist of images of people without gun. The dataset contains a total of 17684 images. Among them, 8843 images contain handguns, whereas rest 8841 images do not contains handguns (see Fig. 18).

Fig. 18. Sample images from the dataset of D. Ramerio et al. [44].

6.1.6. ITU firearm dataset

ITUF dataset consists of images of Guns and Rifles from different scenarios of practical importance such as being pointed, being carried, lying on tables, ground, or in racks. These variations allow machine learning algorithms to overcome dress variations, body pose variations, firearm pose and size variations, varying light conditions, and both indoor & outdoor scenarios making a strong prior for data-driven algorithms. Some sample images from the dataset are shown in Fig. 19. The images of the dataset are collected using web scraping by incorporating keywords such as weapons, wars, pistol, movie names, firearms, types of firearms, sniper, shooter, corps, guns, and rifles. The results were cleaned to remove images not related to firearms, cartooned images, and duplicated images. The final clean dataset consists of 10,973 fully annotated firearm images containing 13,647 firearm instances.

Fig. 19. Sample firearm images of ITU Firearm Dataset [52].

6.1.7. CT based UBPS dataset

CT based UBPS database contains two threat object classes: guns, and bottles. The number of images within this dataset is 236. Sets of 7 different guns and 35 bottles are used. Therefore, they conducted experiments on the dataset of bottles. Therefore, the training set used in the experiments consists of 24 bottle volumes (positive samples) and 55 non-bottle volumes (negative samples). Fig. 20 shows an example set of positive samples.

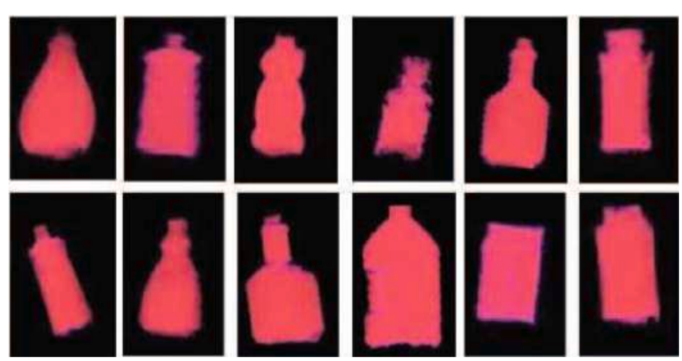

Fig. 20. A representative set of samples used in CT based UBPS database [41].

6.1.8. GD-Xray dataset

GD-Xray Dataset is a baggage X-ray imagery designed by Samet Akcay et al. The dataset is consist of four sub-dataset, such as Dbp_2, Dbp_6, Full Firearm vs. Operational Benign - (FFOB), Firearm Parts vs. Operational Benign - (FPOB). Each image of the dataset is of size 256×256 Each sub-dataset is described as follows:

Dbp_2: The Dbp_2 sub-dataset is consist of 11,627 X-ray images constructed using single conventional X-ray imagery. The dataset is further processed by cropping baggage objects manually and labeling each accordingly. During cropping, negative images consist of objects other than firearm also generated. Following these approaches, the dataset Dbp_2 has 19,398 image patches for a two-class firearm classification problem. The exact statistics as follows: positive class: 3179 firearm images/1176 images of firearm components; negative class: 476 images of cameras, 2750 knives, 1561 ceramic knives, 995 laptops and 9261 cropped images of background clutter. Fig. 21 shows a sample image of X-ray baggage; the image patches extracted from this image has been shown using red boxes.

Dbp_6: Dbp_6 sub dataset is more complex dataset than Dbp_2. The Dbp_6 dataset is constructed for 6-class classification problem. More specifically, each image patches are labelled with one of the six class. DBp_6 dataset can be categorized as follows: **Full Firearm vs. Operational Benign - (FFOB):** comprising 4,680 firearm threat and 5,000 non-threat images, and is denoted as FFOB. **Firearm Parts vs. Operational Benign - (FPOB):** contains 8,770 firearm and parts threat and 5,000 non-threat images (denoted FPOB, comprising of annotations as any of bolt carrier assembly, Pump action, Set, Shotgun, Sub-Machine-Gun).

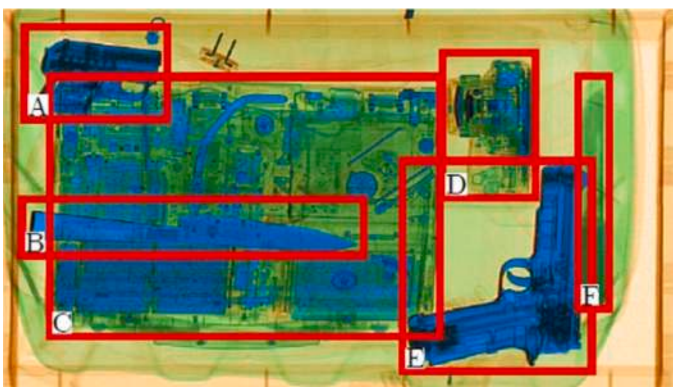

Fig. 21. Exemplar X-ray baggage image [89] with extracted dataset regions including background samples. Type of baggage objects in the dataset is as follows: (A) Firearm Component, (B) Ceramic Knife, (C) Laptop, (D) Camera , (E) Firearm , (F) Knife.

6.2. *Comparison using published results*

In this survey, an extensive comparative study is conducted per category of the mentioned challenges: (DI) Complex background, (DII) High Intra-class Variability, and (DIII) Partial Occlusion. These challenges are defined and discussed in the previous section. Note that the results listed in the following tables or figures are reproduced from the respective publications. In a few cases, we are unable to produced results of some publications due to the imprecise definition of evaluation procedure. Some publications indicate qualitative results instead of quantitative results; in this case too production of the result is not possible. In the next section, we represent a comparative study of weapon detection that compensates for different challenges, as mentioned.

6.2.1. *(DI) complex background*

Based on the survey, we find three publications [1,29,46] under this category for the comparison. The comparison between the methods, also for the following comparison, recall, precision, and accuracy, are calculated. Accuracy, precision, and recall are calculated using True Positive(TP), False Positive(FP), True negative(TN), and False Negative(FN). TP, FP, TN, and FN are the most common measures used in computer vision. The interpretation of these measures differs according to the application. For weapon detection using classifiers, TP referred to the number of images correctly classified as positive images. Positive represent the presence of a weapon in the input image and vice versa for the negative image. False-positive referred to the number of images are correctly classified as negative image as a positive image. Whereas, TN indicates the correctly classified negative image and FN indicates the number of incorrectly classified negative images. In general, a high value of TP and TN, along with lower values of FP and FN considered a good performance of the proposed methods. Consequently, the accuracy, recall, F1, LR+, LR-, Kappa Coefficient are defined as:

- **Accuracy** $= \frac{TP+TN}{TP+FN+FP+TN}$,
- **Recall** $= \frac{TP}{TP+FN}$,
- **F1** $= \frac{2TP}{TP+FP+FN}$,
- **LR+** $= \frac{TPR}{FPR}$,
- **LR−** $= \frac{FNR}{TNR}$ and
- **Kappa Coefficient** $= \frac{Accuracy-RA}{1-RA}$

Where,

- True Positive Ratio (TPR) $= \frac{TP}{TP+FN}$,
- False Positive Ratio (FPR) $= \frac{FP}{FP+TN}$,
- False Negative Ratio (FNR) $= \frac{FN}{FN+TP}$,
- True Negative Ratio (TNR) $= \frac{TN}{TN+FP}$,
- RA $= \frac{(TN+FP)(TN+FN)+(FN+TP)(FP+TP)}{n^2}$, n is the total number of subject.

Definitions of TP, TN, FP, and FN vary for the methods based on localization of the methods as follows: If the center of detected bounding box lies within the corresponding ground truth box and the detected object is the weapon, then the detected weapon is considered as TP. If the detected object is not a weapon, then it is considered as FN. Whereas, if the center of the detected weapon box is outside of the ground truth box, then the detected weapon is considered as FP. Otherwise, the detected weapon is considered as a TN. In Tables 6, 7, 8, 9, 10, 11, the detection methods are shown using a different color.

Few methods, are evaluated in private dataset and few on public dataset. We mentioned the publicly available dataset in the previous section. The Table 6 presents the results of the methods evaluated on private dataset and Table 7 shows the results of the methods evaluated on publicly available dataset. From both the table we can observe that, [42] and [39] outperforms other methods in classifying and localizing weapon.

6.2.2. *(DII) high intra-class variability*

In order to detect weapons, high intraclass variability is the primary challenge, as mentioned before. Thus, the taxonomy is proposed primarily based on the methods that can handle high intra-class variability effectively. In this section, we also show, the comparison of

Table 6
Comparison of detection of weapon from a scene with complex background using private dataset.

Publications	Detection result						Data specifications		Required time
	Accuracy	Recall	F1	LR+	LR−	Kappa coefficient	Negative	Positive	
[18]	88.67	86.67	0.9285	–	0.1333	0.8824	15	13	–
[21]	84	Nil	–	–	–	–	14	12	6.5 s

Table 7
Comparison of detection of weapon from a scene with complex background using publicly available dataset.

Dataset	Publications	Detection result						Data specifications		Required time
		Accuracy	Recall	F1	LR+	LR−	Kappa coefficient	Negative	Positive	
IMFDB	[25]	–	65	0.50	–	–	–	–	–	–
	[55]	92.6	–	–	–	–	–	–	–	–
Olmos et al.	[39]	93.16	0.7760	0.7453	1.311	0.6505	0.1912	3000	–	–
	[54]	90	100	0.9143	1	0	0.6844	3000	–	0.19 s
	[57]	–	80.02	–	–	–	–	618	9421	0.02 s
	[61]	85.44	–	–	–	–	–	4425	7920	–
Knives Image Dataset	[42]	84.6	–	–	–	–	–	1394	309	0.58 s
Gun Movies Dataset	[43]	–	81.8	–	–	–	–	12,000	12,000	–
David Romero et. al.	[44]	90.95	90.97	0.9096	10.07	0.0997	0.8191	8843	8841	–
	[60]	86.12	86	0.8641	6.2656	0.1628	0.7224	1361	1362	–
ITU Firearm Dataset	[52]	84.7	–	–	–	–	–	1253	848	–

Table 8
Comparison of detection of weapon methods using private dataset under high intra class variability category.

Publications	Detection result						Data specifications		Required time
	Accuracy	Recall	F1	LR+	LR−	Kappa Coefficient	Negative	Positive	
[18]	92	86.67	0.9285	–	0.1333	0.8823	15	13	–
[21]	84	–	–	–	–	–	14	12	6.5 s

the methods is conducted based on the proposed taxonomy as well as based on the features. Similar to the previously discussed challenge, the performance of the method under this category, evaluated using accuracy, recall, and precision. The definition and interpretation of TP, TN, FP, and FN are the same as before.

Table 9 lists the publicly available benchmark dataset-specific results reproduced from the respective paper, while Table 8 presents the results on a private dataset.

All the measures are not calculated in each and every publication. Therefore, some measures of specific methods are not present. Note that state-of-the-art methods claimed that proposed methods are able to handle more than one challenge or all the challenges. Therefore, such methods are replicated in comparison tables specific to the challenges.

Fig. 22 represents the best and worst performance of the proposed taxonomy over all datasets. In addition, Fig. 23 presents the best and worst performance of the features over all the datasets. From Fig. 22, it can be noticed that detector based methods show consistent performance. Note that some taxonomy is not included in Fig. 22 as publications are not mentioned clear representation of results. From Fig. 23 , the superiority of region based feature in detection of weapon. Note that, till now, deep learning-based features are not exploited in the detection or localization of the weapon in the input image. Region based and Gradient based features are also used in a few publications.

In Fig. 24 , we present the best performance of methods on all the publicly available datasets. From the Fig, we can conclude that dataset

Table 9
Comparison of detection methods of weapon in public Dataset under High Intra Class Variability category.

Dataset	Publications	Detection result						Data specifications		Required time
		Accuracy	Recall	F1	LR+	LR−	Kappa Coefficient	Negative	Positive	
IMFDB	[25]	–	65	0.5000	–	–	–	–	–	–
	[55]	92.6	–	–	–	–	–	–	–	–
	[49]	89	–	–	–	–	–	2535	218	1.3 s
	[59]	–	84.8	–	–	–	–	185	135	–
Olmos et al.	[39]	93.16	0.7760	0.7453	1.311	0.6505	0.1912	3000	–	–
	[54]	90	100	0.9143	1	0	0.6844	3000	–	–
	[57]	–	80.02	–	–	–	–	618	9421	0.02 s
	[61]	85.44	–	–	–	–	–	4425	7920	–
Knives Image Dataset	[42]	84.6	–	–	–	–	–	1394	309	0.58 s
	[50]	95	–	–	–	–	–	1211	3176	–
Gun movies Dataset	[43]	–	81.8	–	–	–	–	12,000	12,000	–
David Romero et. al.	[44]	90	90.97	0.9096	10.07	0.0997	0.4997	8843	8841	–
	[60]	86	85	0.8612	6.2656	0.1628	0.7224	1361	1362	–
ITU Firearm Dataset	[52]	84.7	–	–	–	–	–	9340	3559	–
CT based UBPS	[41]	98.18	100	–	–	–	–	61	65	–

In [25], recall is calculated from the Recall Precision Graph. As accuracy is not mentioned in [25], F1 measure is considered as accuracy.

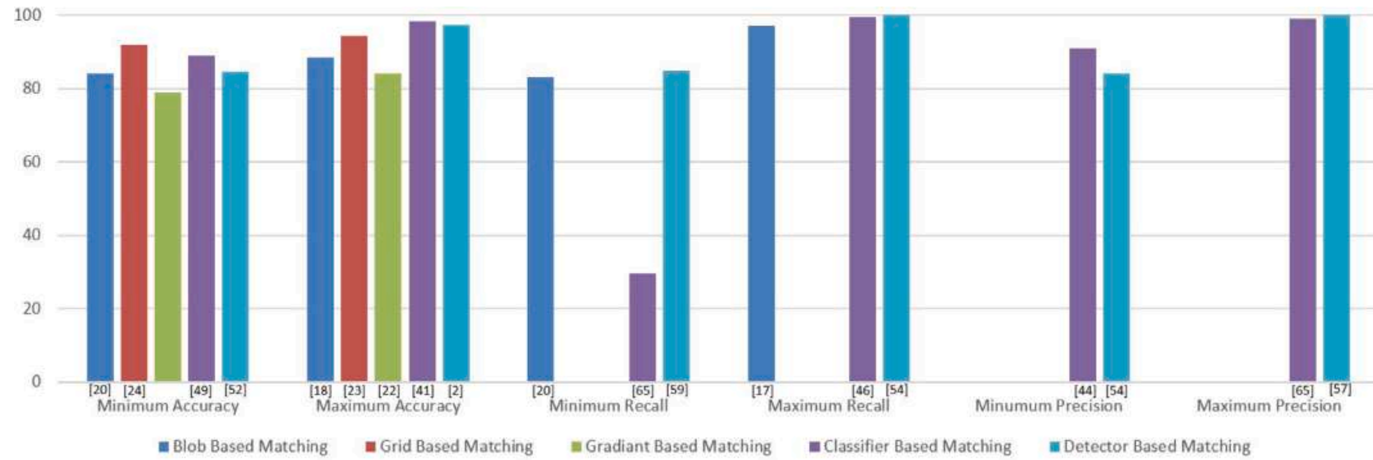

Fig. 22. Performance analysis of proposed taxonomy: not all references have reported accuracy, recall and precision measures.

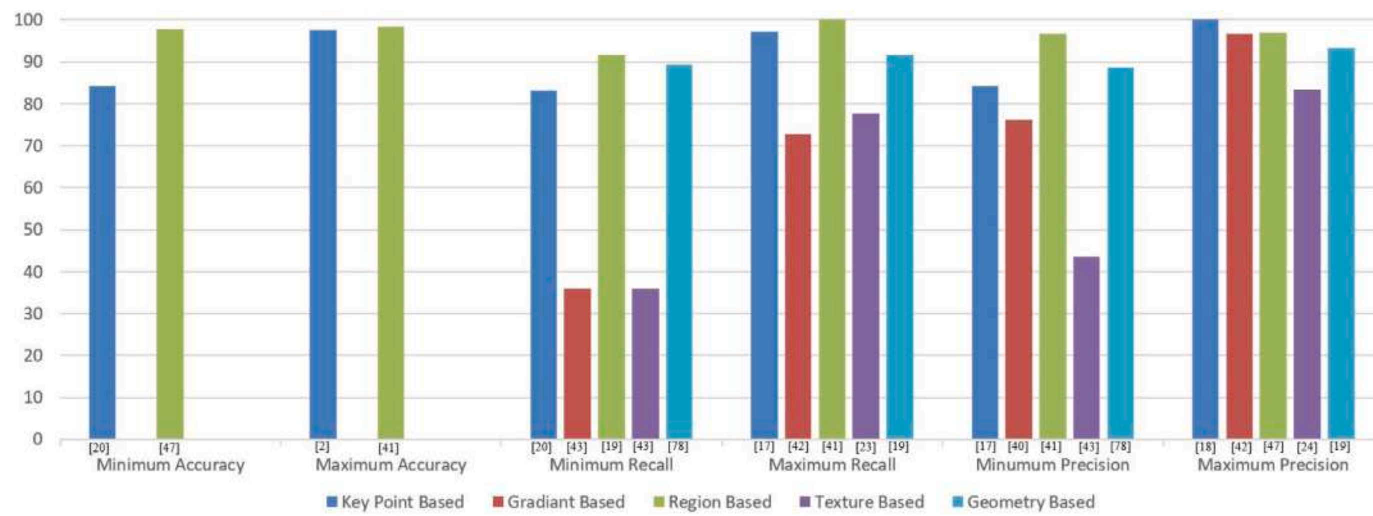

Fig. 23. Performance analysis of features: not all references have reported accuracy, recall and precision measures.

generated from online sources are more challenging, and methods are not able to attain more than recall values in these datasets.

6.2.3. (DIII) partial occlusion

The methods under this category deals with an important challenge, occlusion of weapon. Detection of fully occluded is near to impossible task, even classifying image based on the presence of fully occluded weapon is very challenging task. Even human perception of fully occluded weapon of an image is challenging. Therefore, for machine it is even more difficult task.

Detection of partially occluded weapon is implementable in real time and has a vast application in surveillance & security. Yet the issue is not addressed successfully after a large number of publications. We found out 13 publications work with occlusion. Among these publications, according to results and discussion, we found [48] addressed partial occlusion as a primary challenge in the detection of concealed weapons and claimed to handle partial occlusion successfully.

Other publications addressed the challenges, but cannot handle the partial occlusion successfully, specifically in the detection of visual handguns. These publications are not able to detect handguns when a small part of a gun appears in the image. Table 10 and Table 11 showing the performance of methods that claimed to handle partial occlusion in the private dataset and publicly available dataset respectively.

Table 10
Comparison of detection of weapon algorithm to compensate partial occlusion.

Publications	Detection result		Data specifications	
	Accuracy	Recall	Negative	Positive
[20]	84.26	83.07	65	54

In Table 10, publication [51] detect weapon by employing segmentation method. [51] proposed Multilevel thresholding based segmentation that efficiently segments hidden weapon in X-ray baggage image. They present qualitative results to prove the efficiency of the proposed method.

7. Summary and concluding remarks

Weapon detection is an important application of computer vision in the field of security and surveillance. In recent years, a huge amount of works are published for weapon detection with the aim of increased performance. In contrast, there are few works that exploit certain problems, challenges, and enhance applications. Here, we present a comprehensive survey on state-of-the-art methods according to the performance in handling specific challenges. In addition, we also exploit challenges in weapon detection and analyze the methods under different challenge category.

Table 11
Comparison of weapon detection methods in publicly available dataset under partial occlusion.

Dataset	Publications	Detection result						Data specifications		Required time
		Accuracy	Recall	F1	LR+	LR−	Kappa Coefficient	Negative	Positive	
Olmos et al.	[39]	93.16	77	.7453	1.311	0.5492	0.1912	3000	–	–
Knives Image Dataset	[42]	84.6	–	–	–	–	–	1394	309	0.58 s
Gun Movies Dataset	[43]	–	81.8	–	–	–	–	12000	12000	–
David Romero et. al.	[44]	90	90.97	0.9096	10.07	0.0997	0.8191	8843	8841	–
	[60]	86	85	0.8614	6.2656	0.1628	0.7224	1361	1362	–
ITU Firearm Dataset	[52]	84.7	–	–	–	–	–	1253	848	–
GD-Xray Dataset	[48]	99.60	–	–	–	–	–	8770	5000	0.1 s

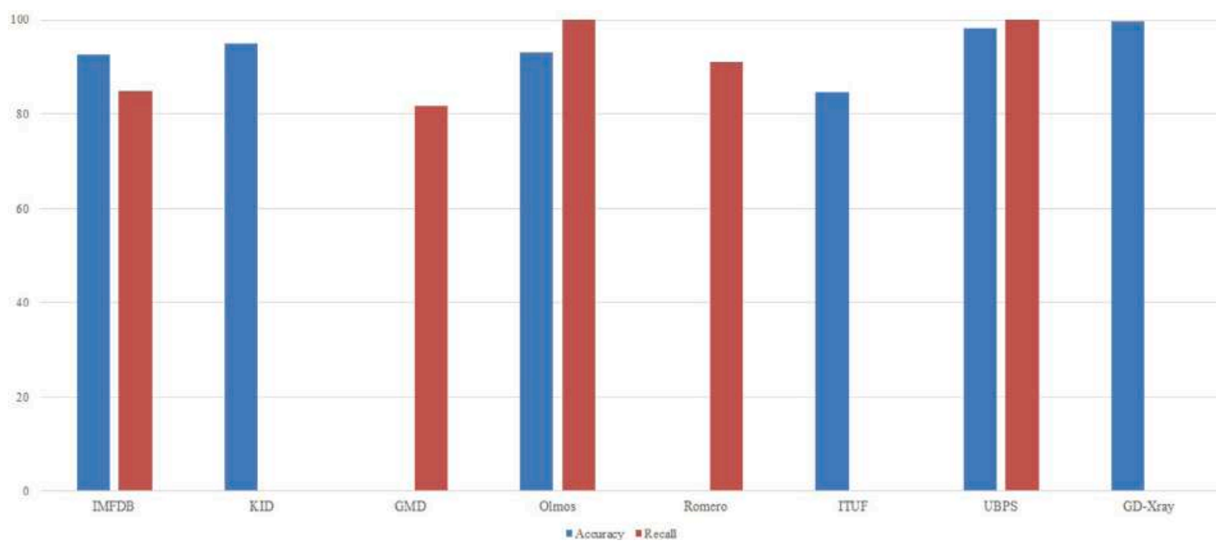

Fig. 24. Best performances on the publicly available datasets.

In Table 12, we present the obtained vital outcomes of the survey. The table answers the following questions. (a) Which methods performed better under different application scenarios? (b) What are the key issues that have been solved? (c) What are the remaining issues that need more attention from computer vision practitioners?

7.1. Desirable key system characteristics

The extensive survey also indicates certain key features should present in an automatic weapon detection system, that it can be implemented in real time.

- **Robustness:** The key challenges in real time weapon detection are multiple weapons, the similarity between all the types of weapons, occlusion of weapons. A real time weapon detection system should be robust in terms of handling all these challenges. From a Machine learning point of view, the major bottleneck is the non-availability real time dataset. Therefore, it requires special attention while designing the machine learning-based method to improve its performance. As machine learning based methods are implemented on memorized video frames or images of movie clips.
- **Real Time:** For real time implementation of a weapon detection system, the automatic weapon detection system should be accurate and fast. In real time, a weapon detection system should operate on CCTV videos. Therefore, speed is most important to reduce the probability of losing objects carrying weapons. For real time implementation, an automatic weapon detection system should able to handle the mass population. If CCTV is operated in public places where a huge number of peoples are present, then detect weapons is difficult. In real time applications, the weapon detection system required to handle this challenge.
- **Consistency:** A real time weapon detection system should be performed well consistently in any situation. Automatic weapon detection means there would be no user interaction.

7.2. Future scope:

Based on the challenges we were discussed in the previous section, we present the following future scope:

(a) A Robust Method: Table 2 shows that there are very few methods that effectively handle all the basic challenges (DI, DII, DIII) of automatic gun prediction in videos. The methods that can handle all the challenges are not fully handled all the challenges such as the detection of any weapon of different shapes, different colors. Therefore, a proposal for a robust automatic weapon detection system should be an important research challenge.

(b) Dataset Creation: From the extensive survey on the automatic detection of weapons, we can conclude that there is a lack of available datasets specific to this problem. Only one real time dataset is designed for weapon detection. Other datasets are collected from the movies, games, etc.. Therefore, a real time dataset generation for weapon detection is a significant research challenge.

(c) Shortcomings of state-of-the-art deep learning architecture is the detection of small objects like a weapon, knife, pen etc.. The State-of-the-art architecture was proposed for the detection of a person, car etc.. Therefore, there is a scope of designing deep architecture specific to the weapons.

Table 12
Key takeaways from the survey.

Best methods in different contexts		
Context	**Methods**	**Publications**
Complex Background	Fusion Based Method [39] & Detector Based Methods	From table it is clear that fusion based method [39] and Detector based method [55] outperform other methods.
High intra class Variability	Deep based classifier and detector [44] and [54], fusion based method [39] and traditional detector based [50]	These methods able to handle the high intra class variations. [50] is able handle different type of weapon shape.
Concealed Weapon detection	Fusion Based Method [39]	Quantitatively [39] perform better because other methods not represent results clearly.
Issues Addressed Successfully		
Issues	**Remarks**	
Handle High-intra variability	This issue has been solved to a great extent as evident in the results of Tables 9 to 10	
Concealed Weapon Detection	This issue has been solved to a some extent as evident in the Table 1	
Issues that Need Attention		
Issues	**Remarks**	
Detection of multiple weapon	This particular issue is addressed only in [55] after the emerging of deep based detection	
Detection of different type of gun	This is an important challenge to consider. Few works [25,49,52] detect different types of gun, since then, we do not find any progress addressing this issue. Even more, deep based detection methods also failed in handling this issue.	
Partial Occlusion	From Table & table shows that, few works addressed this issue of partial occlusion. Still these are challenges.	
Non-availability of Real time dataset	From the Table 5 and discussion of Section 6, it is evident that there is no availability of real time datasets. The datasets that were proposed in several research work are collected from the internet. Hence, there is saliency based meta major scope left behind regarding the research.	

(d) DEEP Learning Architecture for segmentation: As per our knowledge, DEEP learning architecture is not exploited for automatic segmentation of firearms, knives from the input images. It can be considered as an important research scope towards segmentation based weapon detection.

(e) Detection of type of gun: There are different types of guns. Automatic detection of guns and its related details may be beneficial towards crime scene analysis. So, detection of the type of gun is also an important area of research.

(f) Track the position of weapon: Until now, there is no work proposed for real time tracking of weapon position. Information related to the position of weapon is important to know the possibility of crime. Hence, there is a possible research gap.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work presented here is being conducted in the Computer Vision Laboratory of Department of Computer Science and Engineering of Tripura University (A central university), Tripura, Suryamaninagar-799022.

References

[1] United nations office on drugs and crime (UNODC). Global study on homicide 2013. Data: UNODC homicide statistics, 2013, 2013.
[2] A. Glowacz, M. Kmieć, A. Dziech, Visual detection of knives in security applications using active appearance models, Multimedia Tools Appl. 74 (12) (2015) 4253–4267.
[3] S.A. Velastin, B.A. Boghossian, M.A. Vicencio-Silva, A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, Transp. Res. C 14 (2) (2006) 96–113,.
[4] T. Ainsworth, Buyer beware, in: Security Oz, Vol. 19, 2002, pp. 18–26.
[5] IMFDB: Internet movie firearms database. http://www.imfdb.org/wiki/Main_Page.
[6] K.N. Choi, Characteristics of metal sensor using variable-frequency, J. Korea Inst. Electron. Commun. Sci. 9 (2) (2014) 161–166.
[7] K.N. Choi, Two-channel metal detector using two perpendicular antennas, J. Sens. 2014 (2014).
[8] M.-A. Slamani, P.K. Varshney, D.D. Ferris Jr., Survey of image processing techniques applied to the enhancement and detection of weapons in mmw data, in: Infrared and Passive Millimeter-wave Imaging Systems: Design, Analysis, Modeling, and Testing, Vol. 4719, 2002, pp. 296–305.
[9] R. Mahajan, D. Padha, Detection of concealed weapons using image processing techniques: A review, in: 2018 First International Conference on Secure Cyber Computing and Communication, ICSCCC, 2018, pp. 375–378.
[10] B.-N. Vo, B.-T. Vo, N.-T. Pham, D. Suter, Detection of concealed weapons using image processing techniques: A reviewjoint detection and estimation of multiple objects from image observations, IEEE Trans. Signal Process. 58 (10) (2010) 5129–5141.
[11] M. Villamizar, A. Garrell, A. Sanfeliu, F. Moreno-Noguer, Interactive multiple object learning with scanty human supervision, Comput. Vis. Image Underst. 149 (2016) 51–64.
[12] K. Oh, M. Lee, G. Kim, S. Kim, Detection of multiple salient objects through the integration of estimated foreground clues, Image Vis. Comput. 54 (2016) 31–44.
[13] Z. Haladova, E. Sikudov, Multiple instances detection in rgbd images, in: International Conference on Computer Vision and Graphics, 2014, pp. 246–253.
[14] G. Aragon-Camarasa, J.P. Siebert, Unsupervised clustering in hough space for recognition of multiple instances of the same object in a cluttered scene, Pattern Recognit. Lett. 31 (11) (2010) 1274–1284.
[15] A. Torralba, K.P. Murphy, W.T. Freeman, Sharing visual features for multiclass and multiview object detection, IEEE Trans. Pattern Anal. Mach. Intell. 29 (5) (2007) 854–869.
[16] M.K. Uner, L.C. Ramac, P.K. Varshney, Concealed weapon detection: An image fusion approach, in: Investigative Image Processing, Vol. 2942, 1997, pp. 123–132.
[17] I.T. Darker, P. Kuo, M.Y. Yang, A. Blechko, C. Grecos, D. Makris, J.-C. Nebel, A.G. Gale, Automation of the CCTV-mediated detection of individuals illegally carrying firearms: combining psychological and technological approaches, in: Visual Information Processing XVIII, Vol. 7341, 2009, p. 73410.
[18] R.K. Tiwari, G.K. Verma, A computer vision based framework for visual gun detection using harris interest point detector, Procedia Comput. Sci. 54 (2015) 703–712.
[19] N.J. Hussein, F. Hu, F. He, Multisensor of thermal and visual images to detect concealed weapon using harmony search image fusion approach, Pattern Recognit. Lett. 94 (2017) 219–227.

[20] R.K. Tiwari, G.K. Verma, A computer vision based framework for visual gun detection using harris interest point detector, Procedia Comput. Sci. 54 (2015) 703–712.

[21] R. Gesick, C. Saritac, C.-C. Hung, Automatic image analysis process for the detection of concealed weapons, in: Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, Vol. 7341, 2009, pp. 1–4.

[22] S.-Y. Cho, N.-P. Tin, Using infrared imaging technology for concealed weapons detection and visualization, in: TENCON 2010-2010 IEEE Region 10 Conference, 2010, pp. 228–233.

[23] P. Vijayalakshmi, M. Sumathi, Design of algorithm for detection of hidden objects from Tera hertz images, IOSR J. Comput. Eng. 13 (2) (2013) 25–32.

[24] N. Nandhini, M. Sowmya, P. Malarvizhi, Spontaneous detection of weapons in ATM banks, Int. J. Eng. Dev. Res. 5 (1) (2017) 436–449.

[25] E. Ardizzone, R. Gallea, M.L. Cascia, G. Mazzola, Combining top-down and bottom-up visual saliency for firearms localization, in: 2014 International Conference on Signal Processing and Multimedia Applications, SIGMAP, 2014, pp. 25–32.

[26] S. Bandyopadhyay, B. Datta, S. Roy, Identifications of concealed weapon in a human body, 2012, arXiv preprint arXiv:1210.5653.

[27] T. Xu, Q.M.J. Wu, Multisensor concealed weapon detection using the image fusion approach, in: 6th International Conference on Imaging for Crime Prevention and Detection, ICDP-15, 2015, pp. 1–7.

[28] N.J. Hussein, F. Hu, F. He, A.A. Ayoob, Enhancing the seismic histogram equalization of multi-fusion for infrared image of concealed weapon detection, in: 2016 International Conference on Innovative Material Science and Technology, IMST 2016, 2016.

[29] B. Khajone, V.K. Shandilya, Concealed weapon detection using image processing, Int. J. Sci. Eng. Res. 3 (2012) 1–4.

[30] Z. Liu, Z. Xue, R.S. Blum, R. Laganiere, Concealed weapon detection and visualization in a synthesized image, Pattern Anal. Appl. 8 (4) (2006) 375.

[31] M. Kowalski, M. Kastek, M. Piszczek, M. Życzkowski, M. Szustakowski, Harmless screening of humans for the detection of concealed objects, in: Safety and Security Engineering VI, Vol. 151, 2015, pp. 215–223.

[32] Z. Xue, R.S. Blum, Concealed weapon detection using color image fusion, in: Proceedings of the 6th international conference on information fusion, Vol. 1, 2003, pp. 622–627.

[33] L.C. Ramac, M.K. Uner, P.K. Varshney, M.G. Alford, D.D. Ferris, Morphological filters and wavelet-based image fusion for concealed weapons detection, in: Sensor Fusion: Architectures, Algorithms, and Applications II, Vol. 3376, 1998, pp. 110–119.

[34] A. Toet, Color image fusion for concealed weapon detection, in: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Defense and Law Enforcement II, Vol. 5071, 2003, pp. 372–379.

[35] M.K. Uner, L.C. Ramac, P.K. Varshney, M.G. Alford, Concealed weapon detection: an image fusion approach, in: Investigative Image Processing,International Society for Optics and Photonics, Vol. 2942, 1997, pp. 123–132.

[36] E.M. Upadhyay, N.K. Rana, A novel multifocus and multi exposure approach for concealed weapon detection, Int. J. Latest Technol. Eng. Manage. Appl. Sci. 1 (10) (2013) 20–25.

[37] Z. Xue, R.S. Blum, Y. Li., Fusion of visual and IR images for concealed weapon detection, in: Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. IEEE Cat. No. 02EX5997, Vol. 2, 2002, pp. 1198–1205.

[38] Z. Zhang, R.S. Blum, Region-based image fusion scheme for concealed weapon detection, in: Proceedings of the 31st annual conference on information sciences and systems, 1997, pp. 168–173.

[39] R. Olmos, S. Tabik, A. Lamas, F. Pérez-Hernández, F. Herrera, A binocular image fusion approach for minimizing false positives in handgun detection with deep learning, Inf. Fusion 49 (2019) 271–280.

[40] A.J. Lingg, B.D. Rigling, An algorithm for the detection of handguns in terahertz images, in: Terahertz Physics, Devices, and Systems III: Advanced Applications in Industry and Defense, Vol. 7311, 2009, p. 73110J.

[41] N. Megherbi, G.T. Flitton, T.P. Breckon, A classifier based approach for the detection of potential threats in CT based baggage screening, in: 2010 IEEE International Conference on Image Processing, 2010, pp. 1833–1836.

[42] R. Vajhala, R. Maddineni, P.R. Yeruva, Weapon Detection in Surveillance Camera Images (Master's thesis), Department of Applied Signal Processing, Bleking Institute of Technology, 2016.

[43] M. Grega, A. Matiolański, P. Guzik, M. Leszczuk, Automated detection of firearms and knives in a CCTV image, Sensors 16 (1) (2016) 47.

[44] D. Romero, C. Salamea, Design and proposal of a database for firearms detection, in: The International Conference on Advances in Emerging Trends and Technologies, 2019, pp. 348–360.

[45] T.H. Deepthi, R. Gaayathri, S. Shanthosh, A.S. Gebin, R.A. Nithya, Firearm recognition using convolutional neural network, in: 2018 First International Conference on Secure Cyber Computing and Communication, ICSCCC, 2019.

[46] F. Gelana, A. Yadav, Firearm detection from surveillance cameras using image processing and machine learning techniques, in: Smart Innovations in Communication and Computational Sciences, 2019, pp. 25–34.

[47] A. Goyal, S.B. Anandamurthy, P. Dash, S. Acharya, D. Bathla, D. Hicks, A. Bhan, P. Ranjan, Automatic border surveillance using machine learning in remote video surveillance systems, in: Emerging Trends in Electrical, Communications, and Information Technologies, 2020, pp. 751–760.

[48] S. Akcay, M.E. Kundegorski, C.G. Willcocks, T.P. Breckon, Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery, IEEE Trans. Inf. Forensics Secur. 13 (9) (2018) 2203–2215.

[49] J. Lai, S. Maples, Developing a Real-Time Gun Detection Classifier, World Academy of Science, Stanford University, 2017.

[50] A. Matiolański, A. Maksimova, A. Dziech, CCTV object detection with fuzzy classification and image enhancement, Multimedia Tools Appl. 75 (17) (2016) 10513–10528.

[51] X. Shen, C.R. Dietlein, E. Grossman, Z. Popovic, F.G. Meyer, Detection and segmentation of concealed objects in terahertz images, IEEE Trans. Image Process. 17 (12) (2008) 2465–2475.

[52] J. Iqbal, M.A. Munir, A. Mahmood, A.R. Ali, M. Ali, Orientation aware object detection with application to firearms, 2019, arXiv preprint arXiv:1904.10032.

[53] F. Gelana, A. Yadav, Firearm detection from surveillance cameras using image processing and machine learning techniques, in: Smart Innovations in Communication and Computational Sciences, 2019, pp. 25–34.

[54] R. Olmos, S. Tabik, F. Herrera, Automatic handgun detection alarm in videos using deep learning, Neurocomputing 275 (2018) 66–72.

[55] G.K. Verma, A. Dhillon, A handheld gun detection using Faster R-CNN deep learning, in: Proceedings of the 7th International Conference on Computer and Communication Technology, Vol. 275, 2017, pp. 84–88.

[56] G.V.S. Cardoso, P.M. Ciarelli, R.F. Vassallo, Use of deep learning for firearms detection in images, in: Anais do XV Workshop de Visão Computacional, 2019, pp. 109–114.

[57] A. Castillo, S. Tabik, F. Pérez, R. Olmos, F. Herrera, Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning, Neurocomputing 330 (2019) 151–161.

[58] S. Dubey, Building a gun detection model using deep learning, in: M. Afzal Upal (Ed.), Program Chair & Proceedings, PhD Chair of Computing & Information Science Department Mercyhurst University 501 E 38th St Erie, PA, 16546, 2019.

[59] J. Olsson, A client-server solution for detecting guns in school environment using deep learning techniques, 2019.

[60] D. Romero, C. Salamea, Convolutional models for the detection of firearms in surveillance videos, Appl. Sci. 9 (15) (2019) 2965.

[61] M.M. Fernandez-Carrobles, O. Deniz, F. Maroto, Gun and knife detection based on faster R-CNN for video surveillance, in: Iberian Conference on Pattern Recognition and Image Analysis, 2019, pp. 441–452.

[62] Y. Elmir, S.A. Laouar, L. Hamdaoui, Deep learning for automatic detection of handguns in video sequences, in: JERI, 2019.

[63] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.

[64] O. Barnich, M.V. Droogenbroeck, ViBe: A universal background subtraction algorithm for video sequences, IEEE Trans. Image Process. 20 (6) (2010) 1709–1724.

[65] M.K. el den Mohamed, A. Taha, H.H. Zayed, Automatic gun detection approach for video surveillance, Int. J. Sociotechnol. Knowl. Dev. 12 (1) (2020) 49–66.

[66] R. Mahajan, D. Padha, Detection of concealed weapons using image processing techniques: A review, in: 2018 First International Conference on Secure Cyber Computing and Communication, ICSCCC, Vol. 49, 2018, pp. 375–378.

[67] D.G. Lowe, Object recognition from local scale-invariant features, in: Computer vision, 1999, in: The Proceedings of the Seventh IEEE International Conference on, Vol. 2, 1999, pp. 1150–1157.

[68] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[69] A. Kaur, L. Kaur, Concealed weapon detection from images using SIFT and SURF, in: 2016 Online International Conference on Green Engineering and Technologies, IC-GET, 2016, pp. 1–8.

[70] H. Bay, T. Tuytelaars, L.V. Gool, Surf: Speeded up robust features, in: Computer Vision–ECCV 2006, 2006, pp. 404–417.

[71] H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, Speeded-up robust features (surf), Comput. Vis. Image Underst. 110 (3) (2008) 346–359.

[72] C. Harris, M. Stephens, A combined corner and edge detector, in: Alvey Vision Conference, Vol. 15, No. 50, 1988, pp. 10–5244.

[73] J. Malik, R. Dahiya, G. Sainarayanan, Harris operator corner detection using sliding window method, Int. J. Comput. Appl. 22 (1) (2011) 28–37.

[74] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, 2005, pp. 886–893.

[75] C.S. Won, D.K. Park, S. Park, Efficient use of MPEG-7 edge histogram descriptor, ETRI J. 24 (1) (2002) 23–30.

[76] T.P. Wallace, P.A. Wintz, An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors, Comput. Graph. Image Process. 13 (2) (1980) 99–126.

[77] G.H. Granlund, Fourier preprocessing for hand print character recognition, IEEE Trans. Comput. C-21 (2) (1972) 195–201.

[78] B.P. Adhi, R.S. Anugrah, Indonesian traditional weapons recognition using backpropagation, in: 2018 Indonesian Association for Pattern Recognition International Conference, INAPR, 2018, pp. 67–70.
[79] C.P. Papageorgiou, M. Oren, T. Poggio, A general framework for object detection, in: Computer Vision, 1998. Sixth Interna1510 tional Conference on, 1998, pp. 555–562.
[80] M.-K. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inf. Theory 8 (2) (1962) 179–187.
[81] M.R. Teague, Image analysis via the general theory of moments, J. Opt. Soc. Amer. 70 (8) (1980) 920–930.
[82] A. Khotanzad, Y.H. Hong, Rotation invariant pattern recognition using Zernike moments, in: [1988 Proceedings] 9th International Conference on Pattern Recognition, 1988, pp. 326–328.
[83] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. 6 (1973) 610–621.
[84] A.K. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, in: 1990 IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings, 1990, pp. 14–19.
[85] T. Sikora, The MPEG-7 visual standard for content description-an overview, IEEE Trans. Circuits Syst. Video Technol. 11 (6) (2001) 696–702.
[86] J.J. Koenderink, A.J.V. Doorn, Surface shape and curvature scales, Image Vis. Comput. 10 (8) (1992) 557–564.
[87] M. Lillholm, L.D. Griffin, Statistics and category systems for the shape index descriptor of local 2nd order natural image structure, Image Vis. Comput. 27 (6) (2009) 771–781.
[88] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, et al., Query by image and video content: The QBIC system, Computer 28 (9) (1995) 23–32.
[89] D. Mery, V. Riffo, U. Zscherpel, G. Mondragon, I. Lillo, I. Zuccar, H. Lobel, M. Carrasco, GDXray: The database of X-ray images for nondestructive testing, J. Nondestruct. Eval. 34 (4) (2015) 1–12.

Rajib Debnath received his B.Tech (IT) degree from BITM, Shantiniketan, West Bengal, India, in 2010 and M.Tech (CSE) degree from Tripura University (A Central University), Suryamaninagar, Tripura, India in 2012. Currently he is Pursuing Ph.D under Computer Science and Engineering department of Tripura University (A Central University). His topics of interest are related to the field of Computer Vision, Machine Learning, Image and Video Processing etc.

Dr. Mrinal Kanti Bhowmik obtained his Bachelor of Engineering in Computer Science & Engineering from Tripura Engineering College in 2004 and Masters of Technology in Computer Science & Engineering from Tripura University (A Central University), India, in 2007. In 2014, he received Ph.D. (Engineering) degree from Jadavpur University, Kolkata, India. He has successfully completed two DeitY funded project, one DBT-Twinning funded project and one SAMEER funded project as Principal Investigator. Currently, he is the Principal Investigator of the two Govt. of India projects, one DRDO funded project and one ICMR funded project. From July 2010 onward, he is serving in the Department of Computer Science & Engineering at Tripura University as an Assistant Professor. His current research interests are in the field of computer vision, medical imaging, biometrics etc. He is also a Senior Member of the IEEE (USA).