Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 54 (2015) 703 – 712

Procedia Computer Science

Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector

Rohit Kumar Tiwari* and Gyanendra K. Verma

Department of Computer Engineering, National Institute of Technology-Kurukshetra, Kurukshetra 136 119, Haryana, India

Abstract

Today's automatic visual surveillance is prime need for security and this paper presents first step in the direction of automatic visual gun detection. The objective of our paper is to develop a framework for visual gun detection for automatic surveillance. The proposed framework exploits the color based segmentation to eliminate unrelated object from an image using k-mean clustering algorithm. Harris interest point detector and Fast Retina Keypoint (FREAK) is used to locate the object (gun) in the segmented images. Our framework is robust enough in terms of scale, rotation, affine and occlusion. We have implemented and tested the system over sample images of gun, collected by us. We got promising performance of our system to detect a gun. Further, our system performs very well under different appearance of images. Thus our system is rotation, scale and shape invariant.

Peer-review under responsibility of organizing committee of the Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

Keywords: FREAK; Gun detection; Harris; Interest point detector; Video surveillance.

1. Introduction

Today, Closed Circuit Television (CCTV) is used as monitoring and surveillance tool for fighting crimes. The aim of CCTV is to reduce the crime and social offence by monitoring the scene under surveillance. Its use varies from user to user. For example, CCTV is being used in street surveillance for monitoring various activities like finding missing person, indentifying anti social behavior, drug misuse etc[1]. It is being also used for capturing evidence of crime and presenting evidences to courts for prosecution. A CCTV involves the use of an unmanned and remotely mounted camera and an operator. A CCTV camera captures the video and transmits it to the television screen of base station that is monitored by the operator to detect suspicious activity or to capture evidences. But, the detection of suspicious activity is proportional to the operator attention on each video feed of screen. So, it is not possible for a CCTV operator to effectively monitor each activity of video feeds all the time with complete attention due to low operator to screen ratio, concurrent run of multiple video feeds at same screen[2] and environmental condition of operational room. There is a possibility for them to miss the detection of some abnormal activity. According to Velastin *et al.*[16], A CCTV operator suffers from video blindness after 20 to 40 minutes of active monitoring due to which he is not able to recognize objects in video feeds. According to a study published in Security Oz Magazine[17] "an operator will often miss up to 45% of screen activity after 12 minutes of continuous monitoring and this miss rate increase up to 95% after

*Corresponding author. Tel. +918950909698.
E-mail address: rohitkushinagar@gmail.com

Peer-review under responsibility of organizing committee of the Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)
doi:10.1016/j.procs.2015.06.083

22 minutes". So to make the CCTV monitoring and surveillance efficient, there is a need to automate the detection of suspicious activity in video material which in turn reduces the operator overload. The automated system will raise an alarm if any abnormal activity takes place under CCTV surveillance, due to which the operator will allocate his attention on video feed and will take appropriate action.

In the recent past there have been some accidents which have brought the demand of an automated visual weapon detection system, in picture. Specially, the detection of visual gun is of particular interest because guns are broadly available and are used as a weapon. According to a survey of US Department of Justice, on average about 75% of the homicide involves use of hand gun from 1994 to 2011[18]. According to another survey of US Department of Justice, hand gun is the main weapon used for different crime like robbery, rape, etc. in comparison to other weapon like knife, blunt object etc.[19]. Some of the countries like Ireland carrying gun at public place are either prohibited or restricted[3]. So the detection of gun is of high importance for safety concern. Our work focuses on automatic detection of gun in the images which can be further used in tracking of gun by use of optical flow. As the electronic devices are becoming cheaper day by day, it seems to be a good idea to have an automated Gun Detection System (GDS) for surveillance of public using CCTV. The automatic GDS raise an alarm signal if it detects gun in a frame extracted from video feed, after that CCTV operator can allocate his attention on the video scene and either accept or discard that detection. So, GDS will provide great assistance to CCTV operator in long working hour as well as monitoring multiple video feeds concurrently on same screen. The automated gun detection system can be used in baggage scanning using X-ray to detect guns in bags at airport or railway stations.

Our proposed scheme of automatic detection of gun uses hybrid approach of color based segmentation and interest point detector. We have used combination of Harris interest point detectors and FREAK descriptor for detecting interest point and extracting features which is used for matching with gun descriptor. Color based segmentation is performed using k-means clustering algorithm to eliminate unrelated color or object present in the image. After that morphological processing is applied on each object to bridge small gaps and for boundary extraction. Then interest point features of object boundary is extracted which is used for matching with stored descriptor to find out similarity with gun. If similarity score is greater than 50%, then system will raise an alarm signal.

There are two contributions of our visual gun detection framework-first one is that there has been no other research work on visual gun detection (to the best knowledge of authors) and second it can detect multiple objects in the images unlike any other point detector such as SIFT[4], SURF[5], BRISK[6] etc. Point detectors are able to detect only one object in the image even though there is multiple object of same kind. This paper is organized as follows, Section 2 discuss about various challenges of visual gun detection followed by weapon detection literature in Section 3. Section 4 presents our gun detection approach followed by experimental setup and experimental result and discussion in Section 5 and Section 6 respectively and then conclusion in Section 7.

2. Challenges

Guns are widely available in different color like black, silver etc. in the world which makes its detection in images a challenging task. Its shape is another issue because one model of gun varies from another model by some amount due to variation in body parts like hammer, grip safety, back sight, front sight etc. and different surface characteristics. Two different shapes of gun are shown in Fig. 1(a) and Fig. 1(b). Some other issues are scale, rotation and view variation of gun in images. Scale variation arises due to change in distance of gun from the CCTV when it captures the video. Rotation and view variation occurs due to deviation in orientation of gun and its plane respectively. However, there are many techniques to deal with scale and rotation variations but view variation is most challenging problem in field of object detection even today. Figure 1(c) and Fig. 1(e) show the illustration of scale, rotation and view variation. Another challenge that make the problem more complex are partial or full occlusion of gun, deformation, loss of information due to transformation from 3-D world into 2-D world, illumination change, shadow, presence of noises in image, and real-time processing requirement[7]. Some of these are shown in Fig. 1. Partial or full occlusions occur because guns are mostly carried in either hand or in holster. Use of equipments to target long distance object causes deformation of gun. Illumination change and shadow arises due to change in lighting condition in operational environment. Noises in image occur during acquisition and transmission of image. There are various types of noises like salt and paper noise, Gaussian noise, Raleigh noise etc that are introduced during different conditions of acquisition and transmission

Fig. 1. Illustration of different issues of Gun Detection (a) Partial occlusion; (b) Different shape; (c) View change; (d) Illumination change; (e) Scale and rotation variation; (f) Noisy image.

of image. Real-time processing requirement includes low space and response time requirement. As the GDS system comprises of real-time application so it must have low response time.

3. Related Work

The use of CCTV has become general for monitoring various activities. Typical applications include traffic monitoring like statistics of traffic flow, traffic rule violations, people counting etc and intruder, suspicious activity detection in automated surveillance[7]. All of the above applications utilized the information carried in consecutive frames of video. For gun detection we have focused the detection of gun in single frame then using optical flow method to track it in later frame and perform gun detection again to reduce the chances of false detection.

Although we have already stated that no work has been done on visual gun detection according to our knowledge, but there is a lot of research in field of concealed weapon detection and few in field of visual knife detection. Concealed weapon detection (CWD) involves detection of weapons concealed underneath a person's clothing. It only detects that person carrying any object under his clothes or not but it did not decide whether the object belong to weapon or not and which type of weapon carried out by the person. Most of the methods of it are based on some imaging techniques like infra red imaging, millimeter wave imaging etc. Here we will discuss some of them.

D. M. Sheen *et al.*[8] proposed a method of CWD for airports and other secure location based on three dimensional millimeter (mm) wave imaging technique. Millimeter wave imaging is derived from the microwave holography techniques that uses phase and amplitude information recorded over two dimensional apertures to reconstruct a focused image of target object. Millimeter waves are nonionizing and easily penetrate the common clothing material and reflect from the human body part or weapons. Z. Xue *et al.*[9] proposed a method of CWD based on fusion of infra red (IR) imaging and color visual image. He uses the fusion of infra red image and visual image method to maintain the natural color of original image. R. Blum *et al.*[10] developed a method of CWD based on fusion of IR or mm wave image and visual image using multi resolution mosaic technique where concealed weapon is first detected by fuzzy k-means clustering method from IR or mm wave image. E. M. Upadhyay *et al.*[11] proposed a method of CWD using image fusion. They used homomorphic filter, entropy of blocks and blending approach of fusion to generate multi exposure-multi modal image from a set of visual and IR image with multiple exposures.

A. Glowacz *et al.*[12] proposed a method for visual knives detection based on active appearance model and Harris corner detector. The limitation of above method is based on the accuracy of Harris corner detector because their approach first finds the tip of knives using Harris corner detector and uses that for active appearance modeling of knife.

4. Methodology

In this section we explain our approach of visual gun detection. Figure 2 shows the block diagram of our proposed solution. Our approach of visual gun detection is based on color based segmentation using k-means clustering algorithm and interest point detector. Combination of Harris interest point detector and FREAK descriptor is used to recognize whether a given blob is gun or not. The algorithm of visual gun detection is given below. Our detection approach consists of several steps which will be discussed in following subsection.

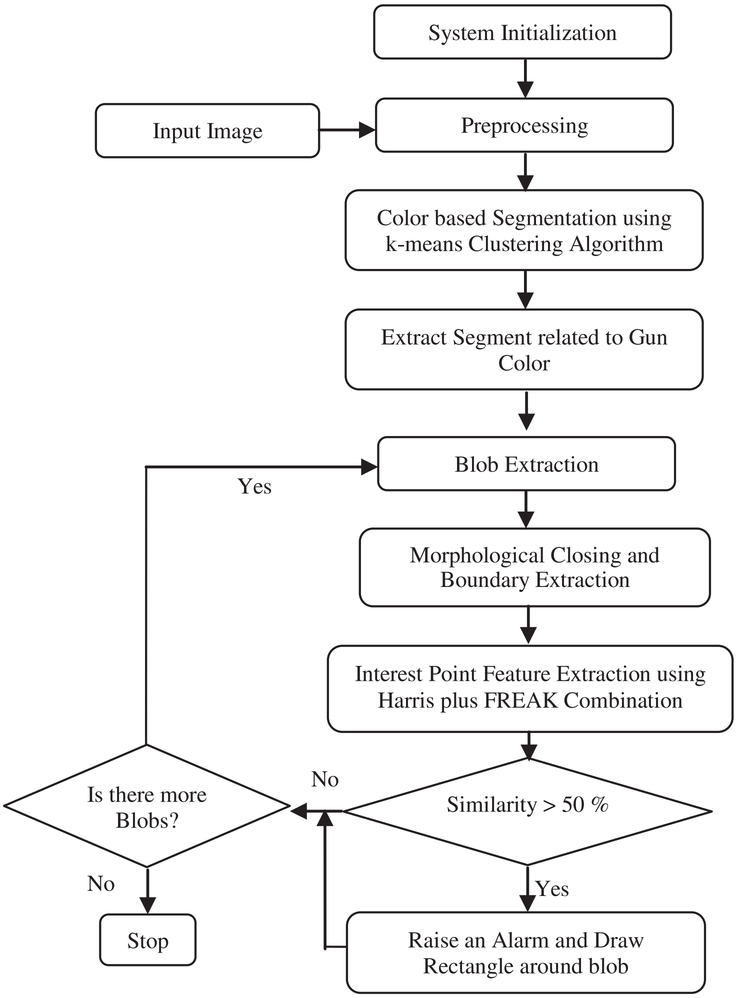

Fig. 2. Flowchart of proposed method.

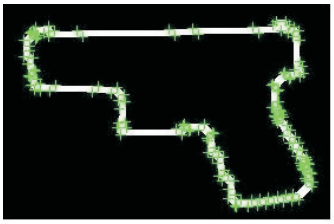

Fig. 3. Harris plus FREAK descriptor of gun.

4.1 System initialization

System initialization loads the stored interest point descriptor of the gun that is used to measure the similarity score with extracted interest point features of blobs. The combination Harris interest point detector plus FREAK is used for finding similarity. The FREAK descriptor of each Harris interest point of gun is shown in Fig. 3.

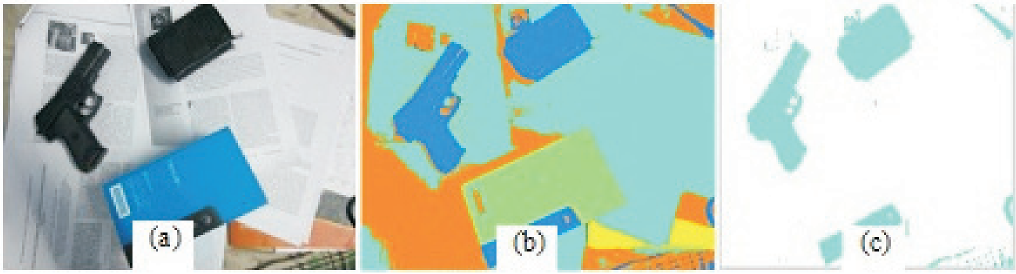

Fig. 4. (a) Referenced image; (b) Segmented image; (c) Extracted gun color part from referenced image.

4.2 Preprocessing

Preprocessing steps involve removal of different noises from the image which arises during the acquisition or transmission of image. We have used median filter to remove noise in the frame extracted from the video. As the images are acquired through camera of high resolution so image resizing is performed to make it appropriate for processing of system. In our developed system we used image with resolution of 400×300 pixels after resizing.

4.3 Color based segmentation

Color based segmentation is performed to extract the color related to gun, i.e. black if the gun is of black color. It eliminates color or objects which do not resemble gun. Color based segmentation using k-means clustering[13] is performed on image obtained from preprocessing to segment the image into different parts based on color. k- means clustering algorithm takes image and 10 number of cluster as input and then quickly approximates 10 or less clusters of image color such that the sum of the squared distances between each point and its closest cluster center is minimized. Figure 4(a) and (b) shows the input image and segmented image results after color based segmentation respectively. After performing segmentation on the input image system takes a cluster of those colors which are similar to that of gun. Figure 4(c) shows the cluster which resembles the gun color where green colored part shows the black colored part of the original image.

4.4 Blob extraction

This step extracts a blob from the segmented image that is of area greater than 1000 pixels. Here we are extracting blob of area greater than 1000 because many blobs of small area arise during segmentation due to noises. A blob is a connected component of pixel where some properties are constant or varies lightly.

4.5 Morphological closing and boundary extraction

Morphological closing is performed to eliminate the presence of small gaps in the blob. Then boundary of blob is extracted. The boundary of object is used for matching process because inner texture varies from object to object of same kind, but their outline remains constant or varies lightly. The concept of using boundary for matching is taken from the human visual system as it can recognize any object from its boundary structure. Morphological closing is achieved by erosion of image with structuring element followed by dilation with structuring element (eqn. 1). Boundary is obtained by subtracting the result of erosion of image from the original image (eqn. 2).

$$I_c = (I \oplus SE)\,\theta SE \tag{1}$$

$$B = I - (I\theta SE) \tag{2}$$

where I is the original image, I_c is the output of closing, SE is the structuring element, B represents the boundary of image.

4.6 Interest point feature extraction

We have used Harris[14] plus FREAK[15] combination of creating descriptor for each blob. This method of feature extraction involves two steps- identification of interest point by Harris corner detector and creation of descriptor of each interest point using FREAK feature extractor.

We have used Harris detector because it is invariant to geometric transformation as well as resistance to illumination change and noise up to some extent. It is based on the principal that there is more noticeable change around corner than any other part of image when a window is shifted a little in all direction. It detects the corner point based on the difference on the pixel intensities caused by window shift of (u, v) which is measured using auto correlation function given as:

$$C(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2 \tag{3}$$

where $w(x, y)$ is the Gaussian window function at centered at (x, y), $I(x, y)$ is the image intensity at (x, y) and $I(x + u, y + v)$ is the image intensity at window shift of (u, v) from (x, y). Equation 4 is approximated by Taylor Series which is given below:

$$C(u, v) = [u, v]\; M \begin{bmatrix} u \\ v \end{bmatrix} \tag{4}$$

where M is the matrix of gradients in all direction in specified neighborhood of point (x, y). The Eigen values of matrix M decided whether a point is corner or not. If the both Eigen values are high then that point is classified as corner.

FREAK is a keypoint descriptor. It is inspired by the retina of human visual system. A retinal sampling pattern is used for the computation of cascade of binary strings by efficiently comparing image intensities. FREAK is very fast as well as accurate than other keypoint descriptors. It is motivated by the working of BRIEF and BRISK keypoint descriptors. The binary representation is used in order to perform dimensionality reduction. FREAK employs smoothing of the input image with Gaussian kernel for noise suppression. A binary descriptor F is constructed by thresholding the difference between pairs of receptive fields with their corresponding Gaussian kernel. F is a binary string formed by a sequence of one-bit Difference of Gaussians (DoG). An Oriented FAST and Rotated BRIEF (ORB) like algorithm are used to select the useful pairs in the receptive fields.

4.7 Matching

Matching has been performed to calculate the similarity score between the stored descriptor of gun and blob. Interest point features of boundary of object are being used for matching because inner texture of object body varies from object to object of same kinds which will results into different features for same interest point of different images whereas the outline of object remains constant or varies lightly. Similarity score of matched boundary features will be used to find out up to what extent the shape of object is similar to shape of gun. Nearest neighbor ratio algorithm[4] is used to calculate matching between two descriptors using sum of square difference (SSD) as metric:

$$\mathrm{SSD} = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{5}$$

where x and y are two feature vector. Two feature vectors are said to be matched if the ratio of SSD of first to second nearest neighbor is less than 0.6 otherwise matching is ambiguous. Matching results of each blob and gun descriptor is shown in Fig. 5. A blob is categorized as gun if half of the interest point features of gun are matching with the blob. Here, the matching threshold is 50% because partial occlusion as well as light variation in shape of gun is also considered.

After applying all of the above procedures, we have the bounding box of those blobs which are categorized as gun. A bounding box is a smallest rectangle which covers a blob. In input image, a rectangle is inserted for each bounding box whose result is shown in Fig. 5.

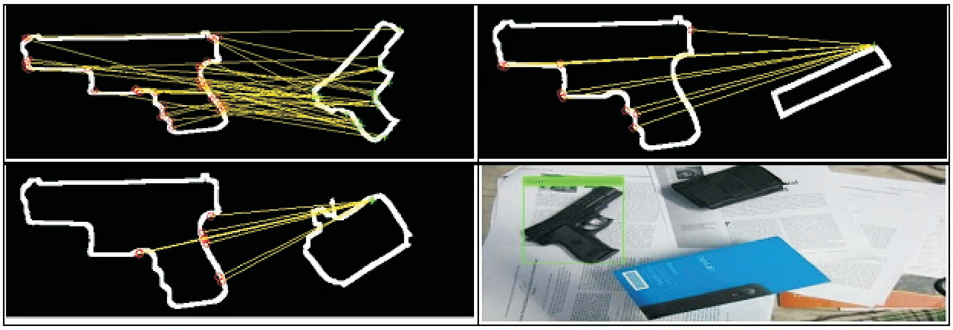

Fig. 5. Matching result of each blob of image and final gun detection result.

5. Experimental Setup

For implementation of our system we used MatLab 2013a on computer with Intel Core i3 processor and 3GB RAM. As there is no standard dataset of gun images so we created our own dataset of 65 positive images in which gun is present and 24 negative images in which gun is absent. Our dataset is prepared in such a way that it consists of image of different type of gun with different scale, rotation and orientation. In some images, some of gun is occluded by either hand or some other object and consist of multiple guns. In some images, some other objects are also present besides gun with different background.

Matching ratio and threshold are two parameters which we considered in our experiment. Matching ratio is the ratio of SSD of first and second best match of a feature. The value of matching ratio considered in our experiment is 0.83. Figure 6 shows the graph between matching ratio and correctly matched features which represent the number of correctly matched features on different values of matching ratio. Graph is drawn on basis of matching of interest point descriptor of two gun one with 28 features and other with 22 features. We find out that at value 0.83 of matching ratio, maximum numbers of features are matching correctly. If matching ratio is greater than 0.83 means one features of gun are closely matching with the two features of other object which means the matching is ambiguous because one feature cannot be similar to two features of another object. If matching ratio is less than 0.83 means first and second best match are quite different and matching is accurate. Matching threshold is the percentage of interest point features of gun that are matching correctly with the other object features. Here, we have taken matching threshold as 50% because guns are of different shape which varies lightly and another reason is of partial occlusion of gun.

The assumptions which we have considered in our approach are uniform gun color and orientation. Gun should be of uniform color because we are using color based segmentation which segments the image based on the color. If a single gun is made up of different color, then segmentation part of our algorithm will segment the gun into different clusters based on their color where each cluster will consist of only some part of gun. If there are many guns of different uniform color in a single image, in that case there is a need to run algorithm parallel on each cluster of color related to gun to detect each gun of different color in image. There should not be too much variation in orientation of gun with respect to CCTV because it will result in different view of gun as well as different visual shape.

To evaluate our system performance we have used three metric namely true positive rate, false positive rate and accuracy.

5.1 True positive rate

It is the measure of percentage of actual positive images which are correctly detected by the system. In machine learning it is also termed as sensitivity or recall.

$$T_{\text{pos}} = \frac{I_{PD}}{I_{PD} + I_{PU}} \tag{6}$$

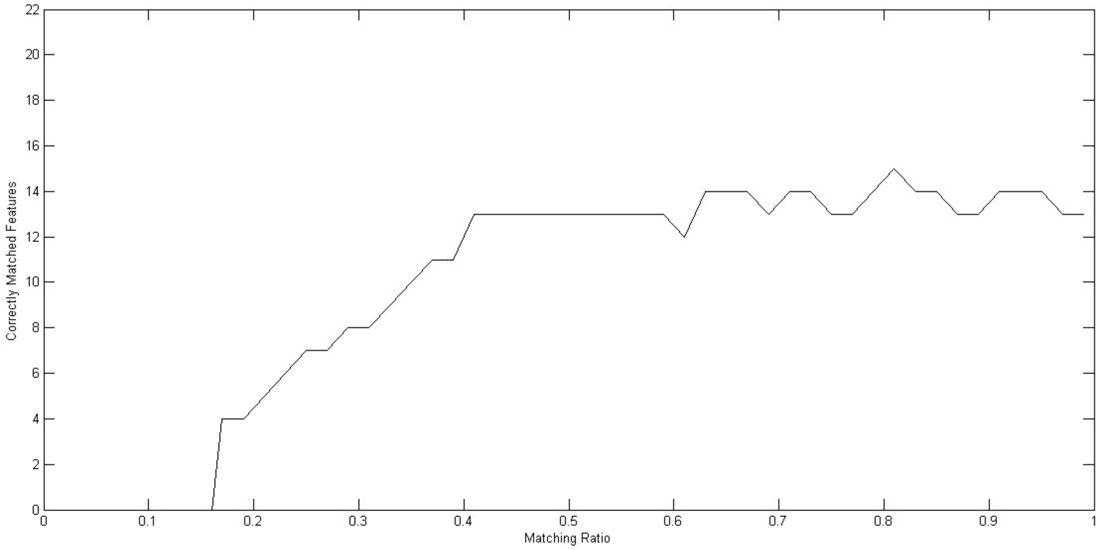

Fig. 6. Graph between max ratio threshold and correctly matched features.

5.2 False positive rate

It is the measure of percentage of negative images which are wrongly detected by the system. In machine learning it is also termed as specificity.

$$F_{\text{pos}} = \frac{I_{ND}}{I_{ND} + I_{NU}} \tag{7}$$

5.3 Accuracy

It is the measure of proportion of total number of images that were correctly detected by the system.

$$\text{Accuracy} = \frac{I_{PD} + I_{NU}}{I_{PD} + I_{PU} + I_{NU} + I_{ND}} \tag{8}$$

where I_{PD} is the number of positive images in which gun is detected, I_{PU} is the number of positive images in which gun is not detected, I_{ND} is the number of negative images in which system wrongly detects gun and I_{NU} is the number of negative images in which system does not detect gun.

6. Results and Discussion

In this section we will discuss the results as well characteristics of visual gun detection system. We have evaluated our system performance in different sessions with varying condition like gun with different background, occlusion etc. Table 1 shows the description of six different sessions in which performance of visual gun detection system is carried out. We have found promising result in different sessions.

The performance of visual gun detection system is shown in Table 2. We found out that our system works well in all situations except illumination change. It works with 91.66% true positive rate in case of images having gun with different background. The reason for working well in all other case except illumination change is due to accuracy of Harris interest point detector. Our system is not able to perform well in illumination change because our color based segmentation algorithm is not able to segment the image in accurate way due to which when we are extracting gun color from the input image we are getting only some part of the gun which affects the performance of our system. Figure 7 shows the result of gun detection in sample images in different sessions.

Table 1. Session description.

Session	Description
Session 1	Gun images with different background
Session 2	Gun images with different degree of illumination
Session 3	Gun images with interclass variation
Session 4	Gun images with different degree of occlusion
Session 5	Gun images with scale and rotation variation
Session 6	Gun images with multiple gun

Table 2. Performance under Harris plus FREAK descriptor matching.

Session	Number of positive images	Number of images correctly classified	True positive rate
Session 1	12	11	91.66%
Session 2	9	7	77.77%
Session 3	11	9	81.81%
Session 4	17	14	82.35%
Session 5	10	8	80%
Session 6	6	5	83.33%

Fig. 7. Detection results of sample images in (a) Session 1; (b) Session 2; (c) Session 3; (d) Session 4; (e) Session 5; (f) Session 6.

Table 3. Overall performance of visual gun detection system.

Number of images in positive dataset	Number of correctly detected positive images	True positive rate	Number of images in negative dataset	Number of wrongly detected negative images	False positive rate	Accuracy
65	54	83.07%	24	3	8.33%	84.26%

After evaluation our system in different sessions we evaluated our system against false positive rate using 24 negative images. We found out our system is performing well in case of negative images too. After testing our system against different sessions as well as negative images we evaluated our system in terms of accuracy. The overall performance of our system against three methods is shown in Table 3. The overall accuracy of our system is 84.26%.

7. Conclusion

In this paper, we have proposed a method for detection visual guns in images using color based segmentation and Harris interest point detector. Color based segmentation is used to eliminate unrelated color or objects which are not of interest. Then FREAK features of each Harris interest points are used to find out similarity of each segmented object with the gun descriptor. If more than 50% of features of gun descriptor are matching with the FREAK features of object then that object is labeled as gun. The novelty of our approach is that it is robust to partial occlusion, scale, rotation, affine variation and can detect the presence of multiple guns in an image. However, we believe that there are many possible ways to improve the performance of our system like making it robust to illumination change and minimizing its real-time processing requirement like time and space complexity.

References

[1] A. Maalouf, M. C. Larabi and D. Nicholson, Offline Quality Monitoring for Legal Evidence Images in Video-surveillance Applications, *Multimedia Tools and Applications*, vol. 73, no. 1, pp. 189–218, (2014).
[2] M. Kmiec, A. Glowacz and A. Dziech, Towards Robust Visual Knife Detection in Images: Active Appearance Models Initialised with Shape-specific Interest Points, In *Proceedings of 5th International Conference on Multimedia Communications, Services and Security* pp. 148–158, (2012).
[3] Global Impact of Gun Violence, Online. http://www.gunpolicy.org/firearms/region/.
[4] D. G. Lowe, Distinctive Image Features from Scale-invariant Keypoints, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, (2004).
[5] H. Bay, T. Tuytelaars and L. V. Gool, SURF: Speeded Up Robust Features, In *Proc. ECCV*, pp. 7–13, (2006).
[6] S. Leutenegger, M. Chli and R. Y. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, In *Proceedings of International Conference on Computer Vision*, pp. 2548–2555, (2011).
[7] A. Yilmaz, O. Javed and M. Shah, Object Tracking: A Survey, *ACM Computing Surveys*, vol. 38, no. 4, (2006).
[8] D. M. Sheen, D. L. McMakin and T. E. Hall, Three-dimensional Millimeter-wave Imaging for Concealed Weapon Detection, *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, (2001).
[9] Z. Xue, R. S. Blum and Y. Li, Fusion of Visual and IR Images for Concealed Weapon Detection, In *Proceedings of the Fifth International Conference on Information Fusion*, pp. 1198–1205, (2002).
[10] R. Blum, Z. Xue, Z. Liu and D. S. Forsyth, Multi Sensor Concealed Weapon Detection by using a Multi Resolution Mosaic Approach, In *Proceedings of the Vehicular Technology Conference*, pp. 4597–4601, (2004).
[11] E. M. Upadhyaya and M. K. Rana, Exposure Fusion for Concealed Weapon Detection, In *Proceedings of 2nd International Conference on Devices, Circuits and Systems*, pp. 1–6, (2014).
[12] A. Glowacz, M. Kmiec and A. Dziech, Visual Detection of Knives in Security Applications using Active Appearance Model, *Multimedia Tools Applications*, (2013).
[13] M. N. Wu, C. C. Lin and C. C. Chang, Brain Tumor Detection using Color-based k-means Clustering Segmentation, In *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 245–250, (2007).
[14] C. Harris and M. Stephenes, A Combined Corner and Edge Detector, In *Proceedings of 4th Alvey Vision Conference*, pp. 147–151, (1988).
[15] A. Alahi, R. Ortiz and P. Vandergheynst, FREAK: Fast Retina Keypoint, In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 510–517, (2012).
[16] S. A. Velastin, B. A. Boghossian and M. A. Vicencio-Silva, A Motion based Image Processing System for Detecting Potentially Dangerous Situations in Underground Railway Stations, In *Transportation Research Part C*, Elsevier, pp. 96–113, (2006).
[17] T. Ainsworth, Buyer Beware, In *Security Oz*, vol. 19, pp. 18–26, (2002).
[18] M. Planty and J. L. Truman, Firearm Violence, 1993–2011, Washington: U.S. Department of Justice, (2013).
[19] C. A. Perkins, Weapon Use and Violent Crime, 1993–2001. Washington: U.S. Department of Justice, (2003).