Available online at: https://ijact.in

Date of Submission	16/07/2019
Date of Acceptance	04/09/2019
Date of Publication	03/10/2019
Page numbers	3415-3421(7 Pages)

Cite This Paper: Fadhil K. Al-Sheikh, Israa H. Ali. Performance comparison for local feature extraction algorithms: surf, sift and orb to detect concealed weapons in x-ray images, 8(9), COMPUSOFT, An International Journal of Advanced Computer Technology. PP. 3415-3421.

ISSN:2320-0790

COMPUSOFT
An International Journal of Advanced Computer Technology

PERFORMANCE COMPARISON FOR LOCAL FEATURE EXTRACTION ALGORITHMS: SURF, SIFT AND ORB TO DETECT CONCEALED WEAPONS IN X-RAY IMAGES

Fadhil K. Al-Sheikh[1], Prof. Dr Israa Hadi Ali[2]

[1] Department of Software, University of Babylon, Iraq
[2] Department of Software, University of Babylon, Iraq

Abstract: The process of detecting hidden weapons is an important process right now due to the increase in terrorist operations, so the process of building an automatic weapons detection system is an important process to reduce errors resulting from manual detection. In the proposed work, the pre-processing was given high importance because the x-ray images contain noise and low resolution, therefore image smoothing has been used to reduce the noise where histogram equalization has been used for image enhancement and increase of contrast. The local algorithms: SIFT, SURF and ORB have been used to detect and describe the features from the region of interest, then KNN algorithm has been used to match and index the similarity between the query image and the extracted features from the data set. KNN and Random Sample used a consensus on the three methods to see which local algorithm performs best. RANSAC has been used to reject false matches that may be taken as correct matches. The performance of the SIFT algorithm with the KNN outweighed both of the algorithms in spite of the fact that it was slow. SURF and the ORB algorithms as a position in the result where SURF was the fastest one with high performance and showing its dominance in illumination changes and rotation.

Keywords: SURF, SIFT, ORB (Oriented FAST and Rotated BRIEF Detectors and Descriptors), Point Detector, KNN, Random Sample Consensus (RANSAC), Convolutional Neural Network (CNN).

I. INTRODUCTION

In view of the increase of terrorist organizations and their operations in sensitive and public areas, there is a need for the use of screening systems at airports, important checkpoints and buildings that have a critical priority to protect them. X-ray inspection systems are considered as important systems for inspecting luggage at airports, there are two types of these systems, which are represented by the systems of X-ray examination of the energy of wattage and also the systems of high-energy[17].

X-ray imaging is one of the most important technologies in many fields and has a wide range of detection of hidden weapons and sensitive objects at checkpoints, whether in checked luggage at airports or in sensitive buildings. The need to use inspection systems in these places is to increase the safety of individuals in the community, Airports or public places. However, inspection is a complex process because the threat can be in any position or place and can also be covered by other objects or in a rotated situation. Hence, the need to use algorithms that detect potential threats increases. The process of constructing auto mechanical inspection systems is important, especially the analysis of X-ray imaging because it is considered a challenge in the computer vision. This system must be highly accurate in detecting threats and taking little processing time and reducing the false positive results because this goal is desired from these systems [14].

In the proposed method of automatic detection of possible threat, three local algorithms have been used to extract the key-points from the region of interest such as SIFT, SURF, and ORB and the description for each key point detected. Local features algorithm has been applied on collected images of luggage as these images include different forms of weapons read by a feature extraction algorithm, as it has been mentioned earlier SIFT, SURF, and ORB were applied to describe and detect key-points. A query image is inserted into the inspection system as input to detect the weapon. After feature extraction algorithm detects and describes the region of interest, a Classification algorithm is used to match them with weapon images directory where the same steps are done for the source input image of luggage. If the matched points are greater than the threshold value then the weapon is detected.

II. Literature Review

The critical need to increase security and protection in public places that includes the discovery of hidden weapons in the X-ray images led to the emergence of a lot of researches and various methods, including [1] who proposed a comparison of two methods, the bag of words and CNN, in the first method the SIFT algorithm was used for the generating of the BOF then SVM was used as a classifier. The CNN outperformed BOW, but in CNN the output layer was not considered, only that the layer has extracted feature are considered, CNN also used SVM as a classifier. Mery in [2] proposed an adaptive ISM that consists of 3 steps 1) image acquisition 2) codebook generation 3) occurrence, in training step capture the object from multi-able view, in this paper the adaptive was in 2 steps the first one included only the useful key-point that were extracted from the test image are consider it, the second only the occurrence that has powerful similarity score is considered to be a valid detection.[3] presented an approach that can detect concealed weapons by using Color based segmentation to remove unrelated objects that are not are not of interest. Harris detector was used to identify the key points. Then FREAK features were used to find the similarity between the segments and weapons descriptors. If the similarity scores aremore than 50% then the weapon is detected. However, this approach is not robust against changes in illumination [4] proposed to distinguish 3D objects in a microwave radar image which had three steps, SIFT, matching filter and histogram thresholding this approach segments microwave image to 8 segments the size of the segments is very important, the correlation coefficient. This approach was not robust against a scale or detecting a partial object such as a key.[5] proposed a method to detect weapons that have a less false rate by using shape context descriptor and a Zernike moment to extract features for the threat region that are obtained by connected component analysis technique. The extracted features are used as input to ANFIS to give a decision about the presence of a weapon or otherwise. In paper [6] proposed a method to retrieve an image from a large image dataset and to achieve this, it requires more accurate algorithm retrieval. The features are extracted using HOG descriptor for the points of interest detected by Harris detector. The PCA was used to reduce the HOG dimension and color information was added to improve System Resolution.

III. The Proposed System

The inspection system depicted in Fig 1. represents a distinctive database of multidimensional vectors is formed. Detecting concealed weapons and identifying them by extracting distinctive vectors from images into the raw database, which forms a distinct database. The pre-processing is an important stage because the x-ray image maybe containing noise or be of low resolution. The feature is extracted from the query image

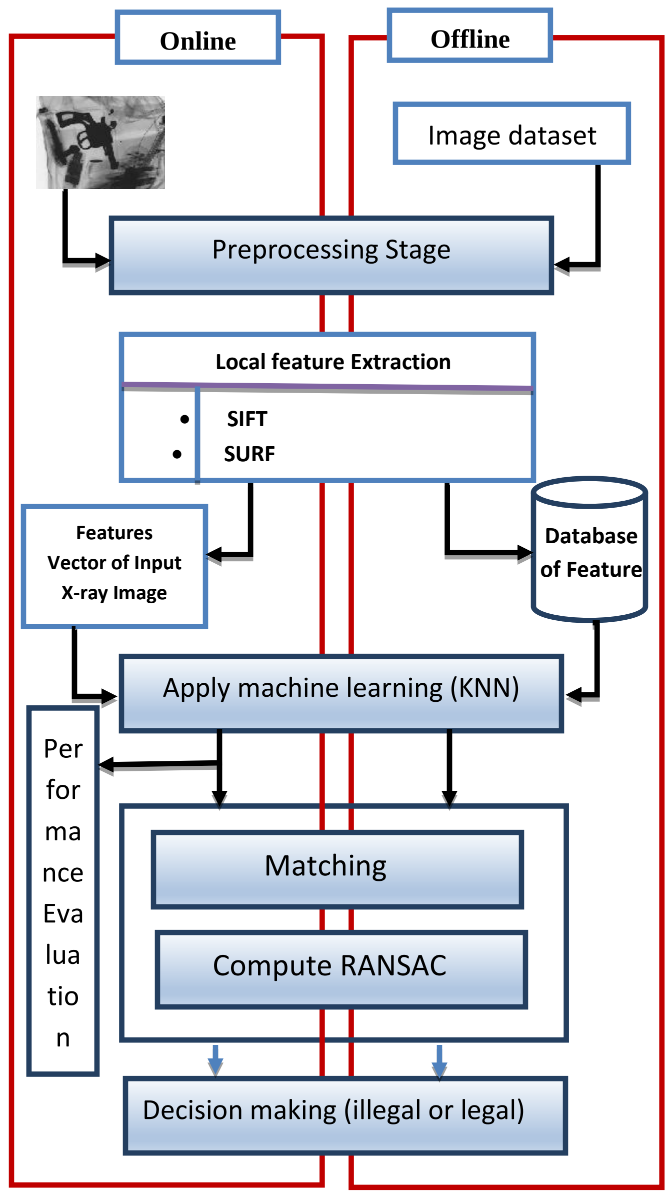


Figure 1: Explanation of the proposed system

A classification algorithm has been applied to find the similarity between image queries vectors and vector database vector, after that KNN algorithm finds similarity and matched features between the query image vectors and the extracted features of the databases that saved in the form vectors of databases. Notice that the dataset in [11] used as a test image

A. *pre-processing*

The pre-processing is an important process, especially with X-ray images, because they contain a lot of noise and can be of low quality. The pre-processing is done for each of the dataset and the query image by image smoothing (blurring) to delete the noise and using the histogram equalization to increase the contrast and enhance the quality of the image after this process comes the feature extraction.

B. *Feature Extraction*

In this work, many local algorithms were used to extract features and the results were compared between these algorithms. The best algorithm was identified for extracting features that were used on X-ray images.

C. *SURF*

Speed-up Robust Feature, it has been developed by Herbert Bay et.al [8]. It is a local feature detector and descriptor used for registration, reconstruction and object recognition. SURF was used to increase the performance. The objective of the integrated image was the speed of calculating any rectangular region. There are two critical steps to calculate the SURF algorithm as follows:

Step 1: Feature Detection.

The SURF algorithm depends on the Hessian matrix to define the descriptors as it is called Hessian blob. The advantage of using Hessian Matrix is to locate the spot and scale of the descriptor.
The definition of Hessian matrix for a given point $\boldsymbol{x} = (x, y)$ is as H(x, σ) in an image by applying equation:

$$H(x,y,\sigma) = \begin{bmatrix} I_{xx}(x,y,\sigma) & I_{xy}(x,y,\sigma) \\ I_{xy}(x,y,\sigma) & I_{yy}(x,y,\sigma) \end{bmatrix} \quad (1)$$

WhereI_{xy}, I_{xx} and I_{yy} are the second order derivatives of the Gaussian function with box filters. using integral image box filters and convolutions Image can compute. The Hessian matrix is extracted and determined by the equation:

$$an\ the\ the\ the\ \det(H) = I_{xx}I_{yy} - I_{xy}^2$$

To localize interest points, it is possible to apply the determinant of the Hessian matrix on all points, the point has local maxima, it can be called interest point and inserted in scale, octaves and image space.

Step 2:Feature Descriptor

Each interest point should have a unique description that does not depend on the features scale and a descriptor is given for each point, it is used to get an exact match in retrieval, matching, and recognition. SURF descriptor is extracted in two steps: in the beginning by using the Haar-wavelet response to appoint the orientation for the interest point to achieve invariance to rotation for circle region around the interest point in both direction (x,y),.that the Haar-wavelet responses are measured and weighted for a Gaussian with σ = 2.5s focused at the interest points. Then estimated the dominant orientation by a summation of the sliding window orientation rotating covering an angle with all vertical and horizontal wavelet responses in the wavelet response. The more stable orientation whichhas maximum size is chosen to describe the orientation of the interest point. In step two, the region of interest is split to 4X4 subdivisions. The horizontal and vertical wavelet responses are calculated for each sub-region to form a first set of entries to the feature vector. Then divided each sub-region is into 5X5 and calculated Haar-wavelet responses to any point from 25 points. Using the equation shown below, four vectors were calculated as the two vectors dx, $|dx|$, two dy , $|dy|$for each critical sub-region.

$$v = \left(\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|\right) \quad (3)$$

Where dx is the horizontal response and dy is the vertical wavelet and the absolute values |dy|and |dy|are summed to acquire information about the polarity of the intensity changes. Moreover, to increase robustness to geometric distortion the Haar-wavelet are weighted with a Gaussian centered at the interesting point. The length of the SURF descriptor vector for all sub-regions is 64. The second step is to construct the scale invariant descriptor on each interest point found in the previous step. To achieve rotation invariance, a gel is aligned to the major orientation. The size of the rectangle is proportional to the scale where the interesting point is detected. The rectangle is then cropped into a 4 by 4 grid. Different informations such as gradient or absolute value of gradient are then subtracted from each of these subsquares and composed into the interest point descriptor.

D. *SIFT*

SIFT is one of the important local feature algorithms in machine vision proposed by D. Lowe, in 2004 in the

University of British Columbia [7], this algorithm solves a major problem such as Scale that Harris algorithm is poorly performed with the scale because a corner could change and stop being a corner if the image is scaled, so Harris is not scale invariant. The main use of SIFT is to recognize and retrieve an object by detecting and describing the key point. The definition of Key point is that it is the image feature that is distinctive and stays reliable and stable after applying a series of filters with more iteration and the key point should be robust against scaling, noise, changes of illumination, and rotation. SIFT descriptor length is 128 dimensional that provides rich information that was used in image matching and gave good accuracy. SIFT included four major steps listed as Scale-space extrema detection, key point localization, Orientation assignment and Key point description. One of the fundamental stages is scale-space that is used to recognize the scale and location for the same object in series Scale. The function named scale space is Responsible for detecting and identifying all locations that are invariant to changes in Scale in any possible Scale. The scale space function is the result of multiplication of I (x,y) by the Gaussian variable-scale G (x,y,σ) as below :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (4)$$

Where I (x, y) is the input image. The scale-space is based on the difference of Gaussian, D(x, y, σ) is used to detect stable areas of key points in the region of interest in scale space, that can be computed of the difference of two nearby scales parted by a fixed factor k[12] as in below:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad (5)$$

Figure (2) explains the approach of difference of Gaussian. The local maxima and minima of difference of Gaussian on the equation above to each point are compared with eight neighbors on each scale, and its nine neighbors down and above one scale. The points are called an extreme point if the value of this point is the minimum or maximum of all these points that are called neighbors.

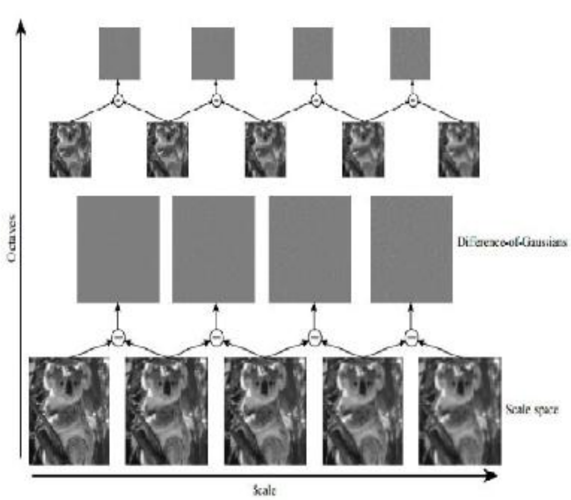


Figure 2: Explains the approach of difference of Gaussian

The next stage is to find the localization of key point, this is done after taking the Candidate points from the first stage where the Taylor expansion is calculated to determine the important points based on their stability under the changes and distortions of the image fixed points which are considered as important sample points, in this stage points that are not considered fixed and flexible against the changes that take place in the image are neglected. After that, specified orientations for each key point are set, based on regional image orientation regions. Then, the goal is to achieve the stability in image rotation, which is done by assigning a constant direction to each key point, whose object vector can be represented in relation to this direction. The direction of this key point is calculated from a histogram of the local gradients from the nearest sleek image. To compute the orientation θ (x, y) and gradient magnitude m (x, y) using the pixel difference as in the equations below:

$m(x, y) = ((L(x + 1, y) - L(x - 1, y))2 + (L(x, y + 1) - L(x, y - 1))2)½ \quad \theta(x, y) = \tan-1 ((L(x, y + 1)-L(x, y-1))/ (L(x + 1, y) - L(x - 1, y)))$

Every point is added to the histogram that has 36 bins covering all degrees of the orientation by the Gaussian with 1.5 times of variance and magnitude m(x,y)to the scale of key points to get more accurate orientation. The powerful heights of the histogram are added with their neighbours. The last stage of the SIFT algorithm is the key point Descriptor, which is used in matching of images and must be robust versus rotation, noise, and change of lighting, which is done by constructing a region of 16x16 about each key point in this area is divided into 4x4 size. A histogram with 8 bin is created for each 4x4 the result is 128 bin values and that is the length of the SIFT Descriptor, which is represented as a key point vector figure (3) explains the vector of each key point.

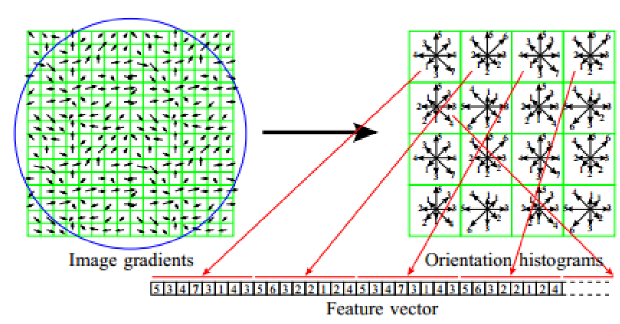


Figure (3) explains the vector of each key

E. ORB

First presented by Ethan Rublee et al. in 2011, the result of the integration of FAST detectors with BRIEF descriptor with the addition of several modifications to enhance performance, accuracy and speed [15]. FAST is used to find the key points and is done by the intensity threshold which is between Centre of the pixels and point in a circle around the Centre, typically, the radius of the circle is 9 and the points that have good performance are detected, Since FAST does not produce a barometer, the Harris corner is used to arrange points according to Harris measurement, and choose the highest N points. The corner orientation is computed by using the intensity centroid, supposing that intensity of corner is from the centre, an orientation is given from the direction of the vector from the corner point to the centroid. Moments are calculated by using the equation below:

$$m_{pq} = \sum_{x,y} x^p y^q I(x, y), \tag{6}$$

The calculation of the centroid with moments is performed by using the equation:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{7}$$

Where a The vector is constructed from the centroid to the centre of the corner. To improve the rotation, invariance moments are computed in both directions (x,y) in the circle region.Algorithm ORB solves the poor performance problem of the BRIEF descriptors in rotation by using and calculating the rotation matrix debugging direction, and the BRIEF descriptors are routed awarding to the orientation.

F. KNN

KNN is one of the classification algorithms that are most commonly used because of the ease of implementation and most of the results are acceptable and do not require prior knowledge of the data[9]. KNN algorithm classifies the data on the basis of similarity, where the search for the nearest similarity where this algorithm stores the training sets without any change and are considered as inputs, which means that the KNN algorithm does not have a learning stage. The nearest neighbour algorithm deals with the training data as n of dimensions, where n-dimensionality contains a set of features. To find and classify the test data, we find the nearest k to this data in the n-dimension of the training data, and this is done by a number of distance metrics such as Euclidean distance as explain in the law below:

$$\sqrt{\sum_{1 \leq i \leq n} (x_i - y_i)^2} \tag{8}$$

The step for the KNN algorithm is below:

Input Parameters: Dataset, k Output: Classification of the test data

1: Save total the data of training.

2: Classify each unknown tuple
loop

- ***Using distance metrics, compute its distance with all the training data***
- ***Find the k nearest training data to the unknown data.***
- ***Specify the class which is the most matching in the k nearest training tuples to the unknown tuple.***

End loop.

We see above that the algorithm of the KNN has two parameters of input, namely, the dataset and the coefficient of K, which enters in deciding the number of neighbours that are taken in the process of classification, and also there are several operations that are taken in the case if the values of the class is real classification [16]. However, there are several things that affect the KNN algorithm such as the type of distance metric and K value. The important thing is the values of a feature.

IV. RESULT AND DISCUSSION

In all cases (feature extraction algorithms), the dataset containing a file that contains various forms of weapons, which are considered as potential threats, have been dealt with. The key points are then calculated, extracted and stored in a vector of features. To classify the Input image as a threat or otherwise, calculate the same previous steps for extracting features and their descriptors then the KNN algorithm is implemented to find the match.An algorithm called random sample consensus (RANSAC) is used to remove false matches and recheck which is much unpretentious, the use of RANSAC filter increased the speed of the system and porduced good results. The performance of the SIFT algorithm with the KNN outweighed both of the SURF and the ORB algorithms as a position in the table. In addition, the KNN algorithm tends to be slower and also requires large storage spaces because the search for matching is done through the entire dataset. Notice that some overlapped x-ray images got misclassified. However, SVM was implemented as a classifier with both (SIFT and ORB) algorithms and gave good results, by comparing ORB which did not produce good results when using the KNN algorithm as a classifier.

Accuracy

The ratio of the total number of the images that are detected correctly by the system is measured

$$Accuracy = \frac{IPD + INU}{IPD + IPU + INU + IND} \tag{9}$$

- Where IPD is positive x-ray image of the gun that has been detected.
- IPU is positive x-ray image of the gun that was not detected.
- IND is negative x-ray image of the gun that was wrongly detected.
- INU is negative x-ray image of the gun that was not detected

Table 1. The accuracy of the proposed Algorithm

ORB+KNN	***SIFT+KNN***	***SURF+KNN***
85.4 ± 0.25	95.32 ±0.16	93.44 ±0.24

Table 2. CPU time to classify one query x-ray image

ORB+KNN	***SIFT+KNN***	***SURF+KNN***
3.7237230	16.732460	6.748203

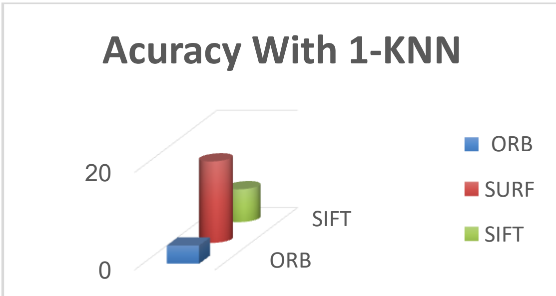


Figure 4: Explains the Accuracy with 1-KNN as classifier

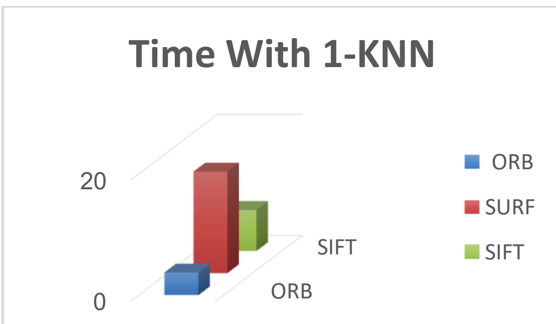
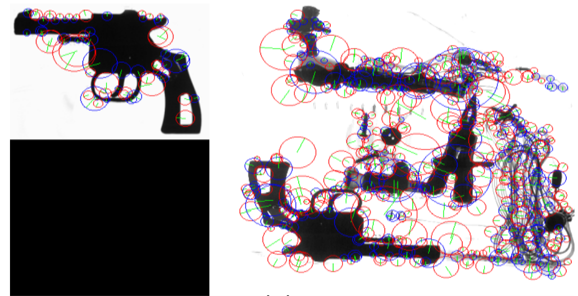


Figure 5: Time consumtig to classify Quray image

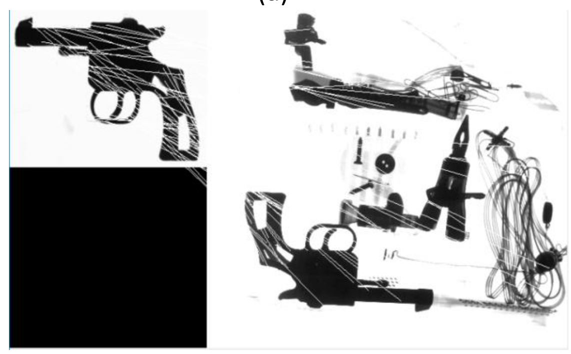
(a)

(b)

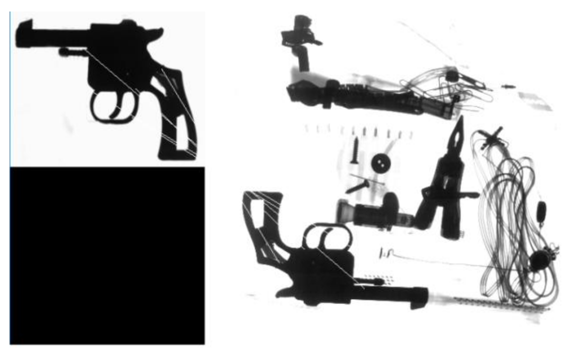
Figure(4)Sampales (a)and (b)detection of the matching points using SIFT algorithm



(a)



(b)



(c)

Figure(5)detect of matching point using SURF algorithm . (a)detect keypoint (b)compute matching (c) Recheak using RANSAC

V. CONCLUSIONS

This paper included the use of the local feature detection and description to detect the potential threat in the x-ray image of the luggage in airports or critical places.

In our approach, we have used the most important factor for the preprocessing, because the X-Ray images have high noise and a low level of clarity, has been used blurring with Kernal (3 x 3) and Histogram equalization to increase the clarity of the image. Three algorithms have been used to extract and describe the Interest point. The algorithm of the nearest neighbour has been selected as a classification algorithm . Three values were tested for K (k = 1 ,k = 3 and k = 5). The lowest K values were the best results.The proposed method compares three algorithms of the local descriptor namely (SIFT, SURF, ORB), in terms of accuracy and speed of processing. SIFT outperformed the rest of the proposed algorithms with high accuracy, SURF also produced good accuracy and gave higher speed than SIFT and robustness against noise, scale, and change in lighting. The ORB algorithm is faster than other methods but it gave poor accuracy when used KNN as a classifier. Notice that some overlap x-ray image got misclassification. For future work, it is planned to use deep learning to improve the detection.

VI. References

[1] Suhad Al-Shoukry, "An Automatic Hybrid to Detect Concealed Weapons Using Deep Learning", ARPN Journal of Engineering and Applied Sciences, 2017.

[2] V.Riffo, D. Mery, "Automated Detection of Threat Objects Using Adapted Implicit Shape Model". IEEE (2016).

[3] Rohit Kumar Tiwari and Gyanendra K. Verma, “A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector”, Procedia Computer Science Vol-54, pp703-712, 2015.

[4] Leonardo Carrer and Alexander G. Yarovoy, “Concealed Weapon Detection Using UWB 3-D Radar Imaging and Automatic Target Recognition”, IEEE, 2014.

[5] A. D. Lopez, E. S. Kollialil and K. G. Gopan, "Adaptive Neuro-fuzzy Classifier for Weapon Detection in X-Ray Images of Luggage Using Zernike Moments and Shape Context Descriptor," 2013 Third International Conference on Advances in Computing and Communications, Cochin, 2013, pp. 46-49. doi: 10.1109/ICACC.2013.16

[6] K.Velmurugan and Lt. Dr.S.Santhosh Baboo, “Image Retrieval using Harris Corners and Histogram of Oriented Gradients”, International Journal of Computer Applications (0975 – 8887), June 2011.

[7] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60(2), 2004.

[8] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, “SURF: Speeded Up Robust Features”, In Proc. ECCV, 2006.

[9] David Marshall. "Nearest Neighbour Searching in High Dimensional Metric Space", sub thesis submitted in partial fulfilment of the degree of Master of Information Technology, June 2006.

[10] Mery, Domingo, et al. "Detection of regular objects in baggage using multiple X-ray views." Insight-Non-Destructive Testing and Condition Monitoring, 55(1) (2013).

[11] Mery, Domingo. X-Ray Image Database. 2012. Raw data.

[12] Abidi, B., Y. Zheng, A. Gribok, and M. Abidi, "Improving Weapon Detection in Single Energy X-Ray Images Through Pseudo colouring".

[13] Abidi, B., Y. Zheng, A. Gribok and M. Abidi, "Screener Evaluation of Pseudo-Colored Single Energy X-ray Luggage Images". Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop, San Diego CA June 2005.

[14] Kaur, A., & Kaur, L. (2016). Concealed weapon detection from images using SIFT and SURF. 2016 Online International Conference on Green Engineering and Technologies (IC-GET).

[15] Ethan Rublee, Vincent Rabaud, Kurt Konolige and Gary Bradski, “ORB: an efficient alternative to SIFT or SURF,” IEEE International Conference on Computer Vision, 2011.

[16] Alka Lamba, Dharmender Kumar ."Survey on KNN and Its Variants", International Journal of Advanced Research in Computer and Communication Engineering, May 2016.

[17] Hossein Pourghassem, Omid Sharifi-Tehrani, and Mansour Nejati, “A NovelWeapon DetectionAlgorithm in X-ray Dual-Energy Images Based on Connected Component Analysis and Shape Features”, Australian Journal of Basic and Applied Sciences, 2011.

[18]Karami, Ebrahim, Siva Prasad, and Mohamed Shehata. "Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images." arXiv preprint arXiv:1710.02726 (2017).