ISSN : 0974 - 7435 Volume 10 Issue 16

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(16), 2014 [9381-9389]

A way of the robust acceleration optimization of image identification of X-ray machine used in airport security check based on the bag of words database model

Zhang Ning[1,2]*, Zhu Jinfu[2]

[1]Guangzhou Civil Aviation College, Guangzhou 510403, (CHINA)

[2]College of Civil Aviation of Nanjing University of Aeronautics & Astronautics, Nanjing 211100, (CHINA)

Email: anlan1008@126.com, daidai@nuaa.edu.cn

ABSTRACT

In airport security check, the demands of the accuracy of image identification of X-ray machine operators have become higher and higher. The different positions of items in the conveyor make the images shown in the computer displays different, which brings difficulty to the accurate identification. This article makes an analysis of the images and puts forward a way of robust acceleration optimization against the classical bag of words model (which has some flaws and needs to be improved). This new way can describe precisely the graphical features of the visual dictionary produced by visual images of X-ray machines, resist the influence of the complicated location and background information and categorize the information that is put in the sorters of support vector machines (SVM). Through experiments and analysis, it is proved that this way can increase the accuracy of the operators' graphical identification and achieve a good effect with few experimental images, which means it can increase the accuracy and the efficiency of the operators' identification of difficult images.

KEYWORDS

Airport security check; Bag of words model; Robust acceleration optimization; X-Ray machine; SVM.

INTRODUCTION

X-ray machine operators may make mistakes when identifying images of dangerous goods in the airport security check. They are highly nervous and concentrated in order to avoid errors and secure the flights while the different angles and positions of the items make the 2D images shown in the displays different due to the effect of the different density and volume of different items on the X-ray penetration. That is why they have difficulties in the identification with a limited time. For example, the different positions of the same pistol in the same box make A and C more difficult to identify than B (see Figure 1).

The main idea of this article is to construct the visual dictionary of X-ray machine images and establish the bag of words model with which the categorization of the images can be finished in order to offer a scientific guidance to the routine practices of X-ray machines. Visual dictionaries will be made for the same dangerous good in different positions; visual lexicons will be made for the same kind of dangerous good of different density; the categorization of the images will finally be finished due to the establishment of the bag of words model which helps operators rapidly identify dangerous good in a limited time, reduces their blindness in image identification training, increases the check efficiency and decreases the possibility of false positives or false negatives in the identification of dangerous goods.

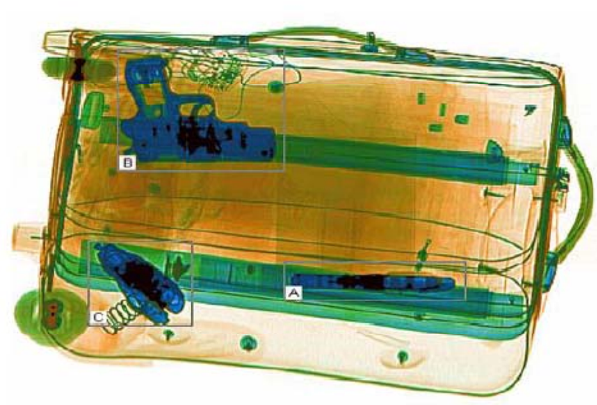

Figure 1: X-ray images of the same pistol at three different positions in the same box

LITERATURE REVIEW

Maryam Jaberi published in 2014 a thesis in Accurate and robust localization of duplicated region in copy–move image forgery applying the classical SIFT method of the extraction of image features[1]. M. G. Ponomarev put forward in 2014 the SIFT method of the extraction of the X-ray machine image features of metal parts[2]. Ashnil Kumar and Jinman Kim put forward an improved way of robust optimization of SIFT in a 2014 thesis[3]. Qizhi Xua, Yun Zhang and Bo Li used the normalization of images and the categorization of images that significantly improved the accuracy of the image matching in their 2014 thesis[4]. Jeongin Seo and Hyeyoung Park in their thesis *Robust identification of face with partial variations using local features and statistical learning* put forward an improved SIFT method of the extraction of the 2D-space image features[5]. Melloni, P. Bestagini and others in their thesis *Attacking image classification based on Bag-of-Visual-Words* believed that the SIFT and SURF methods of the extraction of image features had their limited and suggested to apply the SVM method in the categorization[6].

This article analyzes with details the bag of words model for the extraction of image features based on the results above and categorizes the X-ray machine images in the airport security check with the construction of the bag of visual words dictionary. The SIFT method will be firstly applied to extract the key points of the images and then the SURF method will be applied to figure out the features information of the distribution of items in the images. Finally the bag of words model will be applied in the categorization of the SVM images. After experiments and comparisons, this method is proved more accurate than the classical one.

BAG OF WORDS DATABASE

Bag of words model

Bag of words model is originally the technique of the processing of natural languages applied in the text information retrieval and text categorization. It is sometimes called the bag of features algorithm[7]. The model presumes that a document is the piling of random words which are independent from each other and have no relationship with the emergence of others. It achieves great success in the text categorization due to its rapidness and high efficiency. But this model neglects the connections and the locational relationship between different features which causes some loss of information.

Establishment of the bag of words model

The bag of words of X-ray machine images maps the 2D image information onto the collection of the visual key words which not only reserves the localized features of images but also effectively compresses the image description. The image extraction by X-ray machine of dangerous goods like pistols and knives from different angles is shown below (Figure 2, 3).

Figure 2: Image extraction by X-ray machine of the pistol from different angles

Figure 3: Image extraction by X-ray machine of the knife from different angles

Visual dictionary

To construct the collection of visual key words on the phase of learning: to extract all localized features among the collection of the training images and cluster these features to get the universal features of the training collection. We call the collection of these archetypal features "visual dictionary". With the image-representation method based on visual dictionary, the captured images are considered to be a collection of a series of random localized features. Only the apparent descriptive information of the localized features will be stressed with the space relationship between them ignored. And this offers an effective representation method of exterior scene model for the examination of the visual closed-loops. We have summarized the visual words of cutting tools in the X-ray machine image database (Figure 4).

Figure 4: Various cutting tools image combination in the X machine on the screen of visual words

SIFT ALGORITHM

Principles of SIFT

Scale-invariant feature transform (SIFT) is an algorithm of computer vision that detects and describes the localized features of images. This algorithm searches for extreme points in space scale and extracts their locations, scales and rotation invariant. David Lowe created this algorithm in 1999 and improved it in 2004[8].

The SIFT algorithm matches the features of the same thing in two different pictures and builds the corresponding relationship between them.

To summarize it, SIFT firstly constructs the representation of the space scale of images and then detects the space feature points of images. It then defines the main direction of the feature points and finally produces the descriptive factors of the feature vectors.

Shortages of SIFT

1. The application of henssian matrix in the capture of the extrema of local images in SIFT is quite stable. But SIFT relies too much in the calculation of the main direction upon the graded directions of pixel in local areas which may cause the inaccuracy of the main direction. The feature vector extraction and matching all rely on the main direction whose slight deviation may cause an enlarged error and an unsuccessful matching.

2. All images captured by henssian matrix are not within the same scale. So when the scale of images changes, the corresponding relationship between images may not be easily defined.

3. The algorithm describes the feature points by calculating the histogram of the neighboring area of the feature points thus can deal with the matching problems when the images are translated, rotated or under affine transformation. But the processing data are of a large amount, the complexity of the time of it is high and the spending time is long.

4. Image pyramid is a way of image representation that combines down-sampling operations and smoothing operations. A great advantage of it is a much reduced amount of calculation due to the reduction of pixel layer after layer upwards. The disadvantage is that the upward quantization of the pyramid will become rougher and rougher while the speed is too fast. The insufficient sampling of pyramid layers will also cause errors of the scale on which the feature vector extraction relies. The compromising solution of SIFT is to choose adequate layers and apply interpolation. But SIFT only utilizes the gray-scale features but lacks the information of colors.

5. The same thing in different pictures may have different shapes, size, angles, brightness, even distortion. The knowledge of computer vision shows that the projective correspondence between two flat-shaped objects can be established through the images captured by optical lenses. But the dangerous good in flights like pistol, explosives and drugs are mostly curved. For this reason two images of the same curved object in different angles, distances or with different camera parameters don’t have a linear correspondent relationship, that is, even we have several pairs of matched points within two images of some dangerous good, we still can’t conclude the correspondence of other points.

6. For blurred images and smooth-edged images, the feature points are not enough. For circles, the problem is even severer.

SURF (SPEED-UP ROBUST FEATURE) ALGORITHM

Principles of SURF

Because the SIFT algorithm has many shortages, we used an algorithm that is based on Speed-up Robust Feature (SURF). It can calculate swiftly based on the integral images and use hessian detection to detect the feature points. We will define the main direction and the descriptive points of the feature point through calculating the Haar wavelet transform and achieve the matching of the feature points between images according to Euclidean distance between vectors. SURF is to a certain extent an improvement to SIFT. It is also a detector of feature points that has the invariability in scale and rotation. It not only suffices for accuracy but also has a small calculating amount and a fast calculating speed.

Process of SURF

The three key procedures of SURF are:

1. Using integral images to finish the operations of image convolution;
2. Using Hessian matrix to detect the feature values;
3. Using descriptors (localized information) based on distribution.

We use the following figure to achieve the process (see Figure 5).

Feature extraction

Feature extraction is a concept in computer vision and image processing which refers to the extraction of image information by computer in order to determine whether the points of every image belong to an image feature. The result of it is to divide the points into different subsets which are usually isolated points, continuous curves or continuous regions. Therefore, one of the most important characteristics of image extraction is repeatability, that is, the features of the same object extracted from different images should be the same.

The extraction process of the feature points of the bag of words model is summarized below (see Figure 6):

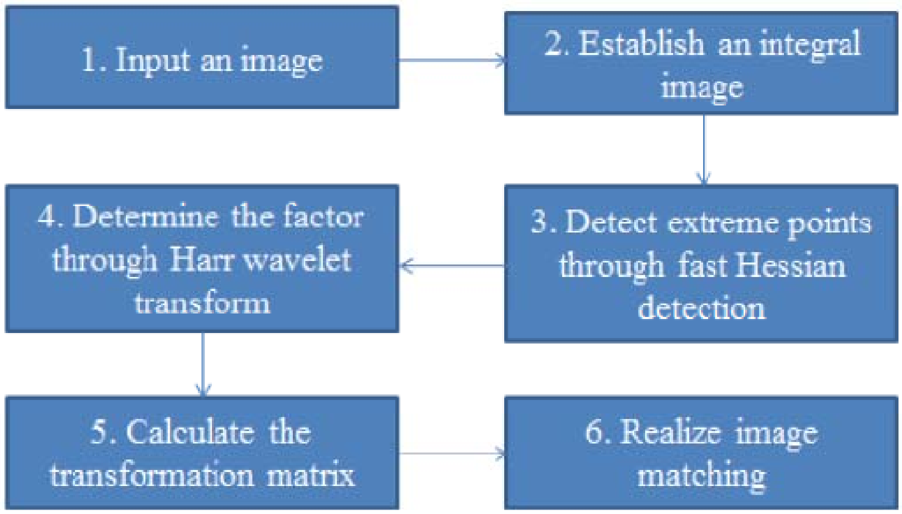

Figure 5: The basic framework of SURF

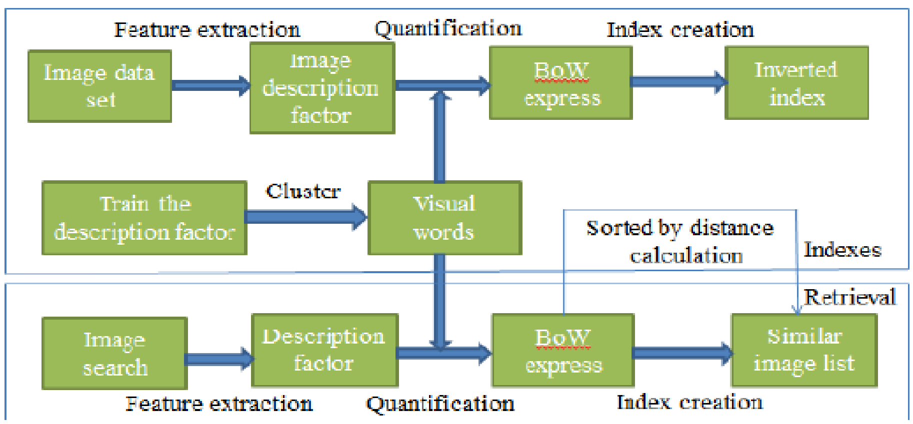

Figure 6: The framework of BoW model retrieval

Matching of the feature points

The representation of low-level visual features of images is the basic procedure for image-based modeling, image retrieval, categorization and identification and other tasks. In the image understanding or other applications, computers mark semantically the correspondent targets and regions and produce cognitive results that are made easy to understand through the calculation, analysis and deduction of the input scenes. Therefore, the high-level semantic acquirement is important for the analysis and understanding of images; the formation of the corresponding relationship of the probabilities between the low-level visual features and the image semantics reflects the transformation between statistics and concepts. Researches based on the image semantics emphasize the analysis, extraction and descriptions of the low-or-middle-level visual statistical features and the modeling of the mapping relationship between the semantics of the localized features[10].

SUPPORT VECTOR MACHINE (SVM)

Introduction

The visual dictionary can describe the image features more precisely and can resist the influence of the constantly changing information of location and background. And then it applies the method of pyramid matching in the representation of the images based on the histogram of the visual dictionary which then is inputted in the SVM sorter for categorization (see Figure 7).

Categories of SVM

According to Statistical Learning Theory, the practical risks of learning machines consist of empirical risk values and fiducial range values. The learning method based on the principle of the minimal empirical risks only stresses the minimal error of the empirical risks of the training sample without the minimal fiducial range values which makes it hard to be generalized. The Support Vector Machine, or SVM, invented by Vapnik takes the training errors as the constraint to the optimization and sets the minimal fiducial range values as the goal of the optimization, that is, SVM is a learning method based on the principle of the minimal structural risks and has a better generalization ability than some classical learning methods[9].

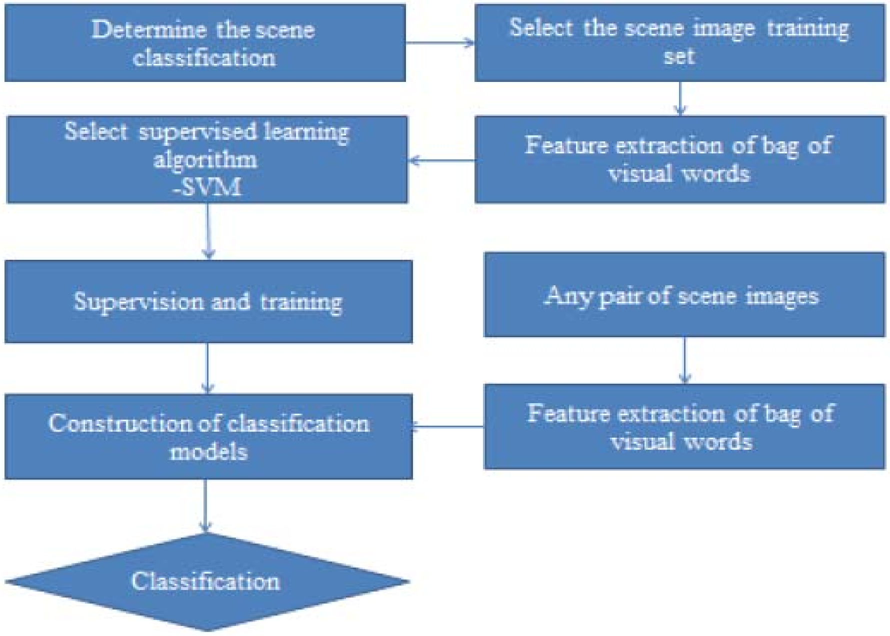

Figure 7: SVM image classification system based on the bag of words model

The core idea of SVM is based on the principle of the minimal structural risks to map the non-linearly separable statistics in the input space through kernel functions onto the high dimensional feature space to make it linearly separable and then to form a low-VC-dimensional optimized categorization hyperplane as the judging plane in the high dimension in order to make farthest the distance between the two types of statistics and the plane. The goal of SVM algorithm is actually to find an optimized categorization hyperplane that can cause the minimal structural risks.

For the two types of categorization problems, make the training sample set $T = \{(x_1, y_1), x_{2,} y_2, \cdots, (x_n, y_n)\}$, in which $x_i \in X = R^n$ is the sample vector, $y_i \in Y = \{+1, -1\}$ is the type number, n is the amount of samples, if there exists a mapping $\phi(x_i)$, that maps x_i from the archetypal feature space X onto the high-dimensional feature space F, and introduces the slack variable ξ_i, then the original problem of SVM can be represented as:

$$\min_{w.b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.}\begin{cases} y_i(w \cdot \varphi(x_i)+b) \le 1-\xi \\ \xi_i \ge 0 \end{cases}, \ i = 1,2,\cdots,n \quad \textbf{(1)}$$

Through the Lagrange function method the dual problem of the archetypal problem can be deduced:

$$\max -\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}\alpha_i\alpha_j y_i y_j k(x_i, x_j) + \sum_{i=1}^{n}\alpha_i$$

$$\text{s.t.}\begin{cases} \sum_{i=1}^{n}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{cases}, \ i = 1,2,\cdots,n \quad \textbf{(2)}$$

in which: $k(x_i, x_j)$ is the positive definite kernel function, polynomial kernel function (POLA), Gaussian radial basis function (GRBF), Sigmoid kernel function and other functions that fulfills the Mercer Theorem[11].

SVM categorization of visual models of X-ray machine images

We use SURF to speed-up the identification speed of X-ray machines and apply the categorization method of SVM. Compared to the classical methods, the functions of detection and the matching of time are greatly improved. We reduce the whole matching time, get proper number of the matched points, has a high matching rate and a more accurate calculating result. The process is below (see Figure 8):

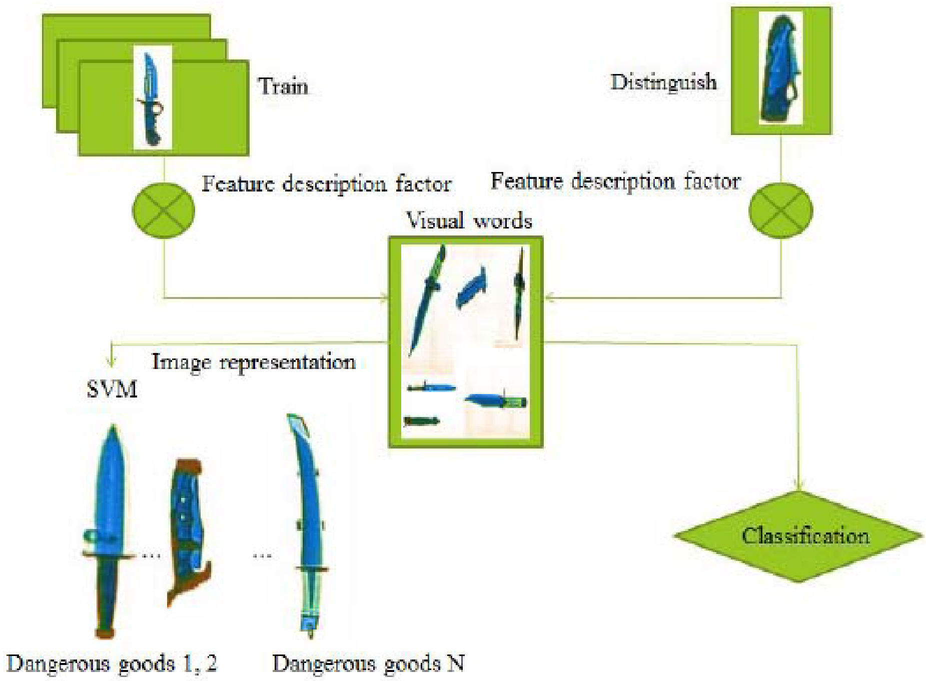

Figure 8: The basic classification of dangerous goods in the bag of words model of security

EXPERIMENT

We found an experienced operator of the X-ray machine and designated 15 dangerous good on the conveyor of the X-ray machine with five different angles in order to test his performance in the identification of the images of dangerous goods. Then we trained him in image categorization for about one month. When the training was over, his performance was generally improved (TABLE 1).

TABLE 1

Dangerous goods	0°		15°		30°		45°		60°	
	Before training	After training	Before training	After training	Before training	After training	Before training	After training	Before training	After training
TNT	67	87	71	90	73	92	71	89	76	93
Knives	61	84	66	86	71	90	71	93	78	94
Bullet	56	78	59	82	63	86	69	88	74	89
Toxic	83	95	87	98	88	96	88	97	80	91
Gun	90	100	93	100	92	99	92	100	89	98
Pistol	88	98	82	95	89	99	89	100	90	99
Alcochol	69	82	73	88	70	91	78	95	74	90
Scissors	96	100	93	99	94	98	96	100	92	100
Hairgel	98	100	92	96	97	100	99	100	96	100
Tear gas	87	95	83	99	88	99	88	92	81	96
Crossbow	93	98	95	97	92	98	95	100	91	100
Detonator	86	93	82	99	88	96	90	99	89	100
Fireworks	91	100	92	99	89	97	92	100	88	96
Lighter gas	95	100	93	99	92	98	96	100	93	99
Explosives	91	99	88	97	86	98	93	99	92	98

The vertical axis of Figure 9 represents the performance before and after the training measured from 0 - 100 points. The horizontal axis represents the performance towards the dangerous goods in 6 different angles. We can see the difficulties of the image identification increases from low to high (Figure 9).

We apply the ROC curve to test the identification effect before and after the training[12]. ROC curve represents the dynamic curve of sensitivity and specificity as the judging dividing values change. It can evaluate objectively and accurately the identification level of the X-ray machine images which adds to the scientificity of the X-ray machine training. See Figure 9, we have here applied the DPD statistics software by Professor Tang Qiyi[13].

The full line in Figure 10 represents the theoretical values, and the imaginary line represents the 95% fiducial space. We can get that the area of ROC is 0.801 which proves that the training effect is pretty good.

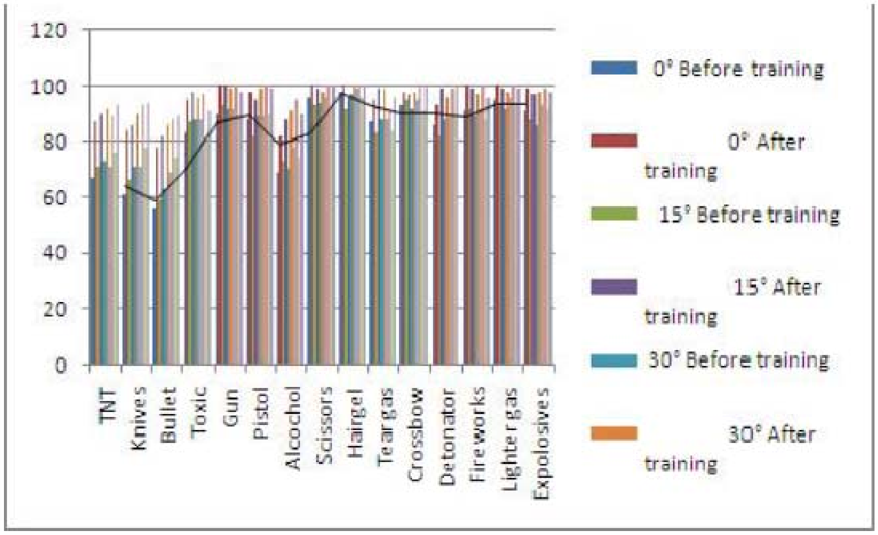

Figure 9: Comparison of the operator's performance of identification before and after the training

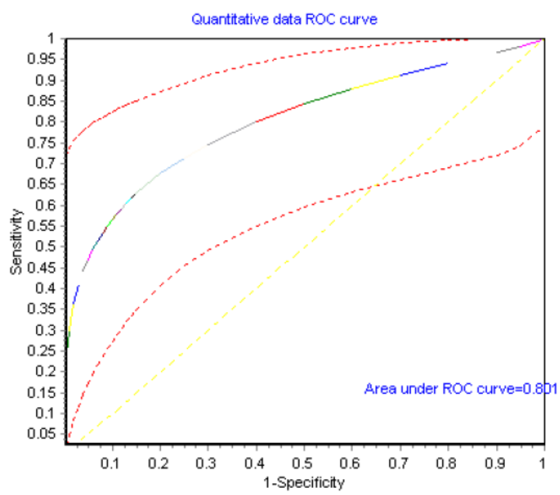

Figure 10: ROC curve of the image identification level of the operator

CONCLUSION

This article improves the bag of words model used in X-ray machine images and uses the visual dictionaries to induce the image features with different angles. We also use SURF algorithm to undertake robust optimization identification and finally categorize the images with SVM categorization method. Moreover, we combine the categorized visual words and the training of the operator and have undertaken some pertinent experiments. The experiments show that the image identification level of the operator after the training is increased.

BIBLIOGRAPHY

[1] Maryam Jaberi, George Bebis, Muhammad Hussain, Ghulam Muhammad; "Accurate and robust localization of duplicated region in copy–move image forgery", Machine Vision and Applications, **25**, 451–475 **(2014)**.

[2] M.G.Ponomarev, C.Selcuk, T.H.Gan1, M.Amos, I.Nicholson, M.Iovea, M.Neagu, B.Stefanescu, G.Mateiasi; "Defect detection and classification system for automatic analysis of digital radiography images of PM parts", Powder Metallurgy, **57(1)**, 17-24 **(2014)**.

[3] Ashnil Kumar, Jinman Kim, Lingfeng Wen, Michael Fulham, Dagan Feng; "A graph-based approach for the retrieval of multi-modality medical images", Medical Image Analysis, **18**, 330-342 **(2014)**.

[4] Jeongin Seo, Hyeyoung Park; "Robust identification of face with partial variations using local features and statistical learning", Neuroconmputing, **129**, 41-48 **(2014)**.

[5] Qizhi Xu, Yun Zhang, Bo Lia; "Improved SIFT match for optical satellite images registration by size classification of blob-like structures", Remote Sensing Letters, **5(5)**, 451–460 **(2014)**.

[6] A.Mellon, P.Bestagin, A.Costanzo, M.Barni, M.Tagliasacchi, S.Tubaro; “Attacking image classification based on Bag-of-Visual-Words”, WIFS '2013. Guangzhou. China, ISBN 978-1-4673-5593-3 ©2013 IEEE, 103-108 November 18-21 **(2013)**.

[7] Kraisak Kesorn, Stefan Poslad; “An Enhanced Bag-of-Visual Word Vector Space Model to Represent Visual Content in Athletics Images”, IEEE Transactions on Multimedia, **14(1)**, 211-221 February **(2012)**.

[8] David G.Lowe; “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, 1-28 January 5 **(2004)**.

[9] Swarnajyoti Patra, Lorenzo Bruzzone; “A Novel SOM-SVM-Based Active Learning Technique for Remote Sensing Image Classification”, IEEE Transactions on Geoscience and Remote Sensing, **52(11)**, 6899-6910 November 2014.

[10] Fan Zhu, Ling Shao; “Weakly-Supervised Cross-Domain Dictionary Learning for Visual identification”, Int.J.Comput.Vis., **109**, 42–59 **(2014)**.

[11] Xiyan He, Gilles Mourot, Didier Maquin, José Ragot, Pierre Beauseroy, André Smolarz, Edith Grall-Maës; “Multi-task learning with one-class SVM”, Neurocomputing, **133**, 416-426 **(2014)**.

[12] Marine Lorent, Magali Giral, Yohann Foucher; “Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality”, Statistics in Medicine, **33(14)**, 2379–2389 June **(2014)**.

[13] Q.Y.Tang, C.X.Zhang; Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research, Insect Science, **20(2)**, 384-390 **(2013)**.