http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 4 (4): 387-390, 2005
ISSN 1812-5638

Hedging Uncertainty in Rough Set-based Approach with Fuzzy Decision

[1]Ning Chen, [1]Bo-Qin Feng, [2]Haixiao Wang and [3]Hao Zhang
[1]Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China
[2]29 Bayowski Road West Orange, NJ 07052, USA
[3]Shenzhen Traffic Control Center, Shenzhen, 518000, China

Abstract: This study introduced a technique for authenticating the vehicle engines by comparing the images of the imprints of the identification number acquired when the vehicle was first registered and the ones acquired from the routine yearly vehicle inspection. The images were taken by rubbing a pencil over a piece of paper covered over the images and then scanned into a computer as binary image. Due to the nature of the acquiring technique, the acquired images have lots of artifacts caused by the shape and the condition of the engine surface and unevenness of rubbing the pencils by hand. The rough set-based approach was used to handle uncertainty arisen from artifacts in the acquired images. But, it has been proved to be NP-hard to find all reductions and a minimal reduction and we generally use different heuristic algorithms to find a set of reductions. By an examination of prior knowledge, Gaussian distribution to describe uncertainty to achieve a minimal reduction was used. The approach can distinguish between two similar images on the basis of Inductive Logic Programming, as is superior to conventional pattern-recognition approach being merely capable of classifier. Furthermore, it can avoid some failures of the approach based on the correlation coefficient to authenticate binary image. The experiments show an accuracy rate close to 93.3%.

Key words: E-government, uncertainty, automatic authentication, rough set

INTRODUCTION

Many techniques were developed for authentication against forgeries. Digital techniques utilizing cheaper and cheaper computation cost, such as digital watermarking technologies, are studied extensively in the passing decade. In a world prevailed with computer-based technologies, many traditional ones are still of great values. For example, labels with bar codes are used for managing commodities sold in stores and imprints (i.e. countermarks) are used to track individual piece of equipments or machineries. To authenticate the labels and imprints now largely relies on discerning eyes of well-trained personals. Presently in China, the vehicle management bureau employs a method of checking the genuineness of a vehicle by comparing the images of countermark imprinted on the engine. Automation of this process would bring a huge benefit socially and commercially by improving the efficiencies, considering the large amount of vehicles. Various techniques have been tried in the past[1,2].

Related research: There are many learning models in the pattern recognition field, which can be used to implement some kinds of automatic authentication task. Examples include template matching algorithm, Artificial Neural Network (ANN) algorithm, Support Vector Machine (SVM) algorithm based on Statistical Learning Theory (SLT) and many others[3]. They can work well in many situations, whereas sometimes they can't distinguish barely noticeable dissimilarity between the image and the template, owing to their mere capability of the classifier, no capability of the authentication. Furthermore, most of them normally require a correlation coefficient or distance metric to judge the closeness of the target image with the template, which often fails to deal with binary image; however the image of countermark for authentication is just binary image in present case. For example, it is obvious that the image of Fig. 1b will match well to image of Fig. 1a based on the correlation coefficient and so does Fig. 1c; thus the approach is incapable of distinguishing between them.

Zdzislaw Pawlak proposed rough set theory in 1982. It can deal with the classificatory analysis of data tables in which the data can be acquired from measurements or from human experts. The main purpose of the rough set analysis is the induction of approximations of concepts from the acquired data. The classical rough set analysis

Correspondign Author: Dr. Ning Chen, Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China

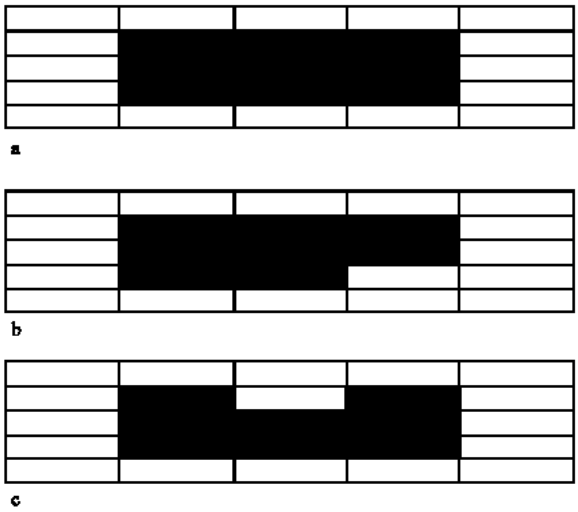


Fig. 1: Both image of (b) and image of (c) are similar to image of (a) with same correlation coefficient

is based on the indiscernibility relation that describes indistinguishability of objects. The concepts are represented by their lower and upper approximations, on the basis of which modeling uncertainty arisen from artifacts in the acquired images has advantage over other approach; hence it is well suited for dealing with image.

This study presents a computational technique for analyzing images taken by rubbing pencil on a piece of paper on countermark imprinted on rigid body surfaces. An approach based on theory of rough sets was chosen as authentication process, which is an extension of one using rough sets as a framework for processing digital image and handling uncertainty in binary image presented by Divyendu Sinha and Phillip Laplante[4]. First, consider finding a template, T, for matching target image of the given identification object (i.e. countermark) in the image space. Divyendu and Phillip[4] proposed a new approach, which rests on theory of rough set to automates the construction of gray-scale templates, thus the approach can avoid the failure arisen from the situation that binary template usually is adopted for binary image. This study proposes the extension of their approach from handling spatial uncertainty in x-direction and y-direction to handling uncertainty in all direction arisen from artifacts in the acquired images; Furthermore, a new method for resolution of uncertainty into x- and y-components is discussed, as has been proved by real-life experiments later. Finally, compute the correlation coefficient for authentication.

Model for authentication: A distorted object is one that has undergone some kind of deformation in shape. Due to nature, authentication problem is considerably more difficult compared with that of conventional invariant object recognition. Furthermore, the images of real-world object acquired under nonlaboratory conditions are often subject to distortions of a more complex nature, which cannot be avoided in real-world images acquisition environment because of equipment limitations and constraints in the viewing geometry.

There are two distinct methods for achieving distorted object authentication. One approach achieves invariant authentication by training classifier, as discussed previous. The other uses features that remain invariant regardless of object distortion, such as fingerprint authentication. However, no such invariant feature for discrimination exists in present case because of existence of forgeries with identical symbol feature shown in Fig. 2.

Combined with Inductive Logic Programming, we present authentication model: images as objects for authentication are described by a fixed repertoire of features, also called attribute. Therefore object can be expressed as a tuple, as well as identical object with identical tuple. Combined fuzzy decision, rough set theory was used to describe uncertainty of attribute and correlation coefficient to describe logical relation between corresponding attributes, i.e. the following form:

$$\text{Pair (attribute11, attribute12 ; attribute21, attribute22)} \quad \text{object1} = \text{object2} \tag{1}$$

The object1 is object with attributes such as attribute11, attribute12, etc and the object2 is another object with attribute such as attribute21, attribute22, etc.

Key techniques: The key techniques include the approach based on theory of rough set and the improved approach based on the correlation coefficient. Alicja and Andrzej[5] presented a rough set based approach to automatic classification, but result is a traditional classification, not authentication. Classification aims to discover similarity among different objects for grouping, while authentication aims

Fig. 2: Rubbing

to discover dissimilarity among similatives for identification. The rough set based approach presented in this paper, based on improvement on approach[4], can work adaptively to construct template for distinguishing between similatives. On the other hand, binary image for authentication can be regarded as gray-scale image to match gray-scale template acquired from the rough set-based approach, hence the approach based the correlation coefficient is improved to avoid failure in present case.

Rough set-based approach: The rough set-based approach describes inexactness by exactness, i.e. exact upper approximation and exact lower approximation (Fig. 3) and generates a gray-scale template, which contains inexactness introduced by artifacts bounded with upper and lower boundaries. These boundaries are determined as below according to the theory of rough set:

$$L(X) = \cup \{Y \subset \Omega \mid Y \subseteq X\} \quad (2)$$

$$U(X) = \cap \{Y \subset \Omega \mid Y \cap X \neq \phi\} \quad (3)$$

It has been proved to be NP-hard to find approximation boundaries on the theory of rough set[6]; hence, it is helpful to use heuristic algorithms to find them. By an examination of prior knowledge, Gaussian distribution was used to describe uncertainty to achieve a minimal reduction (Fig. 4). Considering some constrains in present case, such as the known error arisen from uncertainty, upper and lower approximation can be obtained. The uncertainty of approximation boundaries takes on a Gaussian distribution as below, as is demonstrated by central limit theorem.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad - \quad < x <$$

(4)

For a Gaussian distribution, the empirical rule states that 68% of the data elements are within one standard deviation of the mean, 95% are within two standard deviations and 99.7% are within three standard deviations. The empirical rule is often stated simply as 68-95-99.7. Depending on the rule, boundaries located at 99.7% are defined as approximation boundaries.

Furthermore, the uncertainty arisen from artifacts in the acquired images is in all directions, as can be solved by projecting it onto x-direction and y-direction, a new method presented in the study.

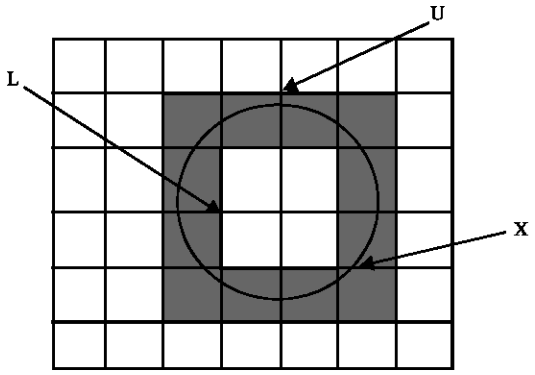


Fig. 3: Upper approximation and lower approximation of uncertainty arisen from artifacts

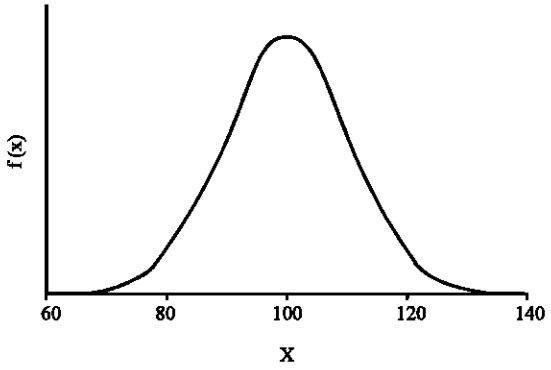


Fig. 4: Gaussian distribution

Correlation coefficient: The correlation coefficient is a concept widely used, referring to pattern recognition; it provides the basis for handling uncertainty and imprecision between objects, for matching spatial entities, for merging spatial data sets, for change detection. Since similarity is the basis, there needs to be a measure to make it quantifiable. Additionally, correlation coefficient is the central notion for any abstraction and has been discussed in the classification controversy as an undecidable problem, thus many methods to compute correlation coefficient are presented. For sake of simplification, a classical method of computing correlation coefficient is discussed here.

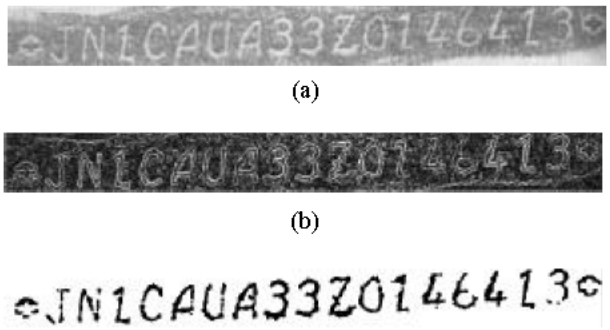
$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad (5)$$

In above equation, r is correlation coefficients, which may range in value from -1.0 to 1.0.

Authentication algorithm: Combined with the knowledge of real-life system, we can acquire an Authentication Algorithm detailed as follow:

Algorithm: Authentication Algorithm
Stage I: Program for construction of template
Step 1: Error of uncertainty is obtained by statistics of error;
Step 2: Resolution of error of uncertainty into x- and y-components;
Step 3: Upper approximation boundary and lower approximation boundary are determined with x- and y-components with error of uncertainty as minimal resolution;
Step 4: For each combination of boundary, the collection of upper and lower rough sets can be determined $\{<L_i, U_i>: I=1,2,\ldots,n\}$;
Step 5: Automatic construction of template by the framework[4].
Stage II: Authentication program
Step 6: Computing correlation coefficients by the function (normxcorr2) in Matlab toolbox;
Step 7: To arrive at a conclusion on the basis of Inductive Logic Programming.

Experiment result: The algorithms were coded in java and MATLAB and the images for authentication are



(a) (b) (c)

Fig. 5: (a) Rubbing (b) Resulting of edge detection (c) Artificial standard image for authentication

shown (Fig. 5). For sake of performance of computing correlation coefficients, edge detection is performed in advance.

The approach with 150 authentication examples was tested and it had 10 mismatches, which attained an accuracy of 93.3%.

CONCLUSIONS

This study was proposed to use a rough set-based approach combined with the correlation coefficient algorithm to authenticate engine imprints on the basis of Inductive Logic Programming. The experiments show an accuracy of 93.3%, which is not perfect but can potentially be used to pre-screen the vast amount of images to be verified.

We would improve the technique in these sections: modification of algorithm of comparisons, such as using Hausdorff distance and appending algorithm to remove noise etc and collect more experiment data to perfect the approach.

REFERENCES

1. Siwei, L. *et al*., 2004. A digital technique for art authentication. PNAS., 101: 17006-17010.
2. Gestal, M. *et al*., 2004. Classification of apple beverages using artificial neural networks with previous variable selection. Analytica Chimica Acta, 524: 225-234.
3. Andrzej, C., 2003. Automatic identification of sound source position employing neural networks and rough sets. Pattern Recogn. Lett., 24: 921-933.
4. Divyendu, S. and L. Phillip, 2004. A rough set-based approach to handling spatial uncertainty in binary images. Eng. Applications of Artificial Intelligence, 17: 97-110.
5. Alicja, A. W. and C. Andrzej, 2003. Rough set based automatic classification of musical instrument sounds. Elsevier Sci. B. V.
6. Guan, J. W. and D. A. Bell, 1998. Rough computational methods for information systems. Artificial Intelligence, 105: 77-103 .