ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Ranking of Resumes Based on Certification Images to Enhance the Efficiency of Hiring Process

Dr.K.Subramanian[1], M. Latha[2]

Assistant Professor, Dept of Computer Science (HEAD), VSS Government Arts and Science College, pulankurichi, Tamilnadu, India

Research Scholar, J J College for Arts and Science (Autonomous), Pudukkottai, Tamilnadu, India

ABSTRACT: The number of resumes posted daily seeking for a job by the job seekers increases gradually and this colossal data intricate the recruitment agency to identify the right candidate for the right job. This paper focuses on extracting the certifications acquired by the candidates and ranks the resumes accordingly to minimize and facilitate the recruitment process considerably. Resumes comprise of unstructured data without any uniform format and few resumes may contain details about the certifications and few resumes may be provided with the certification company logo to denote that the candidate has acquired credential and expertise. This paper deals with extracting the image logo from the document and applies image processing techniques to identify the certifications and provide the weight age to the candidate's resume based on the weight age of the certifications.

KEYWORDS: Resume mining, certification extraction, image processing.

I. INTRODUCTION

Image feature extraction from documents is a new area and needs lot of research attention of late. Thousands and thousands of resumes are posted every day from the candidates seeking job and to classify them and rank them according to the experience, skills, expertise, and qualification is a cumbersome process and even if the classification or clustering is carried out, there arises a major problem during the ranking of resumes. Let us assume two candidates with similar kind of experience, skills, qualification and age, here while ranking both the candidates acquires almost same rank. But the recruiters scrutinize whether there are any advantage or looks for some unique skills present in the resume to call for interview. Here certifications from reputed agencies acquired by the candidates' plays a pivotal role during ranking and the candidate with certification will be awarded higher rank than the one who does not have. But the problem here is, as we are aware of the fact that the resumes posted will be of unstructured format and resumes may contain certification details in textual format or in image format. If the details are provided in textual format, it can be mined to provide ranking accordingly, but what if it is in image format. This paper provides a solution to mine the certification details provided in image format and rank them accordingly.

II. PREVIOUS WORKS

Image matching is an important task to be performed to provide rank to the resumes. The distinction between different matching primitives is probably the most prominent difference between the various matching algorithms. Digital image matching automatically establishes the correspondence between primitives extracted from two or more digital images depicting at least partly the same scene. [1,2]Maximum-.Likelihood Image Matching methods have been improved using a probabilistic formulation for image matching in terms of maximum-likelihood estimation that can be used for both edge template matching and gray-level image matching. With the help of the transformation parameters achieved at low resolution level, we can apply block-wise SIFT extraction and matching to improve the efficiency. [3].The intensity value of a gray scale image varies from 0 (pure black) to 255 (pure white), allowing us to have 256 different shades of gray values. In an image, matching can be defined as the establishment of the correspondence between

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

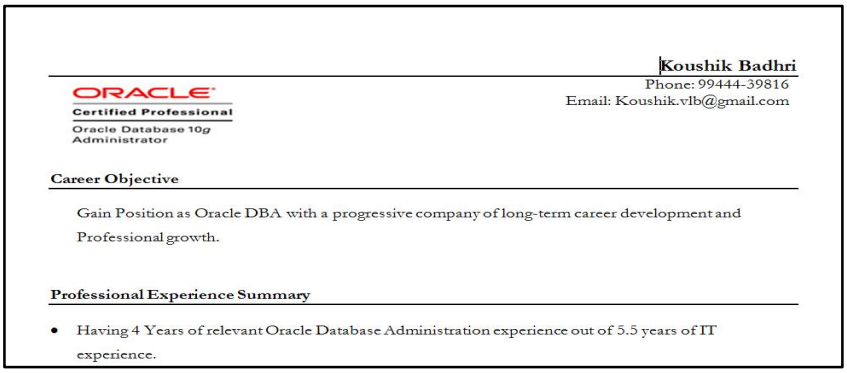
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

various data sets. The matching problem is also referred to as the correspondence problem. The data sets can represent images, but also maps, or object models. Many steps of the photogrammetric processing chain are linked to matching in one way or another. [4] The author introduced a new technique to match two images, in this method the object image is searched through the given reference image where it is searched for its geometric location. For a given image T (Template Image) of size N X N, is an object of interest to be searched in reference image of size PXP image that possibly contains that object. Template image is usually sub part or sub image of the reference image. Dimension of Template image is smaller than Reference Image.

SAMPLE RESUME WITH CERTIFICATION



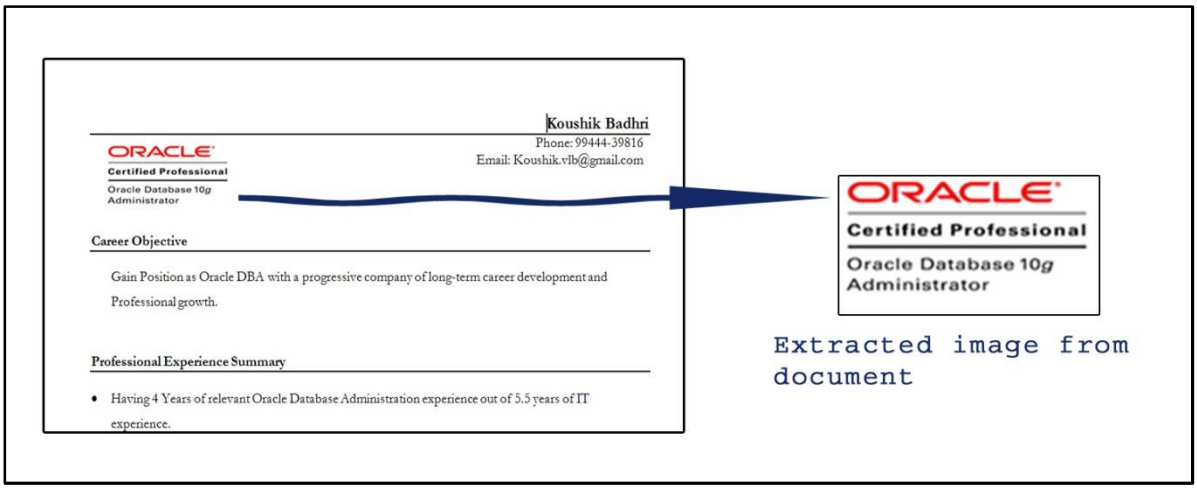


Figure 1,2(Sample resume with certification).(Extracted image for processing and Figure from sample resumes)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

III. PROPOSED WORKCHART

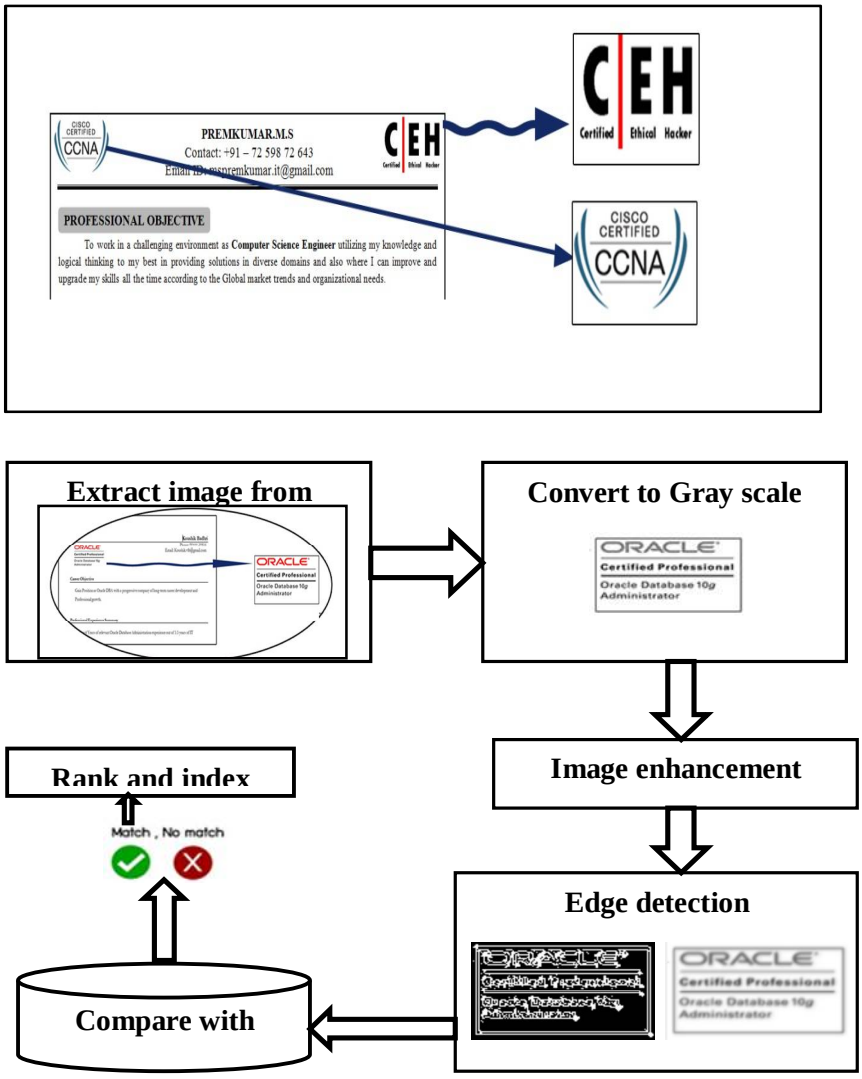


Figure 3: Work chart diagram

The proposed method comprises of few processing namely, extraction of image from the document, Gray scale conversion, enhancement of image quality, detecting the edges and then comparing the input image with the template to rank and index.

(a) Extraction of Images

The images are extracted from the resume document for processing and the images are indexed to the resume ID to store the rank or weight age allotted according to the certified images present in the resume. The resume may contain single or multiple images and all these images are extracted for processing. ***Gray scale conversion***

Since the images extracted will be in RGB color, it is necessary to convert the RGB images into gray scale to make the processing faster, since the gray scale stores the color values as the intensity at one level.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

(a) This process resizes the images uniformly to a 320 x 240 pixel to speed up the process. To enhance the image quality low pass filtering techniques along with smoothing techniques are employed and the image quality is increased considerably.

(b) Edge detection

This process eliminates noise using Gaussian filter and calculate to find magnitude and orientation of gradient. The point at which maximum value and same direction with gradient is the edge of image. Two threshold values are High threshold and Low threshold, pixel value is larger than High threshold will be assigned to be logic 1 (edge) but pixel value is larger than Low threshold will be assigned as logic 0.

(c) Detection and comparison

The input image is fetched for detection and the images from the template are fetched for comparison. Direct mapping technique is carried out, the template image is mapped with the input image and each pixel is subtracted from the template image and the difference is saved. The difference is compared with the permissible tolerance level and if the difference is within the tolerance, the image is matched and the corresponding weight age is stored for ranking.

EDGE DETECTION PROCESS

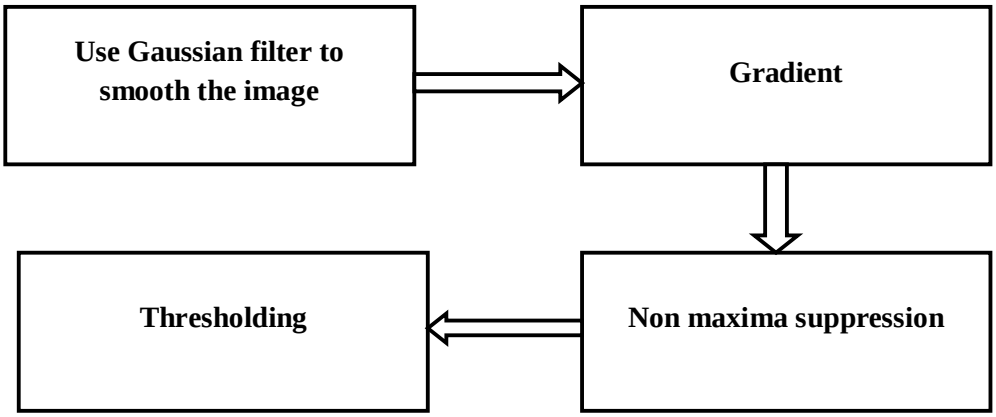


Figure 4: workflow of edge detection

Smoothening the image with Gaussian filter is used to eliminate the noise and the formula for that is

$$S[i, j] = G[i, j, \varsigma] \times I[i, j]$$

Where I[i, j] is the input image, G[i, j, ς] is Gaussian filter, ς is spread of Gaussian, S[i, j] is the smoothened output image.

Gradient of the smoothened image is calculated by

$$M[i, j] = (S[i, j+1] - S[i, j] + S[i+1, j+1] - S[i+1, j]) / 2$$

$$N[i, j] = (S[i, j] - S[i+1, j] + S[i, j+1] - S[i+1, j+1])/2$$

Where M[i, j] is gradient in x direction, N[i, j] gradient in y direction.

Now find the x, y partial directives values by using rectangular to polar conversion and calculate the magnitude and orientation of the gradient.

$Mg[[i, j] = \sqrt{M[i, j]^2 + N[i, j]^2}$ = magnitude

Ori[i, j] = arctan(M[i, j], N[i, j]) = Orientation

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

SAMPLE IMAGE DURING EDGE DETECTION

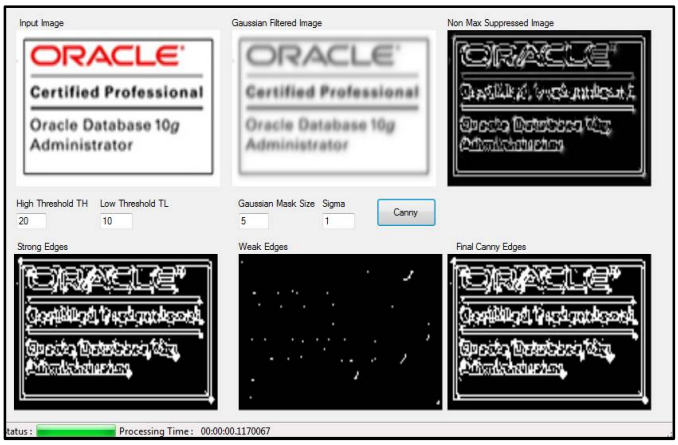


Figure 5: Edge detected sample image

IMAGE TEMPLATE

The image template consists of certification logo along with the weightages for appropriate certifications according to the value of the certification in the corresponding field.

SNO	Certification Image	Weightage
1	ORACLE DATABASE 10g	2.1
2	ORACLE Certified Professional Oracle 11g Administrator	2.3
3	ORACLE 10g Certified Associate	2.0
4	ORACLE Certified Professional Java SE 7 Programmer	2.1

: Sample image template

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

PROPOSED ALGORITHM

Procedure ImageWeightageAllocation(Image I, Imagetemplate Itt)

Input : Image InpImage , ImageTemplate TemImage
Output : Weightage score of image
Begin

1. Load Input image InpImage
2. Load ImageTemplate
3. Convert InpImage to Grayscale
4. Detect Edges
5. While [templateImage ≠ ϕ]
   Do templateImage matching using tolerance
   If matched
   Fetch the weightage value from template and feed it along with the resume ID
   Else
   Fetch next image from template
   End While
6. Return weightage score

END PROCEDURE

Figure 6: ImageWeightageAllocation algorithm

IV. EXPERIMENTAL RESULTS

From the experimental results, various certification images from the resumes are extracted, detected and applied the weightage according to the template and the results procured from the proposed algorithm is up to the satisfactory level. The percentage of mismatch occurred is only 3 to 4 % and the proposed algorithm performed quite well as shown in the figure 7. The canny edge detection technique applied performs reasonably well during edge detection and it enhances the percentage of detection considerably.

Op-

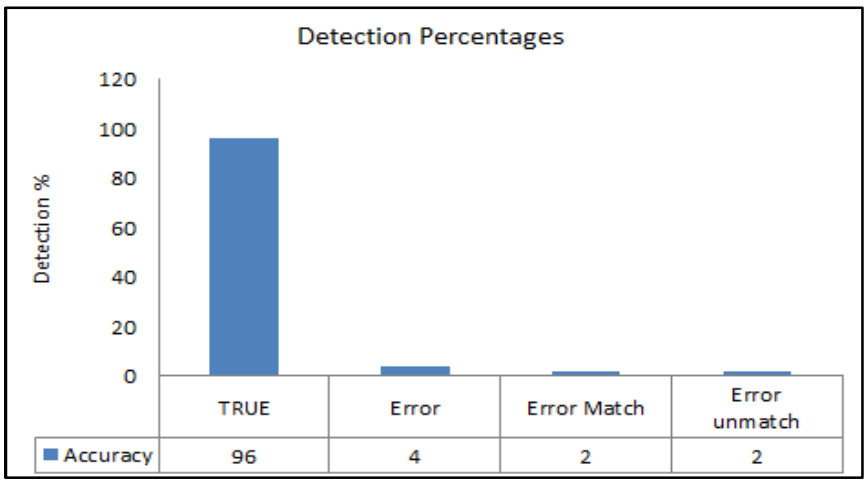


Figure 7: Detection percentage

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

RANKING RESULTS

Resume ID	Certification Weightage
R5	4.4
R7	2.3
R72	2.1
R81	4.2

Table 2: Sample Ranking and Weightage

The resumes will be of unstructured data and importantly some resumes may contain more than one certification and few may not contain any certifications. The weightage calculated here is the sum of the certification values and this helps to provide ranking to the resumes and enhances the recruitment process. This weightage calculated and extracted from the resumes will be added to the experience, skillset, and qualification weightages to rank the resumes. The entire process consists of three phases namely image extraction, image processing and image detection to calculate weightage. From the table 2, it is shown clearly that the weightages calculated from the corresponding resumes are stored in the table for ranking.

V. CONCLUSION

The proposed image matching algorithm implemented uses tolerance value for comparison. From experiments that are given in result section, it has been observed that normally our proposed algorithm detection ratio is far better and faster than many techniques. In future if the higher number of edges is detected, then more features will be selected for detection.The proposed algorithm gives good detection response in case of noisy images and blurred images.

REFERENCES

[1] Clark F. Olson, "Maximum-Likelihood Image Matching" IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 6, June 2002.
[2] Gu Hui, Chen Guangyi, Cao Wenming "Image Match Algorithm Based on Feature Point with Bidirectional Threshold", International Conference on Neural Networks and Brain, Publication Date: 13-15 Oct. 2005, Volume: 3, On page(s): 1464- 1468.
[3]. Chunlei Huo, *Member, IEEE,* Chunhong Pan, Leigang Huo, and Zhixin Zhou" Multilevel SIFT Matching for Large-Size VHR Image Registration , EEE geosciences and remote sensing letters,vol.9,No.2 march 2012.
[4] Rajesh Kumar, Anurag Singh Tomar, An Efficient Image Matching algorithm, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4151 – 4152