Enhancement and Retrieval of Historic Inscription Images

S. Indu[1], Ayush Tomar[1], Aman Raj[1], and Santanu Chaudhury[2]

[1]Delhi Technological University, New Delhi
s.indu@dce.ac.in, ayushtomar@gmail.com, amanraj9992@hotmail.com
[2]Indian Institute of Technology, Delhi
schaudhury@gmail.com

Abstract. In this paper we have presented a technique for enhancement and retrieval of historic inscription images. Inscription images in general have no distinction between the text layer and background layer due to absence of color difference and possess highly correlated signals and noise; pertaining to which retrieval of such images using search based on feature matching returns inaccurate results. Hence, there is a need to first enhance the readability and then binarize the images to create a digital database for retrieval. Our technique provides a suitable method for the same, by separating the text layer from the non-text layer using the proposed cumulants based Blind Source Extraction(BSE) method, and store them in a digital library with their corresponding historic information. These images are retrieved from database using image search based on Bag-of-Words(BoW) method.

1 Introduction

Enhancement, binarization and retrieval of historic inscription images present one of the many challenging issues in image analysis and preservation of digital heritage. Historic inscriptions are an outlook of the past and an essential part of social, economical and scientific studies. However, pertaining to various factors such as environmental change and human intervention, the quality of such inscriptions degrade with the passage of time. Therefore, there is a need to digitize these inscriptions in form of images, which should be free of any noise and unwanted background information and further, store such images in a digital library so that their retrieval with related information could be performed efficiently in real time.

Significant amount of work has been done in the field of digitization of text document images [1][2]. Many of these works are dependent on background light intensity normalization [3] and exploitation of edge information [4]. Generally, the inscriptions are found engraved or projected out of stones or any durable material. Complexity of the background, uncontrolled illumination and minimal difference between foreground (text) and background in camera-held images of such inscriptions pose a challenging problem for text extraction. Direct use of binarization techniques such as Otsu's method do not give good qualitative results due to highly correlated noise present in the images which is not eliminated.

The use of Blind Source Extraction (BSE) and Independent Component Analysis (ICA) [5] is prevalent in the field of digital signal processing; very few applications of the same are found in the field of preservation of our heritage digitally. There are a few ICA-based techniques that have been used for enhancement and digitization of historical inscription images by maximizing text layer information. In the approach as suggested in [6], use of Natural Gradient based flexible ICA (NGFICA) is proposed which minimizes the dependency among the source signals to compute the independent components by considering the slope at each point in the source signals. The authors of [7] suggest a Fast-ICA based method, which aims to enhance the OCR readability of camera-held inscription images.

Image retrieval has been an active field of research. Various techniques involving comparison using local features such as color [8], shape [9] and various other low level features [10] have been proposed but these techniques fail to provide good results under addition of correlated noise or change in scale. There are certain techniques involving comparison of scale invariant features such as SIFT [11] but directly comparing such features is computationally expensive and not realizable in a real time retrieval system as the size of dataset grows. Bag-of-Words (BoW) representation of a images for comparison has been a popular method and has shown excellent results in retrieval of word images [12] and videos [13]. BoW model has been very much suitable for image retrieval from a large database in real time [12].

In our proposed technique, firstly we calculate three independent components i.e. foreground (text-layer), background and middle layer from a linear mixture of unknown sources using cumulant based Blind Source Extraction. After the extraction of three layers, the text information contained in the foreground layer is extracted by further processing which involves local thresholding through Otsu's method [14] , morphological operations and median filter for smoothening purpose and finally a binary image is obtained which is free from any noise or unwanted background information and contain maximum text content. Secondly, these binary images are stored in a database and BoW descriptors based on SIFT features [11] of each image are computed and stored, which are used later for image retrieval from the database that are same or similar to the query image.

2 Methodology

2.1 Blind Source Extraction (BSE)

The ICA based techniques as proposed in [15] [16] has the potential to become computationally very expensive when the number of source signals is at large, let's say an order of 150 or more. Simultaneous Blind Source Extraction (BSE), overcomes this problem by providing a provision and flexibility to extract the desired number of independent components from a set of linear mixtures of large number of statistically independent source signals. The approach is to use a contrast function to handle the third and fourth order cumulants simultaneously to reduce the computational time overhead.

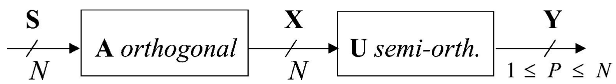


Fig. 1. Signal model of BSE taken from [18]

BSE [15][17] can be explained in a simplest possible way as follows. Let's consider some random sources $s_1, s_2, ..., s_N$ generated by a random process forms the source matrix $\mathbf{S} = [s_1, s_2, s_3, ..., s_N]^T$ as shown in Fig..1 . It is assumed that sources are non-Gaussian with zero mean and have statistical independency. These sources are linearly mixed in a memory-less system represented by a mixing matrix ***A[N X N]*** such that:

$$\mathbf{X{=}AS} \tag{1}$$

Where $\mathbf{X} = [x_1, x_2, x_3, ..., x_N]^T$ are the obtained linear mixtures. It is considered that the mixing matrix **A** is orthogonal and non-singular without much loss of generality. In order to extract $P\,(where\,1 \leq P \leq X)$ number of sources from this mixture matrix **X**, the observations are processed in a linear and memory-less system which is described by a semi-orthogonal de-mixing matrix ***U[P X N]***, such that output of system gives:

$$\mathbf{Y{=}UX} \tag{2}$$

Where $\mathbf{Y} = [y_1, y_2, y_3, ..., y_P]^T$, is the matrix containing the extracted independent components $y_1, y_2, y_3, ..., y_P$ as specified by the user. The semi-orthogonality of matrix **U** is important consideration for having a spatial decorrelation in the outputs as in [9].

2.2 Proposed Method: Enhancement and Binarization of Inscription Images using Cumulants based BSE

The proposed approach uses a contrast functional that captures higher order cumulants, which is maximized by the blind source extraction procedure to calculate the independent components (ICs). The final binarized image is obtained on further analysis of the ICs.

2.2.1 Mixture Acquisition from Various Source Observations

We consider the text containing inscription image as mixture of sources i.e. text layer, semi-text layer and non-text layer. Our goal is to separate the text-layer from the non-text parts using the suggested Simultaneous Blind Source Extraction method based on higher order cumulants. We have refined our proposed technique by using HSV color space rather than the RGB (as used in NGFICA [6] and Fast-ICA [7] based papers) to obtain three samples of observation from the original image.It is considered that the color of each individual source is uniform in the inscription image. The obtained HSV components of Fig. 2 are shown in Figures 3-5, which constitute $\mathbf{X}$ as per equation (2). As proposed in [19], HSV without filtering is better than RGB for ICA algorithms. Data distribution pattern in a HSV model possess strong deviation of the data which makes it an excellent candidate for independent component analysis and the observed results are more comprehensive than RGB model.

Fig. 2. Inscription image taken from mudgal fort gateway Karnatka

Fig. 3. Hue (H) component of original image

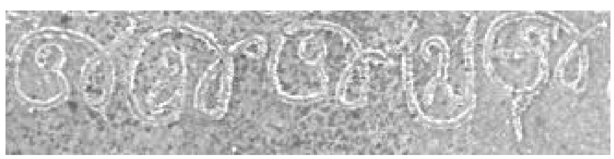


Fig. 4. Saturation (S) component of original image

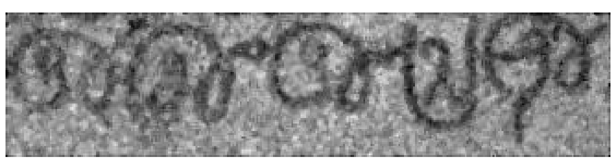


Fig. 5. Value (V) component of original image

2.2.2 Extraction of Independent Components

To find the independent components, Huber [20] suggested finding a contrast functional $\psi(\textbf{.})$ that maps each p.d.f. of a normalized random variable Y_i to real index $\psi(Y_i)$, where $\mathbf{Y}$ is the vector of the output of estimated sources.

Proper optimization of $\psi(Y_i)$ will lead to extraction of one of the independent components.

In general, the blind source separation of the whole set of sources is obtained by maximizing the function below:

$$\max_{U} \sum_{i=1}^{N} \psi(Y_i) \text{ subjected to } Cov(\mathbf{Y}) = I_N \tag{3}$$

I_N is identity matrix with order N (number of mixed sources) and $\mathbf{Y}$ is the vector of estimated sources [17].

In this paper, independent sources are calculated subject to a cumulants based index, which is a measure of the degree on non-gaussianity or the amount of structure present in the outputs. These indexes have their maxima at the extraction of one of the independent sources. The contrast function is a mathematical function which simultaneously handles these high order cumulants [21]. Thus blind extraction of one of the sources is obtained after solving the following maximization problem:

$$\max_{U} \psi(Y_1)\, subjected\, to\, Cov(Y_1) = 1 \tag{4}$$

One generalized form of cumulant index for a random variable is given by [17]:

$$\psi_C um(Y_1) = \sum_{r>2} \omega_r^{\backslash}. \mid C_{Y_1}^{r} \mid^{\alpha_r} \tag{5}$$

where r is order of cumulants and $\alpha_r \geq 1$, $\mid C_{Y_1}^{r} \mid$ denotes modulo of *rth* order auto-cumulant, *Cum(*Y_1 *X r)* and $\omega_r^{\backslash} = \frac{\omega_r}{r\alpha_r}$ are scaled or normalized as non-negative weighting factors. The low order cumulants having *r=1* and *r=2* are excluded from indexing due to normalization constraint.

In case of blind source extraction, we express the equation (3) in terms of cumulant index to calculate P out of N total sources, the corresponding cumulant contrast function with largest cumulant index is given by:

$$\psi_C um(\mathbf{Y}) = \sum_{i=1}^{P} \sum_{r>2} \omega_r^{\backslash}. \mid C_{Y_1}^{r} \mid^{\alpha_r} \; subject\, to\, Cov(\mathbf{Y}) = I_P \tag{6}$$

The global maxima of this contrast function correspond to the extraction of first P sources from the mixture. Here, in our case *P=N=3*. This way, based on the cumulants index, three estimated Sources (ICs) are calculated from the observed mixtures of the sources (H,S,V components of the original inscription image) as shown in Figures 6-8 .

2.2.3 Enhancement of the Text Layer

The estimated sources(ICs) are labeled text, semi-text and non-text layer as per their resemblance with the original text on the inscription image. The text layer is selected and is used for further processing. This layer is then binarized calculating a suitable local threshold level as per Otsu's method proposed in [14]. Further the post-processing morphological operations dilation and erosion are used to enhance the readability of text in the text layer followed by a suitable median filter. The final digitized binary image is shown in Fig. 9. Our method provides better qualitative results than as proposed in existing fast-ICA [7] based technique in which thresholding is done after combining all the ICs which leads to adding up of noise also.

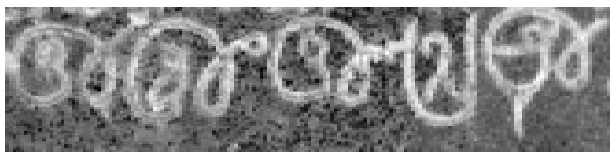


Fig. 6. Text layer after execution of algorithm

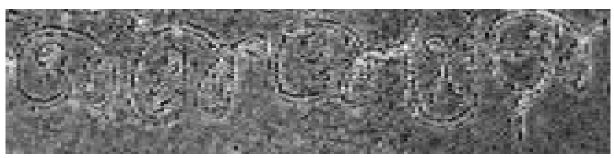


Fig. 7. Semi-text layer after execution of algorithm

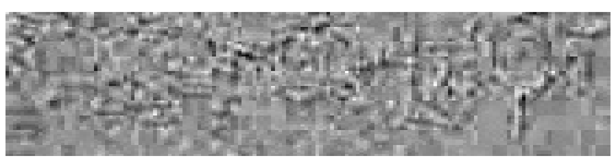


Fig. 8. Non-text layer after execution of algorithm

Fig. 9. Final digitized image

2.2.4 Storage of Digitized Image and its Related Information in a Database

We apply the suggested technique for enhancement and binarization on every inscription image taken from a site and obtain the final image similar to as shown in Fig. 9. These final images are in binary form and contain maximum text content. All the non-text information such as noise and background is removed so that final images contain maximum text information, in this way retrieval of such images from a digital database is more accurate due to ease in matching of only important features. As the images contain maximum text information in form of a binary image, these can be used for an OCR, if available. Any information related to the binary image such as history, era, location, civilization etc. are also attached with the image and stored in the digital database.

Creating a digital database using binary images having maximum text content has several advantages as only the features having text information remain in the images, which provide much better results while matching and retrieval giving a better matching score as unwanted information such as noise and background are removed. OCR readability of such images is more than an unprocessed inscription image. Moreover, having a digital database of binary images has less storage space overhead than unprocessed historic images.

2.3 Extraction and Preparation of Bag-of-Words Database

Bag-of-Words was originally used in text classification and retrieval which has been to extended to use in retrieval of images and videos [12] [13] and has shown excellent results. An image is a set of unordered visual features. A bag of words is a sparse vector which contains occurrence counts of local image features; in other words, BoW is a sparse histogram over a vocabulary where vocabulary is a set of discrete visual features such as Scale Invariant Feature Transform (SIFT) [11] descriptors. SIFT features have a continuous space and are clustered for a fixed number of bags which are trained with clustered feature descriptors using K-means algorithm. Finally an image is represented as a normalized histogram of the clustered SIFT features having vector length k, where k is the vocabulary size; this histogram is BoW descriptor which can be used to match or classify the image. Fig. 10 describes the working of BoW.

2.3.1 Obtaining a Set of Bag-of-Words

We take the set of digitized images from the database and obtain SIFT [11] descriptors based on SIFT feature points that are extracted from each of the image in the database. Further, these feature descriptors are clustered using K-means algorithm into a defined number of bags and are trained, thus descretizing the descriptor space [12] [13]. These clustered descriptors are bag of features and functions as a vocabulary in later stages of retrieval.

2.3.2 Obtaining BoW Descriptor of Images in the Database

We take all the digitized images from the database and again find their corresponding local SIFT features. SIFT features are chosen being invariant to image

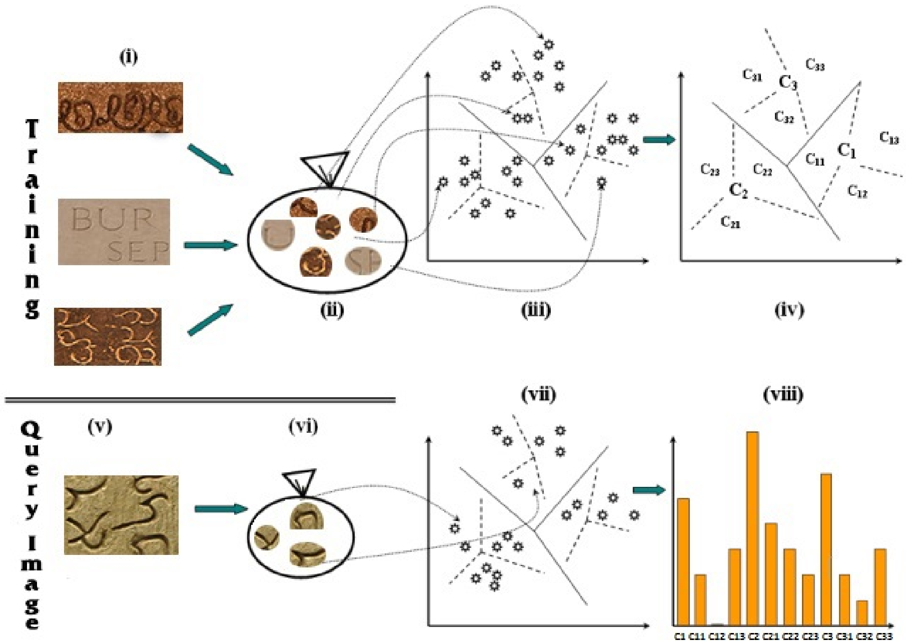


Fig. 10. A Description of working of bag-of-words

scaling, translation, rotation and illumination changes [11]. Based on these feature points we obtain SIFT descriptors. The feature descriptors so obtained are matched with the vocabulary as proposed in sub-section 2.3.1 to quantize the extracted features by assigning these features to the closest clustered centroid [12]. The images are finally represented as a normalized histogram of quantized SIFT features, which is a count of each word (local visual features) belonging to its bag (clustered features). This histogram is a BoW descriptor, which is computed for each digitized image in the database and stored in a descriptor dictionary (.yml file) on a stable media.

2.4 Retrieval of Image and its Corresponding Information using BoW Descriptors

In the final step of retrieval, we take a query image which is an image of a historic inscription. The query image is processed using the method as proposed in sub-sections 2.2.1 - 2.2.3. On the obtained digitized image, BoW descriptors are computed based on SIFT features [11]. BoW descriptor of this image is matched with existing BoW descriptors which are stored in a descriptor dictionary as proposed in the previous subsection. This matching is performed using FLANN (which gives approximate euclidean distance between the descriptors) to compute a dissimilarity score with all the images in the database based on number of "good matches" found. Good matches depend on the euclidean distance between nearest neighbors and represented in terms of dissimilarity score between images. The best image match based on the dissimilarity score was returned and all related information of the matched image in the database was also displayed. Similar images had a dissimilarity score very close to zero.

3 Results

We implemented the proposed algorithm using the OpenCV library for C++ on Windows platform. The results were obtained using a set of 200 historical inscription images based on a number of query images taking k to be 100, where k is the vector length of clustered SIFT features. The statistics of first fifty query images belonging to each category i.e. unprocessed and binarized text using a raw and processed database respectively are shown in Table 1 with their outputs in Figures 11-12. The system is designed such that it returns a best possible match with input query image based on the least dissimilarity score. The results show the robustness and accuracy of the proposed method in retrieval of corresponding similar images with their attached historic information from the digital library. The results are shown in Figures 13-15 , which are accurate even in case of distorted or rotated query images. Fig. 16 shows the result when query image is not present in the database, in this case image having the lowest dissimilarity score is retrieved.

Query Image	Median Dissimilarity Score	% of Correct Matches
Unprocessed (Raw)	0.231	58%
Binarized Text	0.046	92%

Table 1. Statistics of Result

Query image	Retrived Image	Dissimilarity Score
		0.27

Fig. 11. Result of an Unprocessed Image from a Unprocessed Image Database

Query image	Retrived Image	Dissimilarity Score
		0.0192

Fig. 12. Result of a binarized Image from a Processed Image Database

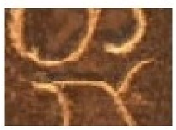
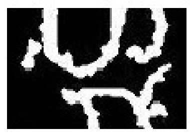
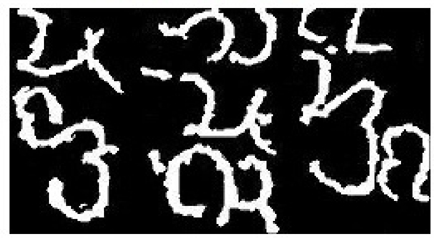
<table>
<tr><th>Query image</th><th>Binarized image</th><th>Retrieved image</th></tr>
<tr><td></td><td></td><td></td></tr>
<tr><th colspan="2">Dissimilarity score</th><th>Retrieved information</th></tr>
<tr><td colspan="2">0.078</td><td>Image taken from Badami caves of Karnataka</td></tr>
</table>

Fig. 13. Results on query image - I

4 Conclusion

In this paper a new robust and efficient technique to enhance, binarize and retrieve historic inscription images and their related information from a database is put forward. The proposed method of enhancement and binarization is found excellent for removal of correlated noise and unwanted background from historic inscription images to extract maximum text content. Further, using a digital library of these images, a suitable method was proposed for retrieval of corresponding inscription image and its related information of historical importance using Bag-of-Words method; the technique was found to be excellent in terms of accuracy of matching based on the obtained results.

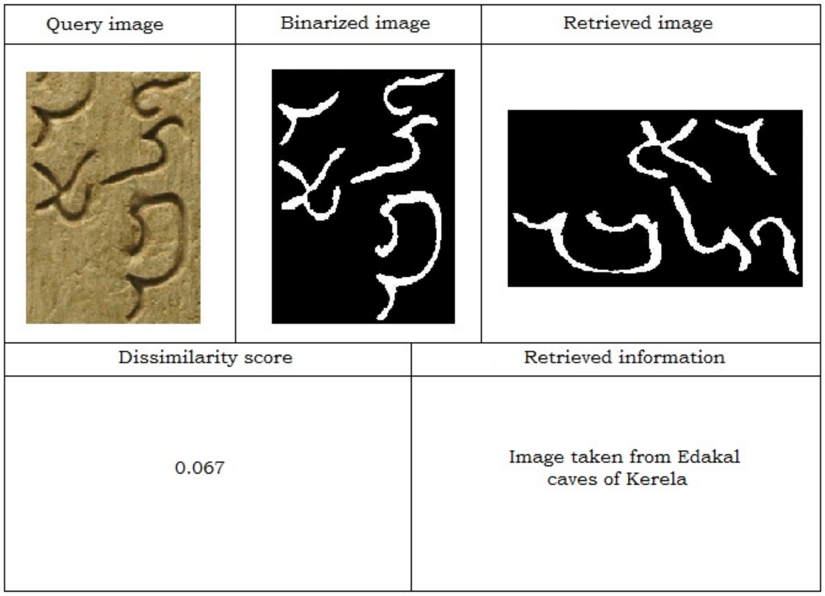


Fig. 14. Results on query image - II

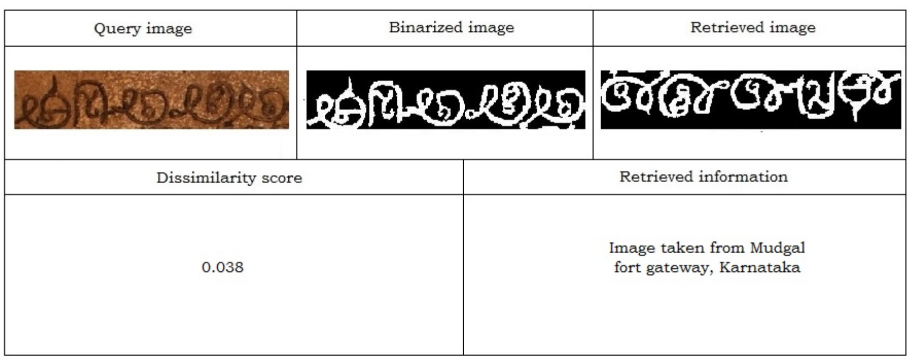
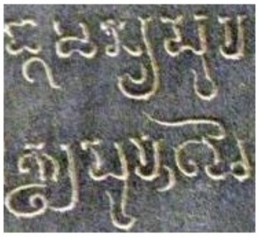
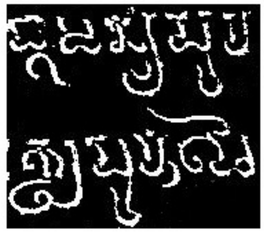
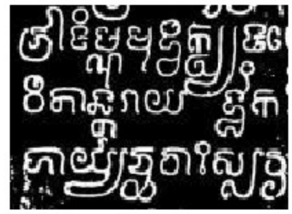


Fig. 15. Results on query image - III

5 Future Work

An extension of this technique by using SURF and ORB-Features over SIFT features to check the accuracy and measure the computational overhead is planned for the future work on a larger database. Detailed comparison of proposed enhancement and retrieval techniques with existing techniques is also planned. Further, a web based tool of the technique will be conceptualized in which researchers can add their images and its related information to the existing database to enhance the matching performance of the retrieval process.

Query image	Binarized image	Retrieved image
		

Dissimilarity score	Retrieved information
0.183	Khmer inscription

Fig. 16. Results on query image - IV

References

1. Doermann, D., Liang, J., Li, H.: Progress in camera-based document image analysis. In: Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on, IEEE (2003) 606–616
2. Wolf, C., Jolion, J., Chassaing, F.: Text localization, enhancement and binarization in multimedia documents. In: Pattern Recognition, 2002. Proceedings. 16th International Conference on. Volume 2., IEEE (2002) 1037–1040
3. Shi, Z., Govindaraju, V.: Historical document image enhancement using background light intensity normalization. In: Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. Volume 1., IEEE (2004) 473–476
4. Epshtein, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE (2010) 2963–2970
5. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis. Volume 46. John Wiley & Sons (2004)
6. Sreedevi, I., Pandey, R., Jayanthi, N., Bhola, G., Chaudhury, S.: Ngfica based digitization of historic inscription images. International Scholarly Research Notices **2013** (2013)
7. Garain, U., Jain, A., Maity, A., Chanda, B.: Machine reading of camera-held low quality text images: an ica-based image enhancement approach for improving ocr accuracy. In: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE (2008) 1–4
8. Swain, M.J., Ballard, D.H.: Color indexing. International journal of computer vision **7** (1991) 11–32
9. Rui, Y., She, A.C., Huang, T.S.: Modified fourier descriptors for shape representation-a practical approach. In: Proc of First International Workshop on Image Databases and Multi Media Search. (1996) 22–23

10. Rui, Y., Huang, T.S., Chang, S.F.: Image retrieval: Current techniques, promising directions, and open issues. Journal of visual communication and image representation **10** (1999) 39–62
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60** (2004) 91–110
12. Shekhar, R., Jawahar, C.: Word image retrieval using bag of visual words. In: Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on, IEEE (2012) 297–301
13. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE (2003) 1470–1477
14. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11** (1975) 23–27
15. Tonazzini, A., Bedini, L., Salerno, E.: Independent component analysis for document restoration. Document Analysis and Recognition **7** (2004) 17–27
16. Cichocki, A., Amari, S.i.: Adaptive blind signal and image processing: learning algorithms and applications. Volume 1. John Wiley & Sons (2002)
17. Cruces-Alvarez, S.A., Cichocki, A., Amari, S.i.: From blind signal extraction to blind instantaneous signal separation: criteria, algorithms, and stability. Neural Networks, IEEE Transactions on **15** (2004) 859–873
18. Cruces-Alvarez, S.A., Cichocki, A., Amari, S.I.: On a new blind signal extraction algorithm: different criteria and stability analysis. Signal Processing Letters, IEEE **9** (2002) 233–236
19. Katsumata, N., Matsuyama, Y.: Database retrieval for similar images using ica and pca bases. Engineering Applications of Artificial Intelligence **18** (2005) 705–717
20. Huber, P.J.: Projection pursuit. The annals of Statistics (1985) 435–475
21. Blaschke, T., Wiskott, L.: Cubica: Independent component analysis by simultaneous third-and fourth-order cumulant diagonalization. Signal Processing, IEEE Transactions on **52** (2004) 1250–1256