Digitization of Historic Inscription Images using Cumulants based Simultaneous Blind Source Extraction

N. Jayanthi*
Delhi Technological University
New Delhi, India
njayanthi@dce.ac.in

Ayush Tomar
Delhi Technological University
New Delhi, India
ayushtomar@gmail.com

Aman Raj
Delhi Technological University
New Delhi, India
amanraj9993@gmail.com

S. Indu
Delhi Technological University
New Delhi, India
s.indu@dce.ac.in

Santanu Chaudhury
Indian Institute of Technology
New Delhi, India
schaudhury@gmail.com

ABSTRACT

In this paper a novel method to address the problem of enhancement and binarization of historic inscription images is presented. 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. The proposed technique provides a suitable method to separate the text layer from the historic inscription images by considering the problem as blind source separation which aims to calculate the independent components from a linear mixture of source signals, by maximizing a contrast function based on higher order cumulants. Further, the results are compared with existing ICA based techniques like NGFICA and Fast-ICA.

Keywords

Blind Source Extraction, ICA, Inscription Images, Binarization

1. INTRODUCTION

Enhancement and binarization 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 degrades with the passage of time. Thus there is a need to enhance these historic inscriptions and binarize them in a way such that OCR readability is maintained for use in future.

Significant amount of work has been done in the field of text extraction from document images and increasing the OCR Readability [5][17]. Many of these works are dependent on background light intensity normalization [14] and exploitation of edge information [6] . Generally, the inscriptions are found engraved or projected out of stones or any durable material. Uncontrolled illumination, complexity of the background and minimal difference between foreground (text) and background in camera-held images of such inscriptions pose a challenging problem for text extraction.

The use of blind source extraction and independent component analysis (ICA) [9] 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 binarization of historical inscription images. In the approach as suggested in [15], 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.

In the proposed technique, the problem of calculating three independent components i.e. foreground (text-layer), background and middle layer from a linear mixture of unknown sources can be considered as an independent component analysis [16] of the sources, which is a subset of blind source extraction problem. 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 [11] , morphological operations and median filter for smoothening purpose.

The sub-section 2.1 in the paper explains the independent component analysis and blind source extraction; subsection 2.2(1-3) describes the proposed technique; further, the results of our method are compared with those of existing NGFICA and Fast-ICA based techniques in subsection 2.3.

2. METHODOLOGY

2.1 Blind Source Extraction

The ICA based techniques as proposed in [16] [2] 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),

*Corresponding author

ICVGIP '14, December 14-18, 2014, Bangalore, India
Copyright 2014 ACM 978-1-4503-3061-9/14/12 ...$15.00
http://dx.doi.org/10.1145/2683483.2683534
.

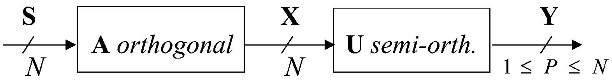

Figure 1: Signal model of BSE taken from [3]

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.

BSE [16][4] 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 in-dependency. 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 $\mathbf{A}$ 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 $\mathbf{X}$, 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 $\mathbf{U}$ is important consideration for having a spatial decorrelation in the outputs as in [9].

2.2 Proposed Method

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 [15] 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 Figure 2(a) are shown in Figures 2(b)-(d), which constitute $\mathbf{X}$ as per equation (2).

As proposed in [10], 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.

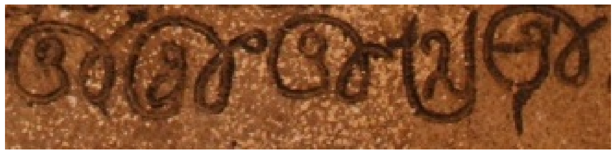

Figure 2a: Inscription image from Mudgal fort in Karnataka

Figure 2b: Hue (H) component of original image

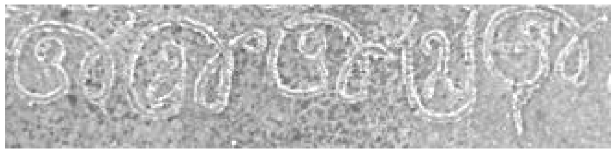

Figure 2c: Saturation (S) component of original image

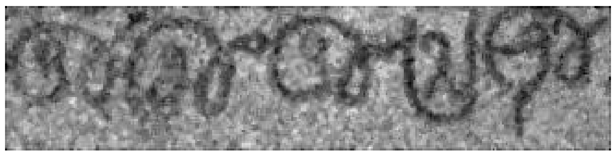

Figure 2d: Value (V) component of original image

2.2.2 Extraction of independent components

To find the independent components, Huber [8] suggested finding a contrast functional $\psi(.)$ 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 [4].

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 [1]. 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 [4]:

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

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:

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

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 3(a)-(c) .

2.2.3 Character Extraction from 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 [11]. 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 binarized image is shown in Fig. 3(d).

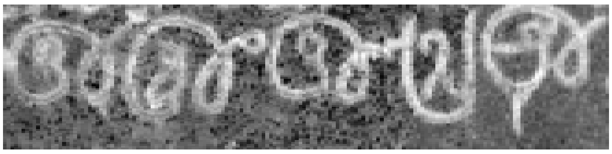

Figure 3a: Text layer after execution of algorithm

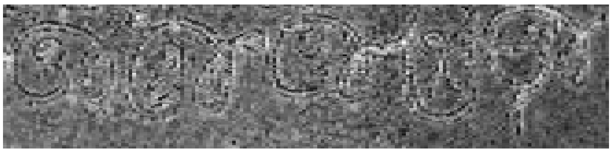

Figure 3b: Semi-text layer after execution of algorithm

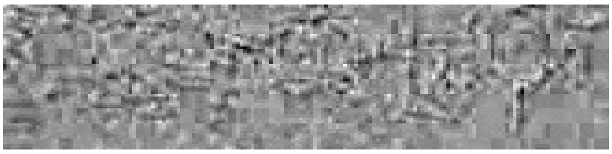

Figure 3c: Non-text layer after execution of algorithm

Figure 3d: Final binarized image

2.3 Performance analysis with NGFICA and Fast-ICA

We have compared the results of our method with the proposed NGFICA based enhancement of historical images [15] and fast-ICA method [7] and results are found to be much better in terms of accuracy and readability of processed text.

The method proposed in [7], based on Fast-ICA analysis on RGB components of the inscription image , combines all the three binarized independent layers (i.e. text, non-text and semi-text layer) of output, which also adds up the noise in final image resulting in very low readability.

The authors of [15] used an NGFICA based method, also based on RGB model as in [7] to calculate the three independent layers. Their method used a threshold criteria determined by local and global average of pixel values of three layers to select text layer. The layer having local average with maximum offset from the global average is used for binarization and extraction of text.

Our method is different from [15],[7] in terms of mixture acquisition from various source observations as it uses H, S and V layers as the observed sources and gives completely de-correlated output. A single clear layer containing text is obtained as evident from Figure 3(d) which is used for further processing. This way our results are better in terms of readability and accuracy, shown in Table 2; as there is no added up noise from the three independent layers of output as seen in [7] and error in selecting the text layer as observed in [15]. Further, it has been observed that the proposed algorithm handles simultaneously the third and fourth order cumulants and as no parameter adjustment is required, thus it has an advantage of much faster result processing in case of data with high dimensionality.

For evaluation and comparison of techniques, we have used similar binarization evaluation performance measures as in used in DIBCO and H-DIBCO contests [12] [13] such as F-Measure and PSNR.

A. **F-Measure**

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{7}$$

where $Recall = \dfrac{TP}{TP + FN}$ and $Precision = \dfrac{TP}{TP + FP}$

TP, FP, FN denote the True positive, False positive and False Negative values, respectively.

B. **PSNR**

$$PSNR = 10log(\frac{C^2}{MSE}) \tag{8}$$

where, $MSE = \dfrac{\sum\limits_{x=1}^{M} \sum\limits_{y=1}^{N} (I(x,y) - I'(x,y))^2}{MN}$

C is the difference between foreground and background. Higher the value of PSNR more is the similarity between two images.

The evaluation was performed using F-Measure and PSNR values described above. Table 1 shows performance of each technique for each encountered measure. Evaluation was performed using five inscription images (Figures 2,5-8) based on manually created ground truth images. Evaluation based on these measures clearly established our algorithm as a better technique for text enhancement compared to other existing ICA based methods.

Technique	**F-Measure (%)**	**PSNR**
Our method	79.02	10.67
NGFICA based method	63.36	9.95
Fast-ICA based method	57.99	8.21

Table 1: Evaluation results based on measures used

Table 2 shows the comparison of our technique with existing techniques based on readability of processed text and time taken by technique to process the images. Analysis of readability as mentioned in Table 2 was performed by processing a variety of inscription images containing a total of 501 English words through all the three techniques and passing the output to the OCR to determine the OCR readability. It was found that 77.60% of words were correctly recognized using our method. 63.20% recognition was achieved using NGFICA based method whereas 14.60% recognition of words was obtained using Fast-ICA based method.
Analysis of time taken was done by running our proposed method in comparison to Fast-ICA and NGFICA based techniques on our dataset containing a variety of historical inscription images of different resolutions on Intel i5 processor clocked at 1.8 GHz. The observations for time taken mentioned in Table 2 are based on careful observation of timing analysis on images used for analysis in Table 1.

Method	**Readability of Processed Text**	**Time Taken**
Our method	Very High	Medium
NGFICA based method	High	Very High
Fast-ICA based method	Very Low	Low

Table 2: Comparison with existing methods

Figure 4a: Output image after Fast-ICA based method

Figure 4b: Ouput after NGFICA based method

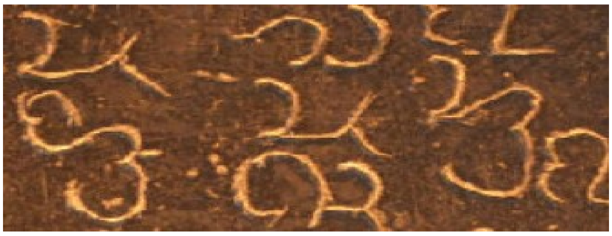

Figure 5a: Inscription image taken from the Badami caves of Karnataka

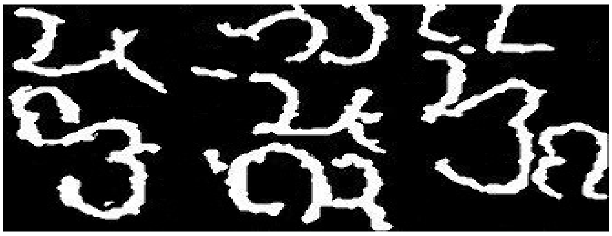

Figure 5b: Output image after proposed method

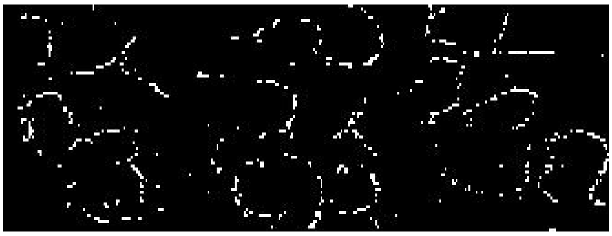

Figure 5c: Output image after Fast-ICA based method

Figure 5d: Output image after NGFICA based method

A number of inscription images taken from different sources were processed through the method proposed in this paper and their result are shown below with respective source image:

Figure 6a: Inscription Image from Edakal cave in Kerala

Figure 6b: Output of proposed method

Figure 7a: Roman Inscription from Mausoleum of Caecilia Metella

Figure 7b: Output of proposed method

Figure 8a: Image taken from India Gate

Figure 8b: Output of proposed method

3. CONCLUSION

A novel technique for enhancing and binarizing historic inscription images has been put forward based on the results obtained by processing various images. The technique outperforms several existing techniques and clearly establishes cumulants based blind source extraction as a significant improvement in this research field. Moreover, the technique shows a much faster result processing in case of data with high dimensionality as it is based on simultaneous handling of higher order cumulants. Later on, extension of this technique into enhancement of images of ancient coins, seals and sculptures is planned.

4. REFERENCES

[1] T. Blaschke and L. Wiskott. Cubica: Independent component analysis by simultaneous third-and fourth-order cumulant diagonalization. *Signal Processing, IEEE Transactions on*, 52(5):1250–1256, 2004.

[2] A. Cichocki and S.-i. Amari. *Adaptive blind signal and image processing: learning algorithms and applications*, volume 1. John Wiley & Sons, 2002.

[3] S. A. Cruces-Alvarez, A. Cichocki, and S.-I. Amari. On a new blind signal extraction algorithm: different criteria and stability analysis. *Signal Processing Letters, IEEE*, 9(8):233–236, 2002.

[4] S. A. Cruces-Alvarez, A. Cichocki, and S.-i. Amari. From blind signal extraction to blind instantaneous signal separation: criteria, algorithms, and stability. *Neural Networks, IEEE Transactions on*, 15(4):859–873, 2004.

[5] D. Doermann, J. Liang, and H. Li. Progress in camera-based document image analysis. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pages 606–616. IEEE, 2003.

[6] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2963–2970. IEEE, 2010.

[7] U. Garain, A. Jain, A. Maity, and B. Chanda. 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*, pages 1–4. IEEE, 2008.

[8] P. J. Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985.

[9] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[10] N. Katsumata and Y. Matsuyama. Database retrieval for similar images using ica and pca bases. *Engineering Applications of Artificial Intelligence*, 18(6):705–717, 2005.

[11] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[12] I. Pratikakis, B. Gatos, and K. Ntirogiannis. H-dibco 2010-handwritten document image binarization competition. In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, pages 727–732. IEEE, 2010.

[13] I. Pratikakis, B. Gatos, and K. Ntirogiannis. Icdar 2013 document image binarization contest (dibco 2013). In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1471–1476. IEEE, 2013.

[14] Z. Shi and V. Govindaraju. Historical document image enhancement using background light intensity normalization. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 1, pages 473–476. IEEE, 2004.

[15] I. Sreedevi, R. Pandey, N. Jayanthi, G. Bhola, and S. Chaudhury. Ngfica based digitization of historic inscription images. *International Scholarly Research Notices*, 2013, 2013.

[16] A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *Document Analysis and Recognition*, 7(1):17–27, 2004.

[17] C. Wolf, J. Jolion, and F. Chassaing. Text localization, enhancement and binarization in multimedia documents. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 1037–1040. IEEE, 2002.