Restoration of Degraded Images for Text Detection and Recognition

Sayali R. Joshi
Department of Computer Engineering, MMCOE.
Pune, Maharashtra, India.

Sankirti S. Shiravale
Department of Computer Engineering, MMCOE.
Pune, Maharashtra, India.

ABSTRACT
The task of text detection natural scene images is very challenging due to the complex background and unpredictable text appearances in the image. Apart from the background and the structure of the text, unpredictability also lies in the image capturing quality. These issues include noise, orientation, low exposure, blurring, and other kinds of degradations.

It is therefore necessary to first restore the target text in the image in order to ensure robust text detection and recognition. This research focuses on removing a maximum number of degradation factors from a natural scene image containing text such that the detection and recognition of the text present in that image becomes very easy. Text Specific Dictionaries will be used in order to restore the text in the images. The sparse representation method is selected with an aim to apply techniques such as denoising, deblurring, sharpening and implementing other forms of enhancement in a single text image restoration system.

General Terms
Image restoration, Image enhancement, Text Detection, Text recognition, Sparse representations, Dictionaries.

Keywords
Text- Specific Dictionary, Natural Scene Dictionary, Image restoration, Image enhancement, Text Detection, Text recognition, Sparse representations, Dictionaries.

1. INTRODUCTION
Text detection from image is one of the important tasks for image processing and computer vision. Image can have needful information and this information is important for fully understanding of the image. Text detection is useful in many content-based image and video applications, such as content-based web image search, video information retrieval, and mobile based text analysis and recognition. As in recent years, image, as the visual basis for perceiving the world, is the key media for information acquisition, expression and transmission.

The text in natural scene images has to be robustly detected before being recognized and retrieved. Many problems need to be solved in order to read text in natural images including text localization, character and word segmentation, recognition, integration of language models and context, etc. [8]. Text detection and localization in natural scene images is challenging due to its complex background, and variations of font, size, color and orientation, low exposure, blurring, masking, etc. To solve these problems, it is necessary to first restore the images to a level at which the text area can be easily detected and text can identified and recognized. Figure 1 shows some sample scene images from different areas.

Figure 1: Sample Scene Images

Many methods for text detection from scene images have been proposed over the past years by various authors. These methods are based on connected components, edges, colors, combination of edges and colors, textures, corners, semi-automatic ground truth generation, strokes, etc. However, these methods are not efficient when dealing with images containing dense information. Therefore we propose an approach that overcomes the limitations of these methods by using learned dictionaries with sparse representation.

In the proposed method, two sequences of learned dictionaries for the text and non-text parts respectively will be empirically constructed first. Then, the sparse representations of all different sizes and non-overlapped document patches in these learned dictionaries will be computed. Based on these representations, each patch can be classified into the text or non-text category by comparing its reconstruction errors. Same sized patches in one category will be then merged together to define the corresponding text or non-text layers which will be combined to create a final text/non-text layer. Finally, in a post-processing step, text regions will further be filtered out by using some learned thresholds.

2. RELATED WORK
Image restoration techniques mainly take into consideration the noise, blur, illumination problems, etc. Most of the work on restoration has been done on generic images; text specific image restoration is less explored. We will divide our survey into parts and discuss each restoration method separately.

2.1 Image Denoising

Numerous methods for denoising generic images have been studied. Patch- based methods are popular among those. Jian Sun et al. [2] apply whole image iterative shrinkage by performing patch group shrinkage and whole image aggregation. Jing Jin [13] et al. focus on noise repression and edge preservation using wavelet domain. Zuo [20] et al. denoise images by enforcing gradient histogram of denoised image to be close to a reference gradient histogram of original image. Their focus is on preservation of texture appearances. Correlated databases are used in [5],[19] to eliminate the need to guess priors for denoising. Huanjing Yue et al. [5] explore both internal and external correlations of a noisy image with the help of web images to perform denoising. Enming Luo et al. [19] use targeted databases to determine basis function of optimal filter by group sparsity and spectral co-efficient of optimal filter by localized priors. Mayank Tiwari and Bhupendra Gupta [12] use spatial gradient based bilateral filter and minimum MSE filtering which does not depend on prior information about amount of noise present in the image. Joint denoising and contrast enhancement is performed by Xianming Liu et al. [11] using Laplacian Operator. Poor lighting conditions are considered in this work but blurring issue is not considered in these works which concentrate on only noise reduction.

2.2 Image Deblurring

Non- uniform deblurring is used in [29],[21],[22],[16] though different methods have been explored in each work. Oliver Whyte et al. [29] used Richardson- Lucy (RL) Algorithm for blind deblurring. They used noisy/ blurry image pairs in their work to remove blur caused by camera shake. C. Paramanand and A. N. Rajagopalan [21] performed non- uniform motion deblurring for bilayer scenes exploiting the burst mode feature of camera. They used the Transform Spread function (TSF) for this purpose. Xin Yu et al.[16] used Total Variation (TV) Regularization to obtain blur kernel and performed patch-wise deblurring. Karteek Alahari et al. [22] performed joint image segmentation and deblurring under defocus and linear motion blur. Blind Deblurring is performed in [23] and [14]. In all the above works the main focus is on deblurring and the systems are sensitive to noise.

2.3 Image Restoration

Image restoration is attempted in [10],[3],[24],[17],[25] considering multiple degradation aspects of an image. Xiaoyong Shen et al. [10] captured color, infra-red, and flash images in different fields and employed them to effectively eliminate noise and visual artefacts. Automatic contrast enhancement by preserving image saliency is done in [3] by exploiting the fact that image saliency is sensitive to noise injection but immune to contrast enhancement. A wide range of algorithms like split Bregman, algebraic multi-grid, Krylov acceleration, are used to convert nonlinear TV model into 3 linear systems for image deblurring and denoising in [24]. Image inpainting problem is solved in [17] based on wavelet frame and weighted sparse representation model. In [25] the relationship of blur and sharp patches and smoothing priors are learned simultaneously from image in multiple scales. But this approach is not suitable for deblurring under sparse kernels.

2.4 Sparse Representations for image Restoration

The use of sparse representations and dictionaries for image restoration are explored in [6],[32],[30],[4],[1]. V. Krithika [6] proposed construction of dictionary by PCA or SSC method and use of Group Sparsity. This approach is used for deblurring, denoising and inpainting, but separately. Julien Mairal et al. [32],[30] learned multiscale sparse representations for color images and video with over-complete dictionaries for denoising and inpainting in [32] and exploited self- similarities of natural images for denoising and demosaicking in [30]. Data dependent denoising procedure using patch- based denoising algorithm is performed with a targeted database instead of a generic one in [4]. Jing Han et al. [1] used sparsity based method preserving local sparse structure; where, invariability of these structures of pure signals in low light level images is exploited.

2.5 Text Image Restoration

We will now discuss work on text specific image restoration. Seeri et al. [9] developed a technique for multilingual text localization using Haar wavelet, edge features, K - means clustering, fuzzy classification and threshold concepts. Pujar et al. [8] used log Gabor DWT wavelet with SVM for text extraction from road signs and achieved an accuracy of 91.14%. Thillou et al. [33] developed a camera based system able to capture characters from photo image, making recognition, and then giving speech output. They used Gabor filters and K-means for this system. Lee et al [31] developed a system based on text stroke width information. They used morphological filters to remove salt and pepper noise.

2.6 Use of Sparse Representations for Text Restoration

Recent works in text restoration have explored the benefits of sparse representations and dictionaries. Vijay Kumar et al. [26] exploited the property of text that there are similar strokes, curves and edges for different characters. They applied their system for flips, blur, cuts and texture blending. Dong Zang [18] used Histogram of Sparse Codes (HSC) for scene text recognition with an accuracy of 72.27%. This work exploits the geometric features of character recognition. Thanh Ha Do et al. [28] empirically constructed two sequences of learned dictionaries for text and graphical parts respectively for efficient text detection. In the work done by Xiaochun Cao et al. [7], Text- specific Multiscale Dictionaries (TMD) and a natural scene dictionary is learned for separately modelling priors on text and non- text fields. This method uses an adaptive version of non- uniform deblurring. In [29], generic HR/LR patch pair database useful for finding a dictionary adapted to the characters properties was established and then partitioned into several clusters by performing an intelligent clustering algorithm. Multiple learned dictionaries based clustered Sparse Coding method was applied to the SR of LR document text images. A similar work is carried out in [15] where sparse coding based resolution enhancement approach using multiple coupled dictionaries is applied. The proposed resolution enhancement approach is applied for the denoising and reconstruction of degraded textual images.

2.7 Comparative Study

In the following table we have compared some of the methods used for text restoration using Sparse Representations.

Table 1. Comparison of Text Restoration Methods

Authors	Methods Used	Results
Thanh Ha Do et al. [28]	K-SVD, Orthogonal Matching Pursuit	Nb. ch-61

Walha, R. et al. [27]	iK-Means, HR/LR patch pairs SISR method;	PSNR-17.24
Vijay Kumar et al. [26]	K-SVD, Averaging overlapping patches	PSNR- 6.69
Dong Zhang et al.[18]	Histograms of Sparse Codes (HSC), K-SVD	Accuracy- 74.34
Walha R. et al.[15]	Multiple coupled dictionaries, iK- Means, Anomalous Pattern Algorithm	PSNR- 17.556
Xiaochun Cao et al. [7]	TMD, SPG	Precision- 56.8%

3. PROPOSED METHOD

3.1 Problem Statement

To design a text image restoration system addressing the entire set of degradations- Image Blur, Noise, Low Light and Masking, using a single method exploiting sparse representations and dictionaries.

3.2 Proposed System Architecture

All the modules of the system architecture rely on the design of the dictionaries created from various types of images. The text specific dictionaries will be created from the sparse representations of degraded text. The natural scene dictionaries will be created from the sparse representations of noisy as well as clean natural images.

The patch- matching module will extract the patches from the input image which are most relevant to the patches from the text specific dictionary. On the basis of these patches, the noise estimation module will estimate the amount of noise present in the image and the blur kernel estimation module will estimate the amount of blur, taking into consideration both the noise and the extracted patches. The input image will then be cleaned and enhanced resulting into a restored image. This image can then be fed to the text detection and recognition module.

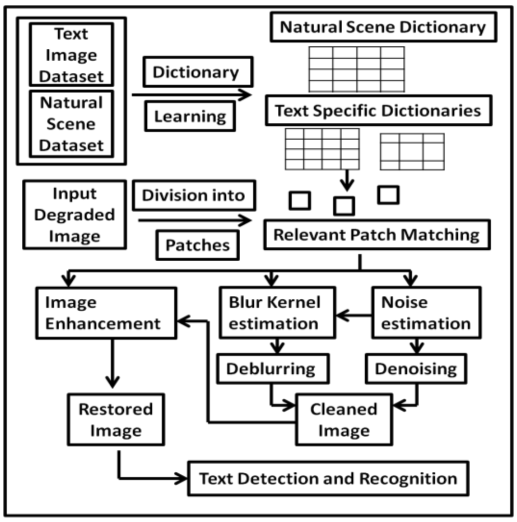

Figure 2: Architecture for Restoration of Images for Text Detection and Recognition System

3.3 Proposed Algorithm

The algorithm will have two phases- The training phase and the testing phase.

3.3.1 Training Phase

Input: Training Images with Degraded Text and Natural Scenes, Clean Text Image Set

Output: Learned Text- Specific and Natural Scene Dictionaries

a) Divide training images and clean text images into patches.

b) Form Text- Specific and Natural Scene sparse representation Dictionaries.

c) Form clean text dictionary.

d) Learn the training image dictionaries for Text extraction.

3.3.2 Testing Phase

Input: Test Images provided by the user

Output: Restored Text in test image

a) Divide test image into patches.

b) Compare patches with Text- Specific and Natural Scene Dictionaries.

c) Extract matching patches with Text- Specific Dictionary for Restoration.

d) Compare extracted patches with clean image dictionary patches and estimate noise and blur levels.

e) Perform image deblurring and denoising.

f) Perform image enhancement.

g) Return restored image to the user.

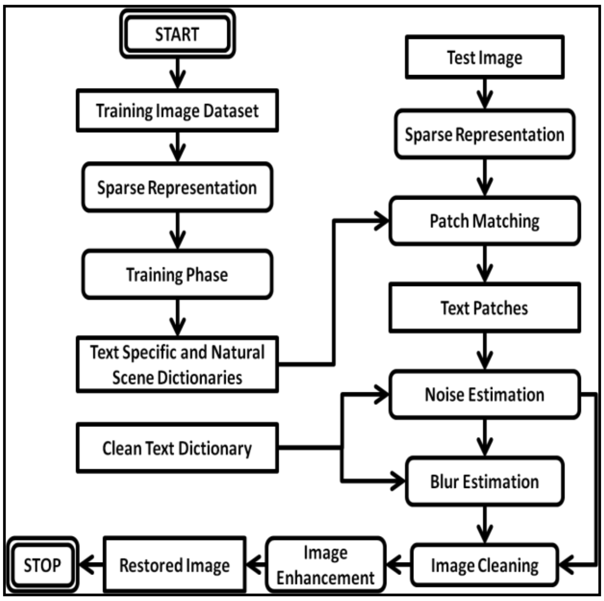

Figure 3: Algorithm for Restoration of Images for Text Detection and Recognition System

4. INPUT AND EXPECTED RESULTS

The proposed work is in the phase of pre- processing and hence no finite results can be shown as of now. We present here our idea of input and the expected results. In the training phase of the algorithm, the input images will consist of the training dataset images of ICDAR 2015. The input images will have text with different amount of noise, blur and illumination. Dictionaries of sparse representations of these images will be formed and given as an input to the algorithm. The algorithm will be trained to detect this type of text.

For the testing phase, the input images will be from the testing dataset of ICDAR 2015. The input test image will be divided into patches and text extraction will be done. The intermediate output will be the identified text region which will then be given as an input to the restoration module. The output of this module will be the restored image. Thus, the proposed system will restore a degraded natural scene text image and make text detection and recognition easy and robust. The system will attempt to overcome challenges of noise, blur, the transparency of text, curvature of text line, low light, multi-coloured text and masking of text to ensure robust text localization and detection.

5. CONCLUSION

This research presents a novel idea of restoring text in natural scene images by using dictionaries of sparse representations. By using this method, the problem of text localization will be solved without using complicated algorithms to remove various forms of degradation in natural scene images. This research will further contribute to the fields of content- based image analysis by making it easier to recognize degraded texts in natural scenes. The proposed method eliminates the need for priors in order to estimate the blur and noise. This work can be further extended by incorporating an inpainting technique using sparse representations.

6. REFERENCES

[1] Jing Han, Jiang Yue, Yi Zhang, and Lianfa Bai, "Local Sparse Structure Denoising for Low-Light-Level Image", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 12, DECEMBER 2015

[2] Jian Sun, and Zongben Xu, "Color Image Denoising via Discriminatively Learned Iterative Shrinkage", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 11, NOVEMBER 2015

[3] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Chang Wen Chen, "Automatic Contrast Enhancement Technology With Saliency Preservation", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 25, NO. 9, SEPTEMBER 2015

[4] Enming Luo, Stanley H. Chan and Truong Q. Nguyen, "Adaptive Image Denoising by Targeted Databases", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 7, JULY 2015

[5] Huanjing Yue, Xiaoyan Sun, Jingyu Yang and Feng Wu, "Image Denoising by Exploring External and Internal Correlations", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 6, JUNE 2015

[6] V. Krithika, "Degradation Removal using Group Sparsity", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 3, Special Issue-1, May 2015

[7] Xiaochun Cao, Wenqi Ren, Wangmeng Zuo, Xiaojie Guo, and Hassan Foroosh, "Scene Text Deblurring Using Text-Specific Multiscale Dictionaries", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 4, APRIL 2015

[8] Ningappa T Pujar, P. B. Madhavanavar, Ramesh M.Badiger, Manjunath Kammar, "Text Extraction from Road Display Boards Using Wavelet and SVM", International Journal of Recent Advances in Science & Engineering Volume 1, Issue 1, March, 2015

[9] Shivananda V. Seeri, J. D. Pujari and P. S. Hiremath, "Multilingual Text Localization in Natural Scene Images using Wavelet based Edge Features and Fuzzy Classification", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 4, Issue 1, January-February 2015

[10] Xiaoyong Shen, Qiong Yan, Li Xu, Lizhuang Ma, Jiaya Jia,"Multispectral Joint Image Restoration via Optimizing a Scale Map", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2015

[11] Xianming Liu, Gene Cheung, Xiaolin Wu, "JOINT DENOISING AND CONTRAST ENHANCEMENT OF IMAGES USING GRAPH LAPLACIAN OPERATOR", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015

[12] Mayank Tiwari and Bhupendra Gupta, "Image Denoising using Spatial Gradient Based Bilateral Filter and Minimum Mean Square Error Filtering", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

[13] Jing Jin, Songyuan Tang, Yi Shen, "An Innovative Image Enhancement Method for Edge Preservation inWavelet Domain",2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)

[14] Guangcan Liu, Shiyu Chang, and Yi Ma,"Blind Image Deblurring Using Spectral Properties of Convolution Operators", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 12, DECEMBER 2014

[15] Rim Walha, Fadoua Drira, Adel M. Alimi, Frank Lebourgeois and Christophe Garcia,"A Sparse Coding based Approach for the Resolution Enhancement and Restoration of Printed and Handwritten Textual Images", 14th International Conference on Frontiers in Handwriting Recognition, 2014

[16] Xin Yu, Feng Xu, Shunli Zhang, and Li Zhang, "Efficient Patch-Wise Non-Uniform Deblurring for a Single Image". IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 16, NO. 6, OCTOBER 2014

[17] Fang Li and Tieyong Zeng, "A Universal Variational Framework for Sparsity-Based Image Inpainting", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 10, OCTOBER 2014

[18] Dong Zhang, Da-Han Wang, and Hanzi Wang, "SCENE TEXT RECOGNITION USING SPARSE CODING BASED FEATURES, IEEE International Conference on Image Processing (ICIP), 2014

[19] Enming Luo, Stanley H. Chan, and Truong Q. Nguyen, "IMAGE DENOISING BY TARGETED EXTERNAL DATABASES", IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP),2014

[20] Wangmeng Zuo, Lei Zhang, Chunwei Song, David Zhang, and Huijun Gao, “Gradient Histogram Estimation and Preservation for Texture Enhanced Image Denoising”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 6, JUNE 2014

[21] C. Paramanand and A. N. Rajagopalan, “Non-uniform Motion Deblurring for Bilayer Scenes”, IEEE Conference on Computer Vision and Pattern Recognition, 2013

[22] Florent Couzinie-Devy, Jian Sun, Karteek Alahari, Jean Ponce, “Learning to Estimate and Remove Non-uniform Image Blur”, IEEE Conference on Computer Vision and Pattern Recognition, 2013

[23] Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee, “Dynamic Scene Deblurring”, IEEE International Conference on Computer Vision, 2013

[24] Y Shi, Q Chang, X Yang, “A Robust and Fast Combination Algorithm for Deblurring and Denoising”, Signal, Image and Video Processing - Springer, June 2013

[25] Purkait, P., Chanda, B., “A FAST AND ROBUST DEBLURRING TECHNIQUE ON HIGH NOISE ENVIRONMENT”, 20th IEEE International Conference on Image Processing (ICIP), 2013

[26] Kumar, V., Bansal, A., Tulsiyan, G.H., Mishra, A., Namboodiri, A., Jawahar, C.V., “Sparse Document Image Coding for Restoration”,12th International Conference on Document Analysis and Recognition (ICDAR), 2013

[27] Walha, R., Drira, F., Lebourgeois, F., Garcia, C., Alimi, A.M., “Multiple Learned Dictionaries based Clustered Sparse Coding for the Super-Resolution of Single Text Image”, 12th International Conference on Document Analysis and Recognition (ICDAR), 2013

[28] Thanh Ha Do, Salvatore Tabbone, Oriol Ramos Terrades, “Text/graphic separation using a sparse representation with multi-learned dictionaries”, 21st International Conference on Pattern Recognition (ICPR), 2012

[29] Whyte, O., Sivic, J., Zisserman, A., Ponce, J., “Non-uniform Deblurring for Shaken Images”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010

[30] Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A., “Non-local Sparse Models for Image Restoration”, IEEE 12th International Conference on Computer Vision, 2009

[31] Huy Phat Le, Gueesang Lee , “Noise Removal from Binarized Text Images”, The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010

[32] J Mairal, G Sapiro, M Elad, “LEARNING MULTISCALE SPARSE REPRESENTATIONS FOR IMAGE AND VIDEO RESTORATION”, Multiscale Modeling & Simulation, 2008 -SIAM

[33] Celine Thillou, Silvio Ferreira, Bernard Gosselin, “An Embedded Application for Degraded Text Recognition”, EURASIP Journal on Applied Signal Processing 2005:13, 2127-2135

[34] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, 3rd ed, Pearson Education, 2008