



ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME



Scientific Report

First name / Family name

MUHAMMAD HANIF

Nationality

PAKISTANI

Name of the *Host Organisation*

Italian National Research Council
(CNR), Pisa (Italy)

First Name / family name
of the *Scientific Coordinator*
Period of the fellowship

Anna Tonazzini

01/03/2017 to 30/11/2018

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Bleed-through degradation is one of the main concern in ancient document image restoration. With the recent developments in digital image processing, the problem posed by bleed-through degradation draw significant attention from the researcher. One of the main attraction in digital restoration of such documents is that the physical restoration of such documents is a cost and time intensive process, with the risk of unintended damage to the original document. Digital image processing based restoration techniques have the advantage of being able to make any number of alterations to the document, whilst leaving the original intact. The aim of bleed through restoration is to improve the readability of affected of affected documents and restore the original form as much as possible. The different restoration approaches in this domain can be broadly categorized into blind and non-blind methods [1]. Blind methods deals with cases where only one side of the document is available, whereas non-blind methods utilize information of both sides (recto and verso).

The bleed-through restoration, with texture estimation, can be posed as image inpainting problem. In this frame the bleed-through is treated as the 'missing' data part and an

estimate is calculated using the local and global image information. Image in-painting is a well-studied image restoration problem with some impressive results. Recently, sparse representation based methods are reported with fast and excellent results in image in-painting applications [3]. Sparsity is an inherent characteristic of natural images and can be obtained by decomposing the signal into its elementary basis components, selected from a predefined basis collection or trained over the data. The collection of basis is generally termed as dictionary [4][5]. In sparse representation the quest is to seek a compact representation of data in terms of linear combination of basis vectors from a dictionary, using some a priori information about the problem in hand.

In my fellowship tenure I worked on different sparse representation based methods to address the bleed-through problem, with the intention to restore the original look of the document and improve the readability. In the literature the bleed-through removal methods mainly focused on the removal of interfering patterns, with no or very little attention to find a befitting replacement for the degraded pixels. In most of the methods a constant value is used to replace the removed degraded pixels, while in some methods a random fill-in is suggested from the surrounding neighbourhood. Unfortunately this random and constant value selection produce visible artefacts, damaging the natural look of the documents.

In recent years sparse representation based methods produce excellent results in estimation of missing data, for example image inpainting, where the surrounding available information is used to find the missing values. In our first method we model the fill-in estimation for the degraded bleed-through pixels as missing data estimation and used sparse image inpainting to find a befitting fill-ins for the degraded pixels to be removed. This modelling produce better results compared to other methods and we published our results in two papers.

In another attempt, we used the colour spectrum information of document image to classify and segment the bleed-through patterns. This colour based segmentation method require only one side (recto or verso) information and belongs to the blind bleed-through restoration methods. A Gaussian mixture model is used for pixel segmentation and then Gaussian texture inpainting model is utilized to estimate a proper fill-in for the pixels classified as bleed-through degradation. The proposed method produce better results compared to the state-of-the-art methods.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

- 1) Muhammad Hanif, Anna Tonazzini, Pasquale Savino, and Emanuele alerno-Blind Bleed-through Removal in Color Ancient Manuscripts Using GMM Based Segmentation. International Journal on Document Analysis and Recognition (IJ DAR), -2019 (Submission stage)

- 2) Muhammad Hanif , Anna Tonazzini, Pasquale Savino, and Emanuele Salerno-Non-local Sparse Image Inpainting for Document Bleed-Through Removal. Journal of Imaging, Vol. 4, 2018.
- 3) Muhammad Hanif, Anna Tonazzini, Pasquale Savino, Emanuele Salerno, Greg Tsagkatakis-Document Bleed through Removal using Sparse Image Inpainting. IAPR Document Analysis Systems (DAS), 2018.
- 4) Franca Debole, Muhammad Hanif, Anna Tonazzini From Digitization to NLP: Manuscript Virtual Restoration. IEEE Machine Learning and Natural Language Processing, 2018.
- 5) Muhammad Hanif, Anna Tonazzini- Sparse Representation Based Inpainting for the Bleed-Through Restoration. European Signal Processing Conference 2017 (EUSIPCO).
- 6) Muhammad Hanif, Anna Tonazzini-Restoration of Ancient Documents using Sparse Image Representation. Digital Humanities ERCIM News letter 1112017.

III – ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

- 1) Arranged one seminar at CNR, Pisa.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

- 1) A 03 Weeks REP visit to the Signal Processing Lab, FORTH, Greece. I work with Dr. Greg Tsagkatakis on tensor based signal processing and its application for bleed-through removal in document images.