SHOW-THROUGH REMOVAL FOR SCANNED IMAGES USING NON-LINEAR NMF WITH ADAPTIVE SMOOTHING

Qingju Liu and Wenwu Wang

Centre for Vision, Speech and Signal Processing, Faculty of Engineering and Physical Sciences,
University of Surrey, Guildford, GU2 7XH, United Kingdom
Emails: [q.liu, w.wang]@surrey.ac.uk

ABSTRACT

Scans of double-sided documents often suffer from show-through distortions, where contents of the reverse side (verso) may appear in the front-side page (recto). Several algorithms employed for show-through removal from the scanned images, are based on linear mixing models, including blind source separation (BSS), non-negative matrix factorization (NMF), and adaptive filtering. However, a recent study shows that a non-linear model may provide better performance for resolving the overlapping front-reverse contents, especially in grayscale scans. In this paper, we propose a new non-linear NMF algorithm based on projected gradient adaptation. An adaptive filtering process is also incorporated to further eliminate the blurring effect caused by non-perfect calibration of the scans. Our numerical tests show that the proposed algorithm offers better results than the baseline methods.

***Index Terms*—** Show-through, NMF, BSS, non-linear NMF, projected gradient

Table 1. SYMBOL NOTATIONS

$\mathbf{X}_i$	the i-th row vector of matrix $\mathbf{X}$
$\mathbf{X}^i$	the i-th column vector of $\mathbf{X}$
x_{ij}	the (i,j)-th element of $\mathbf{X}$
$\mathbf{X} \star \mathbf{Y}$	the element-wise multiplication of $\mathbf{X}$ and $\mathbf{Y}$
$\mathbf{X}^{(k)}$	the updated $\mathbf{X}$ in the k-th iteration
$\|\cdot\|_F$	the Frobenius norm

1. INTRODUCTION

Due to the paper transparency or the non-perfect opacity [1], scanned documents are often degraded by the interferences from the backside. This problem is referred to as show-through shown in Fig. 1. Show-through removal has many potential applications such as in building a digital library. To remove the interference of the verso in the show-through, the techniques of blind source separation (BSS) [4, 7], non-negative matrix factorization (NMF) [5], and adaptive filtering [8] have been considered in the literature [1, 2]. Assuming the mutual independence between the original recto and verso, Tonazzini et al. [1] applied an independent component analysis (ICA) [6] algorithm to the scans, with the assumption of a symmetric and diagonal-dominant linear mixing matrix. In [2], Merrikh et al. used the projected gradient based NMF algorithm in [5], where a modified cost function based on the Euclidean distance between the observation and the reconstructed signals with penalty constraints has been used to compensate the performance degradation caused by the linear model. The adaptive filtering technique proposed in [8] shows good de-blurring performance for the enhancement of the scanned images. Those algorithms are based on the linear model, and in practice, the scanning process is a much more complex non-linear process owning to the transmission medium and other factors such as the *grayscales* of the pixels.

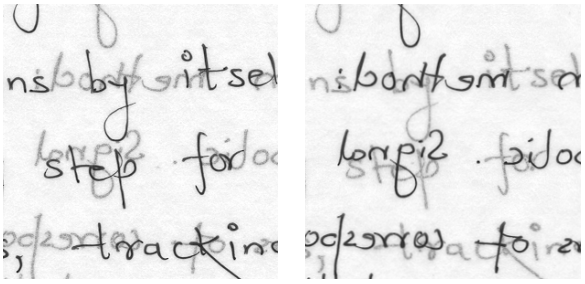


Fig. 1. The recto and verso of a scanned image (from [2, 3], which is also used in our experiment in Section 5).

The additional constraints considered by Merrikh et al. [2] in their proposed cost function suggest that it may offer advantage to consider a non-linear mixing model for the scanning process, or a linear mixing model constrained by non-linear penalty terms. In this paper, we propose a new objective function by considering the non-linear effect of the scanning process. We also derive a new projected gradient NMF algorithm for the optimisation of the proposed cost function. We further apply a post-processing stage based on adaptive filtering to reduce the blurring effect in the scanned image caused by the optical distortions and non-perfect calibration

of the scans. The remainder of the paper is organised as follows. The linear model of the show through and the corresponding NMF solutions are presented in Section 2. Section 3 introduces our proposed non-linear model and the optimisation method, followed by Section 4 on the adaptive smoothing method for the enhanced image. The simulation results are presented in Section 5. Finally Section 6 concludes the paper.

2. LINEAR MODEL AND ITS NMF SOLUTIONS

Scans can be modeled as a linear combination of the recto and the verso as follows:

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

where $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 \\ \mathbf{X}_2\end{bmatrix} \in \mathbb{R}_+^{2\times N}$ is the scanned observation matrix; $\mathbf{X}_1$ (*resp.* $\mathbf{X}_2$) is the concatenated row vector of the recto (*resp.* verso); $\mathbf{A} \in \mathbb{R}_+^{2\times 2}$ is the unknown mixing matrix and $\mathbf{S} = \begin{bmatrix}\mathbf{S}_1 \\ \mathbf{S}_2\end{bmatrix} \in \mathbb{R}_+^{2\times N}$ is the unknown source matrix.

Non-negative matrix factorization is a bound-constrained optimisation process useful for finding representations of non-negative data [5]. To find $\mathbf{S}$ in equation (1), we can minimize the Euclidean distance based cost function with bound constraints defined as follows:

$$J_{\text{nmf}}(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_F^2. \tag{2}$$

The projected gradient descent algorithm [5] has been widely used for the optimisation of this criterion due to its good convergence performance, where $\mathbf{A}$ and $\mathbf{S}$ are iteratively updated as:

$$\begin{aligned}\mathbf{A}^{(k+1)} &= \max\left(\mathbf{0}, \mathbf{A}^{(k)} - \mu\nabla_{\mathbf{A}} J_{\text{nmf}}(\mathbf{A}^{(k)}, \mathbf{S}^{(k)})\right) \\ \mathbf{S}^{(k+1)} &= \max\left(\mathbf{0}, \mathbf{S}^{(k)} - \mu\nabla_{\mathbf{S}} J_{\text{nmf}}(\mathbf{A}^{(k)}, \mathbf{S}^{(k)}))\right),\end{aligned} \tag{3}$$

where $\max(\cdot)$ is the element-wise maximum operator, μ is the step size, and

$$\begin{aligned}\nabla_{\mathbf{A}} J_{\text{nmf}}(\mathbf{A}, \mathbf{S}) &= (\mathbf{AS} - \mathbf{X})\mathbf{S}^T \\ \nabla_{\mathbf{S}} J_{\text{nmf}}(\mathbf{A}, \mathbf{S}) &= \mathbf{A}^T(\mathbf{AS} - \mathbf{X}).\end{aligned} \tag{4}$$

3. NON-LINEAR NMF

The degradation discussed in the previous section is caused by the non-linear scanning where the effect of the reverse interferences on the recto becomes less strong if the front side is much darker [1]. Suppose we use the reverse grayscale in $\begin{bmatrix}0 & 1\end{bmatrix}$ to represent all the pixels (0–white, 1–black), and in the extreme situation where both recto and verso pixels of a certain area are black, then based on equation (1) and the diagonal-dominant symmetry of $\mathbf{A}$ (e.g. $\begin{bmatrix}1 & .5 \\ .5 & 1\end{bmatrix}$), the corresponding elements in $\mathbf{X}$ should be larger than 1, which is however not feasible in practice. Based on this observation, Bayat et al. [2] have proposed a non-linear NMF model and optimisation algorithm for show-through removal.

In [2], however, only the relations between the mixtures have been used in generating the non-linearity in the cost function. In this paper, we propose a new non-linear model in which the relation between the recto and verso is exploited. More specifically, the verso is added to the recto with different factors depending on the grayscales of both sides, and those factors are small when the grayscales are high. Our model is written as follows:

$$\mathbf{X} = \mathbf{AS} - \omega\begin{bmatrix}\mathbf{S}_1 \star \mathbf{S}_2 \\ \mathbf{S}_1 \star \mathbf{S}_2\end{bmatrix}. \tag{5}$$

where ω is a factor that weights the non-linearity from the combination of overlapping pixels, and it is defined in (0, 1), a value close to that of the anti-diagonal elements of $\mathbf{A}$.

Suppose s_1 and s_2 are two registered pixels corresponding to the scanned recto and verso pixels x_1 and x_2. Similar to the mixing matrix structure in [1], we assume $\mathbf{A} = \begin{bmatrix}1 & a_{12} \\ a_{21} & 1\end{bmatrix}$ and $a_{12} \approx a_{21} < 1$. To illustrate the applicability of our model, we show three extreme situations.

1. If $s_1 \approx s_2 \approx 0$,

   then $x_1 = s_1 + a_{12}s_2 - \omega s_1 s_2 \approx s_1 + a_{12}s_2$, and $x_2 \approx s_2 + a_{21}s_1$. Therefore, the linear combination still holds.

2. If $1 \approx s_1 \gg s_2 \approx 0$,

   then $x_1 = s_1 + a_{12}s_2 - \omega s_1 s_2 \approx s_1 + (a_{12} - \omega)s_2$. If $\omega \approx a_{12}$, the second part, i.e. $(a_{12} - \omega)s_2$, can be neglected, which means the reverse-side pixel has no effect on the front-side pixel when the front-side pixel is black. $x_2 = s_2 + a_{21}s_1 - \omega s_1 s_2 \approx a_{21}s_1 + (1 - \omega)s_2 \approx a_{21}s_1$, which means the weak verso pixel s_2 is masked by the show-through of the dark recto pixel s_1, therefore, the factor ω mitigates the influence of s_2.

3. If $s_1 \approx s_2 \approx 1$,

   then $x_1 = s_1 + a_{12}s_2 - \omega s_1 s_2 \approx s_1 + (a_{12} - \omega)s_2$, and if $\omega \approx a_{12}$, $x_1 \approx s_1 \approx 1$ and in the same way $x_2 \approx 1$. Therefore, the scans of overlapped black pixels are still black.

Denote $\dot{\mathbf{S}} = \begin{bmatrix}\mathbf{S}_1 \star \mathbf{S}_2 \\ \mathbf{S}_1 \star \mathbf{S}_2\end{bmatrix}$, we can minimize the following objective function:

$$J(\mathbf{A}, \mathbf{S}) = \frac{1}{2}\|\mathbf{X} + \omega\dot{\mathbf{S}} - \mathbf{AS}\|_F^2. \tag{6}$$

Using the same projected gradient algorithm in equation (3), we can derive the learning equations for the optimisation of function (6). According to equation (4) and replacing $\mathbf{X}$ with $\mathbf{X} + \omega\dot{\mathbf{S}}$, we get the gradient of J with respect to $\mathbf{A}$:

$$\nabla_{\mathbf{A}} J(\mathbf{A}, \mathbf{S}) = \left(\mathbf{AS} - (\mathbf{X} + \omega\dot{\mathbf{S}})\right)\mathbf{S}^T. \tag{7}$$

The partial derivative of the objective function $J(\mathbf{A},\mathbf{S})$ with subject to $\mathbf{S}$, i.e., $\nabla_{\mathbf{S}}J(\mathbf{A},\mathbf{S})$ can be derived as follows.

Suppose $\mathbf{S}_{new}=\mathbf{S}+\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a $2\times N$ perturbation matrix and all its elements are zero except the (i,j)-th element ε_{ij}. Let s_{ij} be the (i,j)-th element of $\mathbf{S}$ and further denote

$$\mathbf{B}=\mathbf{X}+\omega\dot{\mathbf{S}}-\mathbf{AS}. \tag{8}$$

We have

$$\begin{aligned}
&J(\mathbf{A},\mathbf{S}_{new})=\frac{1}{2}\|\mathbf{B}_{new}\|_F^2\\
&=\frac{1}{2}\|\mathbf{X}+\omega\dot{\mathbf{S}}+\omega\begin{bmatrix}0&\cdots&s_{\bar{i}j}\varepsilon_{ij}&\cdots\\0&\cdots&s_{\bar{i}j}\varepsilon_{ij}&\cdots\end{bmatrix}-\mathbf{AS}-\mathbf{A}\boldsymbol{\varepsilon}\|_F^2,
\end{aligned}$$

where

$$\bar{i}=\begin{cases}2,\ i=1\\1,\ i=2\end{cases}$$

and

$$\mathbf{A}\boldsymbol{\varepsilon}=\varepsilon_{ij}\begin{bmatrix}0&\cdots\\0&\cdots\end{bmatrix}\mathbf{A}^{i}\begin{bmatrix}\cdots\\\cdots\end{bmatrix}.$$

Therefore,

$$\begin{aligned}
&\{\nabla_{\mathbf{S}}J(\mathbf{A},\mathbf{S})\}_{ij}=\frac{\partial J(\mathbf{A},\mathbf{S})}{\partial s_{ij}}\\
&\underset{\varepsilon_{ij}\to 0}{=}\frac{J(\mathbf{A},\mathbf{S}_{new})-J(\mathbf{A},\mathbf{S})}{\varepsilon_{ij}}\\
&=\frac{1}{2\varepsilon_{ij}}\Big(\|\mathbf{B}_{new}\|_F^2-\|\mathbf{B}\|_F^2\Big)\\
&=\frac{1}{2\varepsilon_{ij}}\sum(\mathbf{B}_{new}-\mathbf{B})\star(\mathbf{B}_{new}+\mathbf{B})\\
&=\frac{1}{2\varepsilon_{ij}}\sum\Big(\omega\varepsilon_{ij}\begin{bmatrix}0&0&\cdots&s_{\bar{i}j}&\cdots\\0&0&\cdots&s_{\bar{i}j}&\cdots\end{bmatrix}-\varepsilon_{ij}\begin{bmatrix}0&\cdots\\0&\cdots\end{bmatrix}\mathbf{A}^{i}\begin{bmatrix}\cdots\\\cdots\end{bmatrix}\Big)\star\\
&\qquad\Big(2\mathbf{X}+2\omega\dot{\mathbf{S}}-2\mathbf{AS}+o(\varepsilon_{ij})\Big)\\
&\approx\sum\Big(\omega\begin{bmatrix}0&0&\cdots&s_{\bar{i}j}&\cdots\\0&0&\cdots&s_{\bar{i}j}&\cdots\end{bmatrix}-\begin{bmatrix}0&\cdots\\0&\cdots\end{bmatrix}\mathbf{A}^{i}\begin{bmatrix}\cdots\\\cdots\end{bmatrix}\Big)\star\mathbf{B}\\
&=\omega s_{\bar{i}j}[1\ 1]\mathbf{B}^{j}-(\mathbf{A}^{i})^{T}\mathbf{B}^{j}\\
&=\omega\bar{s}_{ij}\mathbf{C}_i\mathbf{B}^{j}-(\mathbf{A}^{T})_i\mathbf{B}^{j}
\end{aligned}$$

where $\bar{s}_{ij}$ is the (i,j)-th element in matrix $\bar{\mathbf{S}}$, and

$$\bar{\mathbf{S}}=\begin{bmatrix}\mathbf{S}_2\\\mathbf{S}_1\end{bmatrix},\qquad\mathbf{C}=\begin{bmatrix}1&1\\1&1\end{bmatrix}. \tag{9}$$

Hence we can obtain the gradient of J with respect to $\mathbf{S}$:

$$\nabla_{\mathbf{S}}J(\mathbf{A},\mathbf{S})=\omega\bar{\mathbf{S}}\star\mathbf{CB}-\mathbf{A}^{T}\mathbf{B}, \tag{10}$$

where the matrices $\mathbf{B}$, $\bar{\mathbf{S}}$ and $\mathbf{C}$ are defined in equations (8) and (9).

As a result, $\mathbf{A}$ and $\mathbf{S}$ can be found by iteratively updating equations (3), (7) and (10). In the proposed algorithm we initialise $\mathbf{A}^{(0)}$ with a positive matrix, then calculate $\mathbf{S}^{(0)}=\max(0,(\mathbf{A}^{(0)})^{-1}\mathbf{X})$. We stop the iteration when the running time or the number of iterations reaches the predefined maximum time or iteration number, or when the current gradient norm is one percent of the initial gradient norm.

4. SMOOTHING

Due to the optical distortions, there may exist blurring effect in show-through. To eliminate these problems, we employ a post-processing stage based on smoothing [8].

To this end, the vector $\mathbf{x}_1$ (.resp $\mathbf{x}_2$) and the recovered $\mathbf{s}_1$ (.resp $\mathbf{s}_2$) via the nonlinear NMF are re-shaped into $M\times N$ matrices $\mathbf{X}_1$ and $\mathbf{S}_1$ (.resp $\mathbf{X}_2$ and $\mathbf{S}_2$). The white paper reflectance R_p^w is estimaed as the mean of the 10 percent highest pixel values of $\mathbf{s}_1$ and $\mathbf{s}_2$. Now we can calculate the density matrix $\mathbf{D}_{x1}$ as:

$$\mathbf{D}_{x1}=\log\frac{\mathbf{X}_1}{R_p^w},\qquad\mathbf{E}_{s1}=\mathbf{I}-\frac{\mathbf{S}_1}{R_p^w}.$$

where $\log(\cdot)$ operates element-wise, and other density matrices $\mathbf{D}_{x2},\mathbf{D}_{s1},\mathbf{D}_{s2}$ can be computed similarly. And the absorption matrix $\mathbf{E}_{s1}$ can be obtained as (similar for $\mathbf{E}_{s2}$)

$$\mathbf{E}_{s1}=\mathbf{I}-\frac{\mathbf{S}_1}{R_p^w}.$$

For each source $i=1,2$ and each pixel with coordinate (m,n), we examine the condition $|D_{si}(m,n)-D_{xi}(m,n)|>\epsilon_1$, where $|\cdot|$ is the modulus. If it is satisfied, we assign $\check{S}_i(m,n)=S_i(m,n)$ where $\check{}$ stands for update; otherwise, we apply smoothing to the pixels as follows.

$$\check{D}_{si}(m,n)=D_{si}(m,n)-\sum\mathbf{H}_i\star\mathbf{E}_{s\bar{i}}^{b}$$

where $\sum$ is the summation over all the elements in the matrix, $\exp(\cdot)$ is the exponential operator, and $\mathbf{H}_i$ is the $P\times P$ smoothing matrix for source i, which can be updated as

$$\mathbf{H}_i=\mathbf{H}_i+\beta\mathbf{E}_{s\bar{i}}^{b}\check{D}_{si}(m,n)$$

$$\mathbf{H}_i=\max(\mathbf{H}_i,\mathbf{0})$$

If $\sum\mathbf{H}_i\star\mathbf{E}_{s\bar{i}}^{b}>\epsilon_2$, $\mathbf{H}_i$ is attenuated

$$\mathbf{H}_i=\alpha\mathbf{H}_i \tag{11}$$

Note that, the elements of $\mathbf{E}_{s\bar{i}}^{b}$ (a $P\times P$ matrix) are taken from the squared $P\times P$ block of $\mathbf{E}_{s\bar{i}}$ centered at the (m,n)-th pixel, and $\bar{i}=\begin{cases}2,\ i=1\\1,\ i=2\end{cases}$. ϵ_1 and ϵ_2 are threshold values, α is the attenuation parameter and β is the step-size. The bold $\mathbf{0}$ is a $P\times P$ matrix whose elements are all zeros.

Different from the filtering method in [8], we calculate the white paper reflectance R_p^w based on clustering (instead of user manipulation). In addition, the show-through density is only calculated for part of the pixels selected by the criterion $|D_{si}(m,n)-D_{xi}(m,n)|>\epsilon_1$. Moreover, the smoothing filters $\mathbf{H}_i$ are attenuated by (11) if the density difference is higher than a threshold after the show-through correction.

5. EXPERIMENTAL RESULTS

We have tested the proposed algorithm on both synthetic and real data. The parameters of the proposed algorithm are set as: $u = 10^{-6}$, $\lambda = 0.5$, $\epsilon_1 = 0.3$, $\epsilon_2 = 1$, $\alpha = 0.8$, $P = 11$ and $\beta = 0.01$. First we demonstrate the effectiveness of our algorithm with synthetic data. We generate the observation matrix with $\mathbf{X} = \mathbf{A}_{\text{real}}\mathbf{S} - \omega_{\text{real}}\dot{\mathbf{S}}$, where $\mathbf{S}$ is the uniformly distributed random source matrix, and $\mathbf{A}_{\text{real}} = \begin{bmatrix} 1 & \omega_{\text{real}} \\ \omega_{\text{real}} & 1 \end{bmatrix}$. We use

$$\mathcal{J}(\mathbf{A}) = \|\mathbf{A} - \mathbf{A}_{\text{real}}\|_F, \tag{12}$$

to evaluate the results, where $\mathbf{A}$ is the estimated mixing matrix after the optimisation with regularized diagonal elements. Table 2 shows the $\mathcal{J}(\mathbf{A})$ when we choose different ω in the optimisation process. From the minimum values in bold font, we found that we can recover the mixing matrix with a high accuracy when we choose the ω close to the anti-diagonal elements in $\mathbf{A}$.

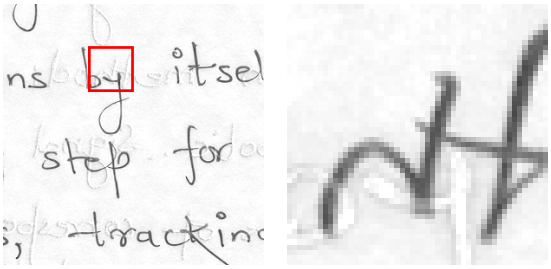
Table 2. Evaluation on synthetic data

$\mathcal{J}(\mathbf{A})$	ω in the optimisation							
ω_{real}	0	.3	.4	.5	.6	.7	.8	1
.4	.14	**.04**	.09	.14	.19	.23	.27	.34
.5	.28	.08	**.02**	.04	.10	.14	.19	.27
.6	.42	.20	.12	.05	**.01**	.06	.12	.20
.7	.57	.32	.23	.16	.08	**.02**	.04	.14
.8	.71	.44	.35	.26	.18	.10	**.03**	.09

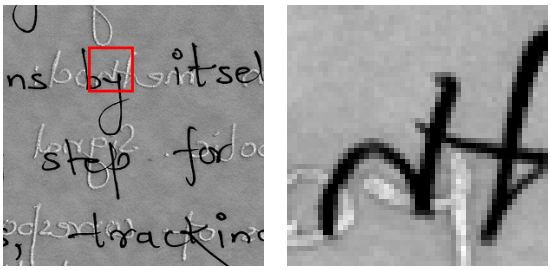
We then applied the proposed algorithm on the same data (`www.site.uottawa.ca/~edubois/documents`) as used in [2] (shown in Fig. 1). This enables us to compare the proposed method with [2]. For the NMF algorithm [5], the step-size $u = 10^{-6}$, the same as used in [2]. The recovered signals are shown in Fig. 2. With the proposed algorithm, we can effectively remove most of the interferences from the other side, and the *lighter* pixels in the overlapping area are compensated by $\tilde{\mathbf{S}}$ in the non-linear NMF. It can be observed that the proposed method performs much better than both the linear NMF [5] and the non-linear NMF in [2].

6. CONCLUSION

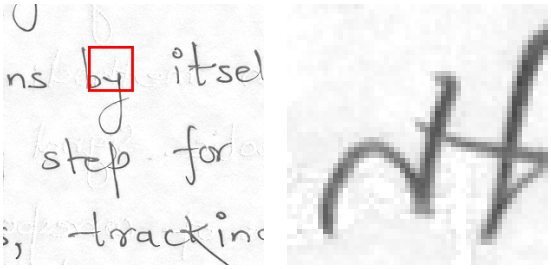
We have proposed a new method for the enhancement of scanned image based on NMF and adaptive smoothing. Using a projected gradient method, we have developed a new non-linear NMF algorithm. To overcome the limitations of the blurring effect, adaptive smoothing has also been applied to enhance the scanned image. Results show the good performance of the proposed method, as compared with several baseline methods.



(a) Linear NMF [5].



(b) Non-linear NMF [2].



(c) Proposed method.

Fig. 2. Comparison of show-through removal for real scanned documents. Left, recto and right, flipped verso. We zoom in the flipped verso in the corresponding area highlighted by the red rectangle in the recto.

7. REFERENCES

[1] A. Tonazzini, E. Salerno and S.L. Bedini, "Fast correction of bleed-through distortion in grayscale documents by a blind source separation technique," *Int. J. on Document Anal. and Recogn.*, vol. 10, pp. 17–25, 2007.

[2] F. M. Bayat, M. B. Zadeh, and C. Jutten, "Using non-negative matrix factorization for removing show-through," in *Proc. LVA/ICA*, 2010, pp. 482–489.

[3] F. M. Bayat, M. B. Zadeh, and C. Jutten, "Linear-quadratic blind source separating structure for removing show-through in scanned documents," *Int. J. on Document Anal. and Recogn.*, vol. 14, pp. 319–333, 2011.

[4] C. Jutten, J. Herault, "Blind separation of sources, part I:

An adaptive algorithm based on neuromimetic architecture," *Signal Process*, vol. 24, pp. 1–10, 1991.

[5] C.J. Lin, "Projected gradient methods for non-negative matrix factorization," *Neural Computation*, vol. 19, pp. 2756–2779, 2007.

[6] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.

[7] J.F. Cardoso, and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc-F. Radar and Signal Processing*, vol. 140, pp. 362–370, 1993.

[8] G. Sharma, "Show-through cancellation in scans of duplex printed documents," *IEEE Trans. Image Process.*, vol. 10, pp. 736–754, 2001.