Independent Component Analysis for Document Restoration

Anna Tonazzini, Luigi Bedini and Emanuele Salerno

Istituto di Scienza e Tecnologie dell'Informazione
Area della Ricerca CNR di Pisa
Via G. Moruzzi, 1, I-56124 PISA, Italy
`<surname>@iei.pi.cnr.it`

Abstract

In the digital images of many documents, the legibility of the text is often compromised by the presence of artifacts in the background. These can derive from many kinds of degradations, such as spots, underwritings, show-through or bleed-through effects. The use of thresholding techniques to remove the background, while can perform well for black and white documents, is not effective for gray level or color documents, since the color values of this background can be very close to those of the text. For the specific problem of bleed-through/show-through, some work has been done, mainly based on the comparison between the front and back page. This, however, requires a preliminary registration of the two images. In this paper, we propose a novel approach, based on viewing the problem as one of separating overlapped texts, and then reformulating it as a Blind Source Separation problem, approached through Independent Component Analysis techniques. Our method and uses the spectral components of the image at different bands, so that there is no need for registration. Examples of bleed-through cancellation and recovering of underwriting from palimpsests are provided.

Categories and Subject descriptors: (ACM) I.4 [Image Processing and Computer Vision]: I.4.3 Enhancement; I.4.8 Scene Analysis *Color*; I.5 [Pattern Recognition]: I.5.4 Applications *Text processing*; I.7 [Document and Text Processing]: I.7.7 Document Capture *Document analysis.*

1 Introduction

In many digital images of documents, either ancient or modern printed texts and manuscripts, often the readability of the writing is made difficult by the presence of strong artifacts in the background. These artifacts can derive from many kinds of degradations, such as scan optical blur and noise, spots, underwriting, overwriting, or the so-called bleed-through or show-through effects. In many ancient documents, bleed-through is intrinsic in the document, because it is caused by seeping of ink from the reverse side, while in modern, double-sided printed documents, show-through appears in the scanned image when the paper is not completely opaque. The use of thresholding techniques to remove these effects can perform well for pure black and white documents, but, in general, it is not effective for gray level or color documents, since the intensities of the unwanted background can be very close to those of the foreground text. In these conditions, thresholding either does not remove bleed-through, or also eliminates part of the information in the front side.

In the case of palimpsests, usually ancient manuscripts that have been erased and then rewritten, the problem is the opposite, and what is desired is to enhance and let "emerge" the traces of the original underwriting. In both cases, extracting the writing of interest from the background, or, equivalently, removing the interfering writing, can be seen as the problem of separating overlapped texts.

For the specific problem of bleed-through/show-through, some work has been done, mainly based on the exploitation of information from the front and back pages. In [15], the physical model of these effects is first simplified for deriving a linear mathematical model, and then an adaptive linear filter is developed that uses scans of both sides of the documents. In [17], the two sides of a grey level manuscript are compared at each pixel, and, basically, a thresholding is applied. In [18] a wavelet technique is applied for iteratively enhancing the foreground strokes and smearing the interfering strokes. In all these methods, a preliminary registration of the two sides is required. In [16] the front side alone of a color image is processed via a multiscale analysis employing adaptive binarization and edge magnitude thresholding.

We adopt the point of view that extracting the text of interest or, equivalently, removing the interfering text, can be seen as the problem of separating overlapped texts. Thus, in this paper we propose a novel approach, based on reformulating the problem as a Blind Source Separation (BSS) problem, where the overlapping texts and the support (paper, parchment, etc.) are the unknown sources to be recovered, and multiple views of the documents in different spectral bands are the observations. In the most natural way, as spectral bands we consider the red, green and blue channels in which an image can always be split. Nevertheless, as it will be clarified later on, extra observations taken in non-visible spectral bands can be useful as well.

As highlighted in [15], the physical model underlying both bleed-through and show-through is very complicated, in that it is non-linear with some unknown parameters, and it should also take into account for the spreading of light or ink in the support, which causes a blurring of unknown PSF of the show-through component. In [15] suitable transformations and simplifying approximations are adopted to linearize the problem. In this first stage of our study, we also adopted a linear approximation, and derived a model in which the observations are seen as linear mixtures of the sources themselves. Unfortunately, however, the coefficients of these mixtures are usually unknown. The problem of separating the overlapped texts becomes thus the highly undetermined problem of jointly estimate the sources and the mixing coefficients. Nevertheless, at least in an idealized setting (e.g. noiseless images), the problem can be efficiently solved through Independent Component Analysis techniques. In particular, we employ the FastICA algorithm [5], which is an efficient, fully blind, and extremely fast procedure. Furthermore, since we use as observations the spectral components of the same image, there is no need for image registration. Preliminary examples of bleed-through cancellation and recovering of underwriting from palimpsests are provided.

The paper is organized as follows. In Section 2, the linear Blind Source Separation problem is formulated in a general context, and the basic principles of Independent Component Analysis and of the FastICA algorithm are stated. Section 3 is devoted to the assessment of the model for images of overlapped texts, and then to the formulation of their separation as a BSS problem of a linear, instantaneous mixture of unknown sources and unknown mixing coefficients. A first, fully artificial, example of the performance of the FastICA algorithm is provided as well. In Section 4, we provide experimental results of the proposed method for the recovering of partially hidden underwritings from palimpsests. We compare our method with the one proposed in [19], and applied to separate the overlapped texts in the famous Archimedes Palimpsest. Finally, in Section 5 we give the results of the separation of foreground text and bleed-through text from artificial images, and from images of ancient documents showing a real bleed-through effect.

2 Blind Source Separation and Independent Component Analysis

Blind source separation (BSS), which became an active research topic in signal processing in the last decade, has only very recently received attention in image processing and computer vision. It consists of separating a set of unknown signals from a set of mixtures of them, when no full knowledge is available about the mixing operator. The most studied BSS problem refers to a linear data model, where the observations are additive mixtures, with unknown coefficients, of the source signals. A well-known application example of linear BSS is the so-called "cocktail party" problem in audio processing. Other applications include the removal of underlying artifact components of brain activity from EEG records, the search for hidden factors in parallel financial data series, feature extraction or noise removal from natural images, and source separation in astrophysical microwave maps.

In order to solve BSS, which is a severely ill-posed inverse problem, many techniques have been proposed so far. Among them, the Independent Component Analysis (ICA) methods are based on the assumption of mutual independence of the sources. Most of these methods were developed in the case of noiseless data, and differ from one another in the way they enforce independence. The Maximum Likelihood (ML) method [1] directly assumes a factorized form for the joint source distribution; in the infomax method [2], entropy is used as a measure of independence; other methods are based on the minimization of contrast functions, related to statistics of order greater than two, still to ensure independence [3]. The strict relationships among the various methods have been investigated as well [4], and in [10][11][12], Bayesian estimation has been proposed as a suitable, unifying framework for BSS, within which the other methods can be viewed as special cases. All the above methods have shown good performances in many practical applications. In particular, a very efficient and fast algorithm, the FastICA algorithm, has been proposed in [5]. However, the ICA solution to BSS presents some drawbacks or not yet explored/solved issues. In fact, the independence requirement can be fulfilled in some practical applications, but in many cases there is a clear evidence of correlation among the sources. Furthermore, ICA algorithms have been developed for noiseless data and most of them do not account for a possible time correlation inside the single sources, and/or for different numbers of sources and data signals. Finally, convolutive or non-linear mixtures could better fit some practical problems. All the above limitations are currently under investigation. In particular, to manage noisy mixtures, the noisy FastICA algorithm [6], and an Independent Factor Analysis method (IFA) [7][8][9] have been developed. Other techniques for the separation of noisy mixtures take advantage from incorporating into the problem available information about auto-correlation properties of the single sources [13][14][21]. Indeed, correlation is an important feature of most real-world signals, and especially of images, and, when used as a constraint, it is known to be able to regularize and stabilize many ill-conditioned inverse problems. Finally, the possibility of separating cross-correlated sources is an emerging topic as well [20].

Although document images present many of the above features (auto- and cross-correlation of the sources, scanner noise, etc.), in this first application of BSS and ICA, we will develop our method in an idealized setting, with the main aim at showing the novelty and the great potentiality of this kind of approaches for the processing and the analysis of degraded documents.

The data generation model for a noiseless linear BSS problem is given by:

$$\mathbf{x}(t) = A\mathbf{s}(t) \qquad t = 1, 2, ..., T \tag{1}$$

where $\mathbf{x}(t)$ is the vector of the measurements, $\mathbf{s}(t)$ is the column vector of the unknown sources, at location t, and A is the unknown mixing matrix. We assume the same number N of measured and source signals, so that A is an $N \times N$ matrix.

Obviously, solving the system in eq. 1 with respect to both A and $\mathbf{s} = (\mathbf{s}(1), ..., \mathbf{s}(T))$, would clearly give an undetermined problem, unless more information is exploited. The kind of information used in the ICA approach to BSS is independence of the sources. Assuming to know the prior distribution for each source, the joint prior distribution for $\mathbf{s}$ is thus given by:

$$P(\mathbf{s}) = \prod_{i=1}^{N} P_i(\mathbf{s}_i) \tag{2}$$

where $\mathbf{s}_i = (s_i(1), s_i(2), ..., s_i(T))$. The separation problem can be formulated as the maximization of eq. 2, subject to the constraint $\mathbf{x} = A\mathbf{s}$. Calling W the unknown matrix A^{-1}, the problem reduces to the search for a W, $W = (\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N)'$, such that, when applied to the data $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$, produces the set of vectors $\hat{\mathbf{s}}_i = \mathbf{w}_i'\mathbf{x}$ that are maximally independent, and whose distributions are given by the P_i. By taking the logarithm of eq. 2, the problem solved by ICA algorithms is then:

$$\hat{W} = arg \max_W \sum_t \sum_i log P_i(\mathbf{w}_i'\mathbf{x}(t)) + T log|det(W)| \tag{3}$$

Once estimated W, the vectors $W\mathbf{x}$ are copies of the original sources $\mathbf{s}$, apart from unavoidable scale and permutation indeterminacies. It has been shown that, besides independence, to make separation possible a necessary extra condition for the sources is that they all, but at most one, must be non-Gaussian. To enforce non-Gaussianity, generic super-Gaussian or sub-Gaussian distributions can be used as priors for the sources, and have proven to give very good estimates for the mixing matrix and for the sources as well, no matter of the true source distributions, that, on the other hand, are usually unknown [1].

The FastICA algorithm [22] gives the possibility of choosing among a number of "non-linearities", to be used in place of the derivatives of the log-distributions. It solves eq. 3 and returns the estimated sources by using a fixed-point iteration scheme [5] that has been found in independent experiments to be 10-100 times faster than conventional gradient descent methods.

3 Formulation of the overlapped text separation as a BSS problem

We assume hereafter that a palimpsest image or a document image affected by bleed-through/show-through can be modelled as the superposition of three different sources, or classes, that we will call "background", "overwriting" and "underwriting", respectively. In the BSS formalism, this means that we have three different sources that combine someway to give the observed image. At the same time, as already mentioned, we can assume to have three observed maps, obtained by splitting the mixture of the three sources into the red, green and blue components. Thus, we have the same number of sources and observations, which is the most classical assumption for BSS problems. Since we consider images of documents containing text, we can also reasonably assume that the color of each of the three sources is almost uniform, i.e we will have mean reflectance indices (r_1, g_1, b_1) for the background, mean reflectance indices (r_2, g_2, b_2) for the overwriting and mean reflectance indices (r_3, g_3, b_3) for the underwriting.

For the superposition of the three classes we developed an approximated linear mixture model assuming that, at each point of the document, the three reflectance indices of the three classes mix linearly to form the actual reflectance indices. In this model, the reflectance indices $(x_r(t), x_g(t), x_b(t))$ of a generic point t of the document can be seen as given by the following equation:

$$[x_r(t), x_g(t), x_b(t)] = \begin{bmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix} \tag{4}$$

where functions $s_i(t)$, $i = 1, 2, 3$ indicate the "quantity" of background, overwriting and underwriting, respectively, that concur to form the color at point t. For instance, the reflectance indices of, say, a pure background point t, will be given by:

$$[x_r(t), x_g(t), x_b(t)] = \begin{bmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

It is immediate to verify that eq. 4 is of the same form of eq. 1, restricted to the 3×3 case, where parameters r_i, g_i, and b_i are the coefficients of the mixing matrix A, and functions $s_i(t)$ are the sources. As already said, in this first application of the method, we assume that noise or blur in the documents can be neglected, i.e. that the data model is a noiseless istantaneous mixture. In this conditions, the ICA principle described in the previous section can be applied to our data model, and both the matrix of the coefficients and the functions $s_i(t)$ can be estimated by using the FastICA algorithm.

Nevertheless, for a correct application of the method, a fundamental constraint is that the mixing matrix in eq. 4 should be non singular. This means that the reflectance indices of each source must be linearly independent, i.e. the sources must have different colors. However, when the mixing matrix is ill-conditioned or singular, extra observations taken in non-visible spectral bands, used in conjunction or in place of the visible channels, could constitute a remedy.

Below we will give a first example of the performance of the FastICA algorithm for separating two overlapped images. The example is completely artificial, though we used as a basis the scan of a real gray level document. We considered this image as first source (Figure 1a) and an its rotated version as second source (Figure 1b). To generate the observations, we linearly mixed the two available sources with the following, randomly generated, 2×2 matrix:

$$A = \begin{bmatrix} 0.6992 & 0.7275 \\ 0.4784 & 0.5548 \end{bmatrix}$$

Note that these observations are not intended to simulate the color components of an image containing overlapped texts, but they represent just the linear superposition of two images, without reference to any specific application. The two mixtures are shown in Figure 1c and 1d, respectively. Applying FastICA to these observations we obtained the images shown in Figure 1e and 1f, and the 2×2 estimated matrix:

$$\hat{A} = \begin{bmatrix} 0.7986 & -0.7152 \\ 0.5542 & -0.5459 \end{bmatrix}$$

It is easy to verify that, apart from a scale factor of around 1.15 affecting the first column (corresponding to the first source), and a scale factor of around -0.99 affecting the second column (corresponding to the second source), the estimated matrix is satisfactorily good, as well as the estimated sources of Figures 1e and 1f, to be compared with the originals of Figures 1a and 1b. Note that, due to the negative scale factor on the second column of the estimated matrix, the second estimated source is correctly the negative image of the original source.

4 Recovery of underwriting from palimpsests

For the case of palimpsests, the information of interest is usually the old text that has been erased before rewriting a new one. In this situation, often the old underwriting is completely hidden in the visible image, but it is very frequent that some traces of it appear when the

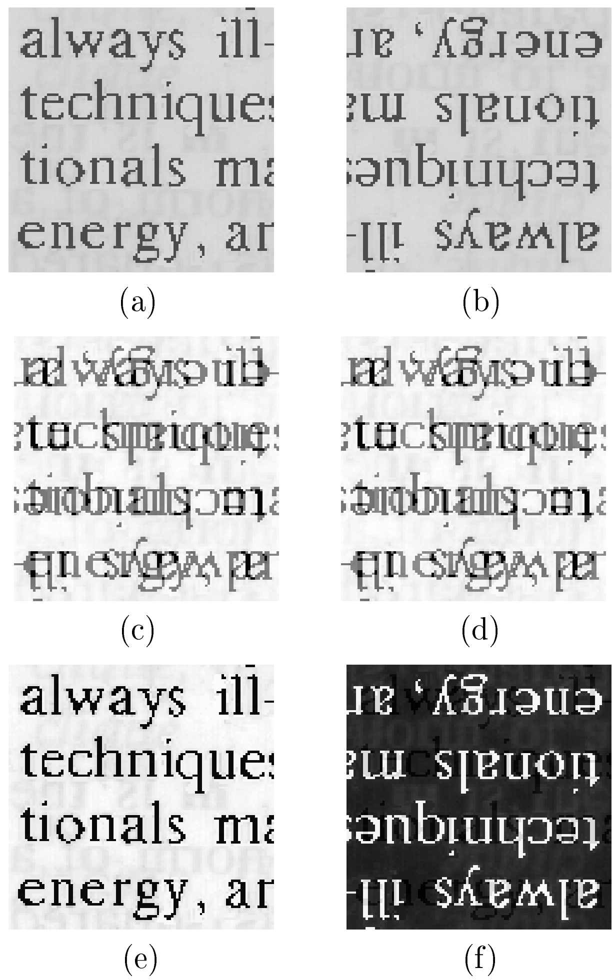


Figure 1: Synthetic example: (a) first original image; (b) second original image; (c) first mixture; (d) second mixture; (e) first FastICA output; (f) second FastICA output.

document is acquired in the infrared band, using a multispectral camera. Thus, the infrared channel has to be used as an extra channel, or in substitution of one of the visible bands, in order to make possible the separation of the three classes.

In the following we will give a first example of the application of our method for the recovering of the underwriting from an artificial palimpsest proposed as an exercise in the web site of Roger L. Easton of the Rochester Institute of Technology [19]. Easton uses this exercise to explain the technique that he and his group adopted for the restoration of the Archimedes Palimpsest. He proposed two methods: a simpler one, that is able to recover only the underwriting, and a more sophisticated one, that is able to recover all the three sources (overwriting, underwriting and background). In this second method, for the color at each point of the image, they adopt a linear mixture model of the same type of the one shown in eq. 4, but assume to know in advance the mixing coefficients. These are estimated by manually reading the values of the red, green and blue components in a number of points of each class, and then computing an average of these three values for each class. The classes are reconstructed by inverting the 3×3 matrix so obtained. Thus, their method is basically the same of ours, with the only, but not negligible, difference that their method is not blind, i.e. it requires the a priori knowledge of the mixing matrix, while our method is fully blind, i.e. it estimates the mixing matrix jointly with the sources. It has to be noted that the manual estimation of the mixing matrix is a cumbersome task, which is possible only when there are areas in the image where the three classes are well separated and distinguishable from each other.

In Figure 2a we show the original color image, while the three channels, the observations, are provided in Figures 2b, 2c and 2d, respectively. Given these observations alone, FastICA is able to reconstruct the sources shown in Figures 2e, 2f and 2g, respectively. Note that, since the recovered mixing matrix contains as coefficients the mean reflectance indices for the fundamental colors of each class, according to eq. 4, it is possible to reconstruct the color of the separated images, as shown in Figure 3. In these color images, the background does not correspond to the "background class", but it has to be intended as a residual, due to the other sources, of a non-perfect separation. In Figure 4a, 4b, and 4c we show the results of the supervised separation, provided in the Easton's web site. It is immediate to note the equivalence between these results and those obtained with our approach, apart from a permutation in the outputs of the FastICA algorithm. The higher contrast and the cleaner appearance of the results of the Web exercise could be due to some post-processing, for instance based on some sort of thresholding, or to the fact that we used a jpeg compressed image downloaded from the Easton's site.

5 Removal of bleed-through and show-through

In this Section we provide the results obtained with our ICA approach for the separation of foreground text and bleed-through/show-through from a synthetic color image generated according to the model of eq. 4, and from a real image of an ancient document. Also in this second case we assumed an underlying additive model.

Figure 2: Synthetic example of a palimpsest: (a) color image; (b) red channel; (c) green channel; (d) blue channel; (e) first FastICA output; (f) second FastICA output; (g) third FastICA output.

(a) (b)

Figure 3: Recovered color images for the separations in Figure 2: (a) first FastICA output; (b) second FastICA output.

Figure 5 shows the results of the experiment performed on the synthetic color image (Figure 5a). This image has been obtained by mixing a black front side text, a black reverse side text, and a light, non-uniform background, according to eq. 4, with coefficients given by the following matrix:

$$A = \begin{bmatrix} 0.2470 & 0.1729 & 0.9000 \\ 0.2510 & 0.1255 & 0.9000 \\ 0.2510 & 0.1004 & 0.9000 \end{bmatrix}$$

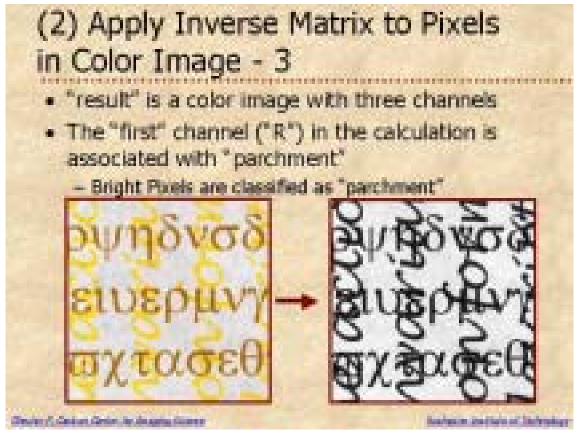
The first column ideally represents the reflectance indices of the front side text in the RGB channels. The second column represent the reflectance indices of the reverse side text, as resulting from paper absorption, assumed different in the three colors. Finally, the third column represents the reflectance indices of the background, equal for the three colors. Clearly, the first output image (Figure 5b) is related to the extracted foreground text, the second output image (Figure 5c) corresponds to the background class, while the third one (Figure 5d) represents the underground bleed-through/show-through. The recovered matrix was in this case:

$$\hat{A} = \begin{bmatrix} 0.1267 & 0.0554 & 0.0864 \\ 0.1271 & 0.0532 & 0.0623 \\ 0.1262 & 0.0520 & 0.0495 \end{bmatrix}$$

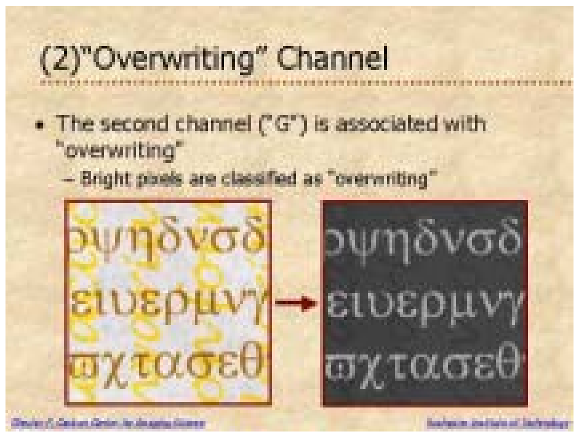
which, as it often happens, differs from the actual one for scale factors and column permutations. In particular, the second and third columns are exchanged, and this is reflected in the order in which the sources are extracted. After restoring the original order, the scale factors can be computed by dividing the two matrices element-by-element:

$$\frac{A}{\hat{A}} = \begin{bmatrix} 1.9495 & 2.0012 & 16.2455 \\ 1.9748 & 2.0144 & 16.9173 \\ 1.9889 & 2.0283 & 17.3077 \end{bmatrix}$$

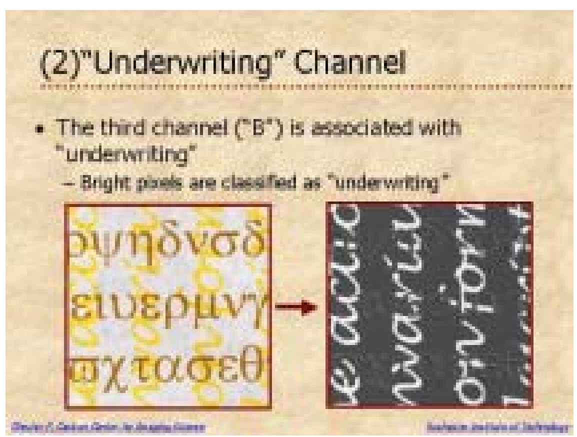
The fact that the scale factor is pretty equal for each column indicates that the separation has been successful. Indeed, a scale factor in a column of the mixing matrix is reflected in an



(a)



(b)



(c)

Figure 4: Treatment of a synthetic example of a palimpsest (from: "Text recovery from the Archimedes Palimpsest", www.cis.rit.edu/people/faculty/easton/k-12/exercise/index.htm): (a) estimated background (b) estimated overwriting; (c) estimated underwriting.

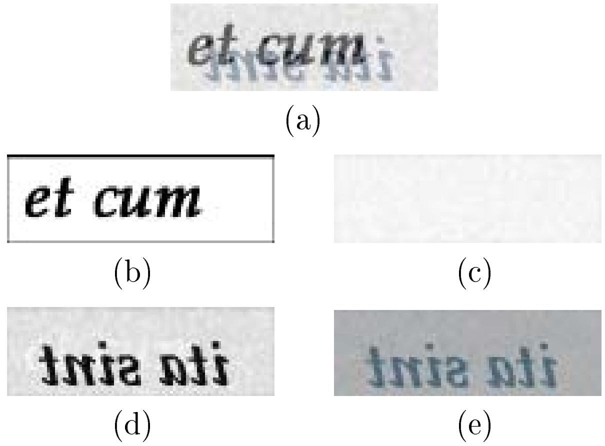
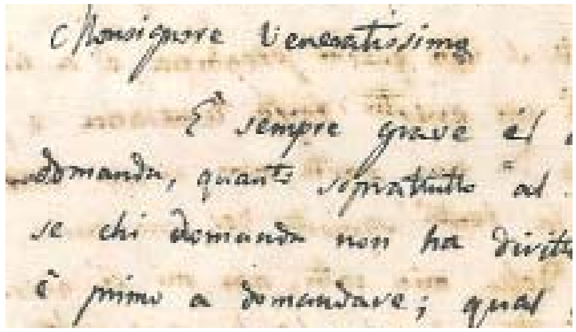


Figure 5: Synthetic image: (a) color image; (b) first FastICA output; (c) second FastICA output; (d) third FastICA output; (e) reconstructed color for the third source.

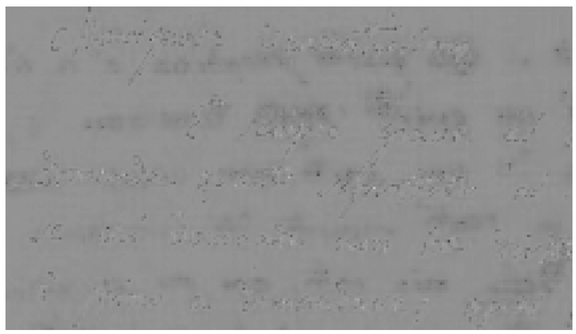
equal, but inverted scale factor in the corresponding sources. Clearly, this does not affect the spatial distribution of pixel values in the source image.

Note that, while in the example of the palimpsest we were able to recover the underwriting only on those zones of the image where it was not masked by the overwriting, in this case it has been possible to recover the bleed-through text also in those points where it was occluded by the foreground text. Indeed, in the palimpsest, at the occlusion points the colors of the two classes do not mix, and the overwriting completely cover the underwriting. Thus, in order to recover the underwriting in those points, an extra observation in the infrared band could be useful. Here, instead, at the occlusion points the colors of the two classes mix each other in perfect accordance with the model adopted. We think that, adopting a model that fits better the physics of the problem, the same result could also be obtained in real situations. In other words, at least in principle and for good quality documents, our method could be able not only to remove, or significantly attenuate, the unwanted bleed-through/show-through effect, but it could also permit its recovery. For example, the show-through effect is very common in the scans of modern, well preserved documents. This could be useful in those cases where the image of the reverse side of a document is not available. With our method, this image can be obtained by simply mirroring the output source related to the show-through.

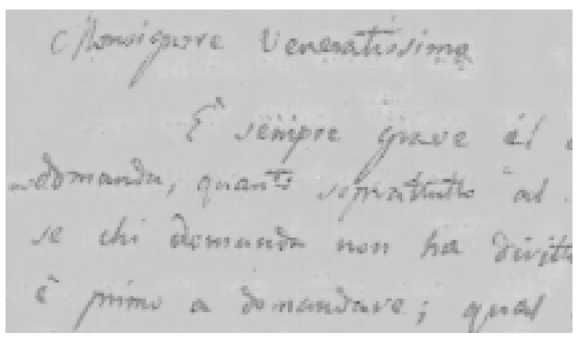
Our second experiment is illustrated in Figure 6. In particular, Figure 6a shows the original color image of an ancient document affected by bleed-through, while Figures 6b and 6c show two of the three separations obtained by using the red, green and blue channels. Also in this case we obtained all the three classes superimposed in the image. The first output image, not shown, is related to the background class, apart from residuals of the other two classes, due to an imperfect separation. Similarly, the second output image corresponds to the underground bleed-through, while the third represents the extracted foreground text. In this case, the recovered bleed-through is completely unreadable, and thus unuseful. This result is quite obvious, keeping in mind that bleed-through is mostly present in ancient, degraded



(a)



(b)



(c)

Figure 6: Real document showing bleed-through: (a) original color image (b) second FastICA output; (c) third FastICA output.

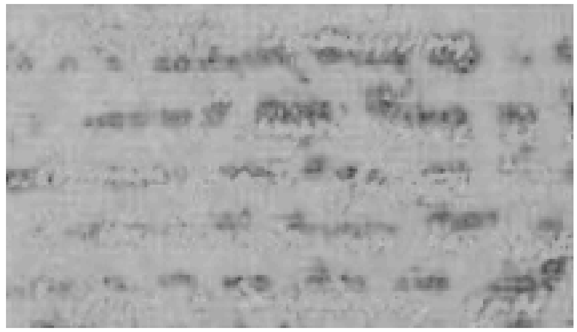


Figure 7: Bleed-through source after post-processing.

documents, and that, being due to spreading of the ink through the support, it appears usually as highly blurred. Some standard post-processing of the image in Figure 6b, such as the median filter and histogram stretching, allowed us to obtain the better quality image of Figure 7. However, we think that, to significantly improve the readability of the bleed-through image, the PSF associated to the ink spreading phenomenon has necessarily to be included into the model, and more sophisticated ICA algorithms have to be used.

6 Conclusions

We presented preliminary results of the application of Independent Component Analysis techniques to the problem of the separation of overlapped texts in documents showing bleed-through or show-through and in palimpsests. Our approach is the first result of a study about the possibility of formulating the problem as a particular kind of Blind Source Separation, which is a well-established discipline in signal processing, but still at the initial stage in the field of image processing and especially of document analysis. Although we adopted a very simplified modellization and standard ICA algorithms, our results are promising. We plan to continue our study in several directions: 1) experimentation of other ICA algorithms, besides FastICA; 2) investigation of more realistic mathematical models (e.g. convolutive or non-linear mixtures) for images of overlapped texts, and related separation algorithms; 3) accounting for auto-correlations of the components, by introducing models of spatial dependence within the single sources; 4) accounting for possible cross-correlations among the component sources and turning towards the emerging field of Dependent Component Analysis; 5) accounting for possible noise in the mixtures.

References

[1] A.J. Bell and T.J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution", Neural Computation, vol.7, pp. 1129-1159, 1995.

[2] T. Lee, M. Lewicki and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources", Neural Computation, vol. 11, pp. 409-433, 1999.

[3] J.-F. Cardoso, "High-order contrasts for independent component analysis", Neural Computation, vol.11, pp. 157-192, 1999.

[4] S. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches", Proc. IEEE, vol. 86, pp. 2026-2048, 1998.

[5] A. Hyvarinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis", IEEE Trans. on Neural Networks, vol.10, pp. 626-634, 1999.

[6] A. Hyvarinen, "Gaussian moments for noisy independent component analysis", IEEE Signal Processing Letters, vol. 6, pp. 145-147, 1999.

[7] H. Attias, "Independent Factor Analysis", Neural Computation, vol. 11, pp. 803-851, 1999.

[8] E. Moulines, J. F. Cardoso, E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models", Proc. ICASSP'97, vol. 5, pp. 3617-3620, 1997.

[9] E. Kuruoglu, L. Bedini, M.T. Paratore, E. Salerno, A. Tonazzini, "Source separation in astrophysical maps using independent factor analysis", Neural Networks, Vol. 16(3-4), pp. 479-491, 2003.

[10] K. Knuth, "Bayesian source separation and localization", Proc. SPIE'98, Bayesian Inference for Inverse Problems, pp. 147-158, 1998.

[11] S. E. Lee and S.J. Press, "Robustness of Bayesian factor analysis estimates", Communications in Statistics - Theory and Methods, vol.27, 1998.

[12] A. Mohammad-Djafari, "A Bayesian approach to source separation", Proc. 19th Int. Workshop on Maximum Entropy and Bayesian Methods, 1999.

[13] A. Tonazzini, L. Bedini, E. E. Kuruoglu and E. Salerno, "Blind separation of time-correlated sources from noisy data", Tech. Rep. IEI-CNR No. TR-42-2001, 2001.

[14] A. Tonazzini, L. Bedini, E. E. Kuruoglu and E. Salerno, "Blind separation of auto-correlated images from noisy mixtures using MRF models", Proc. IV Int. Symp. on Independent Component Analysis and Blind Source Separation (ICA 2003) (1-4 April, 2003, Nara, Japan), 2003.

[15] G. Sharma, "Show-through cancellation in scans of duplex printed documents", IEEE Trans. Image Processing, Vol. 10(5), pp. 736-754, 2001.

[16] H. Nishida, T. Suzuki, "Correction show-through effects in document images by multiscale analysis", Proc. 16th Conf. on pattern Recognition (August 11-15, 2002, Quebec City, Canada), Vol. 3, 2002.

[17] E. Dubois, A. Pathak, "Reduction of bleed-through in scanned manuscript documents", Proc. IS&T Image Processing, Image Quality, Image Capture Systems Conf. (April 2001, Montreal, Canada), pp. 177-180, 2001.

[18] Chew Lim Tan, R. Cao, Peiyi Shen, "Restoration of archival documents using a wavelet technique", IEEE Trans. PAMI, Vol. 24(10), pp. 1399-1404, 2002.

[19] R. L. Easton, "Text recovery from the Archimedes Palimpsest", www.cis.rit.edu/people/faculty/easton/k-12/exercise/index.htm

[20] A. K. Barros, "The independence assumption: dependent component analysis", in: Advances in Independent Component Analysis, M. Girolami ed., Springer-Verlag, London, 2000.

[21] L. Tong, R. W. Liu, V. C. Soon and Y-F. Huang, "Indeterminacy and identifiability of blind identification ", IEEE Trans. on Circuits and Systems, Vol. 38, pp. 499-509, 1991.

[22] A. Hyvarinen et al., "The FastICA Package for MATLAB", www.cis.hut.fi/projects/ica/fastica/