



Plug-and-play approach for blind image separation with application to document image restoration

Xhenis Çoba, Fangchen Feng, Azeddine Beghdadi

► To cite this version:

Xhenis Çoba, Fangchen Feng, Azeddine Beghdadi. Plug-and-play approach for blind image separation with application to document image restoration. [Research Report] L2TI. 2020. <hal-02978489>

HAL Id: hal-02978489

<https://hal.science/hal-02978489v1>

Submitted on 26 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Plug-and-play approach for blind image separation with application to document image restoration

Xhenis Çoba, Fangchen Feng, Azeddine Beghdadi

Laboratoire de Traitement et Transport de l'Information (L2TI)

Université Sorbonne Paris Nord

Villetaneuse, France

xhenis.coba@etu.univ-grenoble-alpes.fr, fangchen.feng@univ-paris13.fr, azeddine.beghdadi@univ-paris13.fr

Abstract—In this paper we propose a new method for document image restoration based on Blind Source Separation. The existing separation methods rely on the general properties of source images such as independence, sparsity, and non-negativity. In this work we show that by exploiting some characteristics of image denoising methods in a play-and-plug scheme, efficient BSS results could be achieved. In particular, we show that the BM3D and Non-local Means denoising methods, which exploit the non-local property, lead to better image separation in terms of the convergence rate. We then propose to use the dictionary-learning approach to accelerate the algorithm which also takes the visual chirality into consideration. We finally apply the proposed approaches to the document image restoration problem and show the advantages of the proposed approaches through some experiments and objective performance evaluation of the proposed scheme.

Index Terms—Blind image separation, Document restoration, Plug-and-play, Image denoising, Dictionary learning

I. INTRODUCTION

Old documents often suffer from the problem of the show-through and bleed-through effect (see the middle images in Fig. 1 as an illustration). Show-through is a front-to-back interference, mainly due to the scanning process and the transparency of the paper, which causes the text in the verso side of the document to appear also in the recto side (and vice versa) [1]. Bleed-through is an intrinsic front-to-back physical deterioration due to ink seeping and produces an effect similar to that of show-through [2]. Despite the complex physical model of the process which depends on the features of the paper, the transmittance parameters, the reflectance of the verso, and the spreading of light in the paper [3]–[7], the following nonstationary linear model is well adapted to these effects locally [2]:

$$\mathbf{x}[i] = A\mathbf{s}[i], \quad (1)$$

where i is the index of the image pixels, \mathbf{s} is the concatenation of the source images in grayscale $\mathbf{s}[i] = [s_r[i], s_v[i]]^T$ where $s_r[i]$ and $s_v[i]$ are the recto and verso sides (with a horizontal flip) of the source images respectively, $\mathbf{x}[i] = [x_r[i], x_v[i]]^T$ with $x_r[i]$ and $x_v[i]$ being the observed recto and verso side images and $A \in \mathbb{R}^{2 \times 2}$ is the mixing matrix. The goal of the blind document restoration is to recover the source images $\mathbf{s}[i]$ from the observed recto-verso mixtures $\mathbf{x}[i]$ without knowing

the mixing matrix A . This problem falls in the general framework of Blind Source Separation (BSS) [8].

Among classical methods that aim at creating independent output components from an input signal, the ICA [9] could be considered as a BSS solution. It assumes that the sources are independent and the key idea is to look for a demixing matrix such that the separated components are as independent as possible [10]. Despite the good results, these kinds of solutions do not work if more than one of the sources follow the Gaussian distribution since ICA-based methods rely on the higher-order statistics. Sparsity Component Analysis (SCA) is another well studied approach for BSS [11]–[13]. This approach assumes that the sources are sparse in the spatial or in a transform domain. The separation problem is then formulated in an optimization framework and the estimation of the mixing matrix and the source images are performed simultaneously. One of the advantages of the SCA compared to the ICA is that it leads to better results if the mixtures are degraded with additive noise [13]. It is also important to notice that several works have been dedicated to exploiting the links between the ICA and SCA methods [14]–[16]. Besides the independence and sparsity properties of the source images, other criteria have also been investigated for source separation such as Non-negative Matrix Factorization (NMF) [17] and Canonical Correlation Analysis (CCA) [2] for document image restoration.

For blind image separation, in the determined scenario ¹, the demixing matrix is estimated based on different source image characteristics. On the other hand, image denoising methods seek to remove perturbations or errors from the observed images based on the same image characteristics. Extensive research has been carried out on image denoising over the past three decades, which has led to highly optimized algorithms [18]–[21]. In this paper, instead of searching for other characteristics for blind image separation, we propose to use the existing denoising approaches in a plug-and-play way for the source separation. In particular, we show that the BM3D [20] and the Nonlocal Mean [18] denoising methods lead to better separation than the existing approaches in terms of the number of iterations. To the best of our knowledge this paper introduces for the first time the non-local image

¹The number of observations equals the number of sources.

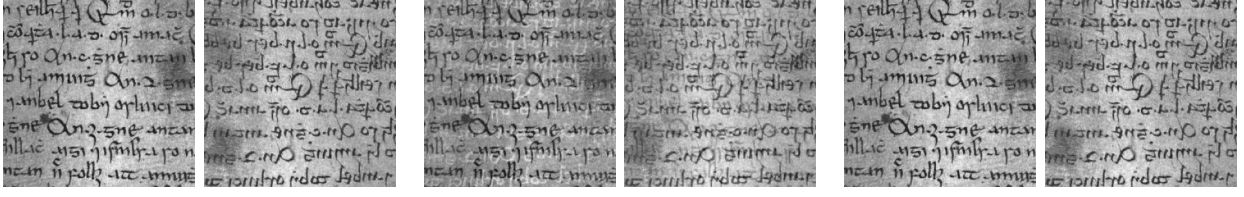


Fig. 1: From left to the right by pairs: the original sources, the mixed images and the estimated images using Double Dictionary Learning. See Section IV for implementation details.

characteristics for blind image separation.

The main contributions of this paper are the following:

- A blind source separation framework, that encompasses several existing approaches, is developed and described in the context of document image restoration.
- The idea of using nonlocal characteristics of the image signal in this BSS scheme is proposed for the first time.
- A dictionary learning strategy to accelerate the algorithms which improves the separation by considering the visual chirality of text images is proposed.

The reminder of this paper is organized as follows. In Section II, we describe the proposed separation method with plug-and-play approach. We describe the proposed dictionary learning strategy in Section III. The performance evaluation of the proposed method is presented in Section IV. Section V is dedicated to conclusion and some open problems.

II. PLUG-AND-PLAY APPROACH FOR IMAGE SEPARATION

A. Optimization framework

Both ICA and SCA schemes in the determined scenario can be formulated in an optimization framework [14] as follows:

$$\operatorname{argmin}_{W, \mathbf{s}} \frac{1}{2} \|W\mathbf{x} - \mathbf{s}\|_F^2 + \sigma \mathcal{P}(\mathbf{s}) + \mathcal{G}(W), \quad (2)$$

where W is the demixing matrix. $\|\cdot\|_F$ denotes the Frobenius norm of matrix. $\mathcal{P}(\mathbf{s})$ is the source penalty term based on the source image characteristics. σ is the hyperparameter that is a tradeoff between the data term and the source penalty term. $\mathcal{G}(W)$ is the constraint of the demixing matrix. For ICA approaches, $\mathcal{P}(\mathbf{s})$ is the independence measurement for the sources and $\mathcal{G}(W)$ is the unitary constraint which leads to the source decorrelation with the whitening pre-processing. For SCA, $\mathcal{P}(\mathbf{s})$ is the sparse penalty in the spatial or a transformed domain and $\mathcal{G}(W)$ is the unit-norm constraint of the mixing matrix to regularize the scaling ambiguity between the mixing matrix and the source images.

The above optimization problem (2) in the general form can be solved with the iterative alternating minimization [22]. In each iteration, the following sub-problems are solved in an alternating way until convergence:

$$\operatorname{argmin}_{\mathbf{s}} \frac{1}{2} \|W\mathbf{x} - \mathbf{s}\|_F^2 + \sigma \mathcal{P}(\mathbf{s}), \quad (3)$$

$$\operatorname{argmin}_W \frac{1}{2} \|W\mathbf{x} - \mathbf{s}\|_F^2 + \mathcal{G}(W). \quad (4)$$

The sub-problem (3) is directly related to the classical denoising problem for additive Gaussian noise which motivates us to use the plug-and-play approach. The plug-and-play [23]–[25] approach consists of replacing the denoising formulation by existing denoising algorithms in a general sense. It is worth noticing that the considered denoising algorithms do not necessarily correspond to any penalty term in a closed form as in (3). In this paper, we illustrate the separation results using Total-Variation (TV) minimization [26], BM3D [20] and Non-local mean [18] denoising approaches.

In the denoising scheme, the hyperparameter σ is directly related to the noise variance (see Fig. 2 for illustrations). Note that, since we do not consider additive noise in the mixing model (1), the hyperparameter in the proposed framework only plays the role of balancing the data term $W\mathbf{x}$ and the source penalty term $\mathcal{P}(\mathbf{s})$. Theoretically, a small σ should be chosen for the noiseless mixtures. However, as we will show in the experiment section, that the appropriate choice of σ leads to a rapid separation.

The sub-problem (4) can be solved with the gradient descend approach with the unitary projection. The proposed algorithm is summarized as follows where Denoise_σ is any aforementioned denoising method with a fixed hyperparameter σ . \mathcal{U} is the unitary projection.

Algorithm 1: Plug-and-play for separation

```

Whiten the mixture  $\mathbf{x}$ ;
Initialization :  $W \in \mathbb{R}^{2 \times 2}$ ,  $\mathbf{s} \in \mathbb{R}^{2 \times I}$ ;
repeat
    Update the images with a denoising method:
     $\mathbf{s} \leftarrow \text{Denoise}_\sigma(\mathbf{s})$ ;
    Update the mixing matrix by Least-Square
    followed by the unitary projection:
     $W \leftarrow \mathcal{U}(\mathbf{s}\mathbf{x}^T)$ ;
until convergence;
```

B. Related BSS methods

The proposed framework encompasses several existing source separation approaches. When the TV denoiser is used, the proposed approach is related to the gradient-based sparsity

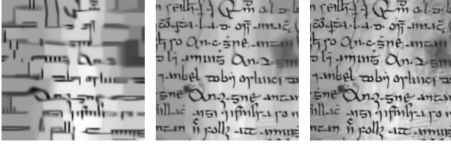


Fig. 2: The result of BM3D denoising of the original recto image in Fig. 1 with different σ . From left to right: $\sigma = 2 \times 10^{-2}$, 8×10^{-3} , 5×10^{-3} . BM3D is capable of capturing image characteristics with adapted σ (5×10^{-3} in this case). A too large σ "smooths" too much the image destroying the details.

method in [11]. The proposed framework is not limited to the aforementioned denoisers. When the dictionary learning denoiser [27] is used, the proposed approach is closely related to the method proposed in [28].

III. DICTIONARY LEARNING FOR VISUAL CHIRALITY

Although BM3D and Non-local mean denoisers lead to better separation, as they exploit the non-local characteristics of images, they suffer computational burden due to their complexity. In order to solve this problem, we propose to use the dictionary learning procedure. More precisely, we propose to learn a dictionary from a document image dataset filtered by BM3D and/or Non-local mean denoiser with a certain predefined hyperparameter σ , and then perform the denoising based on the learned dictionary.

Another advantage of using the dictionary learning as the denoiser is because it can capture the chirality property of text images. For document image restoration from scanned recto-verso document pages, the two original document images are mixed with one of them horizontally flipped. It has been shown recently in [29] that text is a strong signal for high-level visual chirality which is a fundamental property of images. Classical denoisers (TV, BM3D and Non-local mean) are not sensitive to such image property and we show in the experiment section that the dictionary-learning [27] denoiser can well capture the chirality property. More precisely, a dictionary is first learned from a document images dataset where all the images are in the readable direction. The dictionary is then used as the denoiser in the proposed algorithm where the estimated verso image is horizontally flipped in each iteration so that both the estimated images are in the readable direction before the denoising process. We refer to this approach as the Double Dictionary Learning. For the comparison purpose, we also show the image separation results with dictionary learning where such flipping is not applied. We refer to this approach as Single Dictionary Learning which is a baseline

IV. NUMERICAL EVALUATION

In this section, we evaluate the proposed source separation framework. We first evaluate the proposed approach with different parameters of the algorithm in different mixing scenarios. We then concentrate on the old document restoration application.

A. Experimental setup

All the experiments are performed with synthesized image mixtures. We use gray-scale old document images of size 256×256 from the dataset [30]. The image sources are then mixed using a 2×2 rotation mixing matrix according to the model (1) to obtain the mixtures. For the separation, a Principle Component Analysis (PCA) is performed to whiten the mixture and the proposed separation algorithm is initialized with the whitened mixtures. We set the stop criteria as $\|\mathbf{s}^{(k+1)} - \mathbf{s}^{(k)}\|_F^2 < 10^{-8}$ where $\mathbf{s}^{(k)}$ is the estimated source image of the k -th iteration. We also limit the maximum number of iterations to 100.

To better show the performance of the separation, we evaluate both the estimated source images and the estimated demixing matrix. More precisely, we use the Structural Similarity Index Measure (SSIM) [31] to evaluate the perceptual closeness of the estimated images to the original ones. This measure takes values between 0 and 1 and higher SSIM indicates a better performance. Note that for all the experiment, we show the mean value of the SSIM of the two estimated recto and verso source images. To evaluate the demixing matrix, we consider the following Diagonal Ratio (DR):

$$\text{DR}(\hat{W}) = \frac{\sum |[\hat{W}A]_{\text{off}}|}{\sum |[\hat{W}A]_{\text{diag}}|}, \quad (5)$$

where \hat{W} is the estimated demixing matrix. $[\hat{W}A]_{\text{off}}$ and $[\hat{W}A]_{\text{diag}}$ are the off-diagonal and diagonal element of the matrix $\hat{W}A$ respectively. This measure takes values between 0 and 1 and a lower value indicates a better separation result.

B. Separation performance with hyperparameters

The hyperparameter σ in the Algorithm 1 is a trade-off between the data term and the source penalty term. A large σ weights the penalty term, while a small σ penalizes the data term. In this subsection we illustrate the separation performance of the proposed approach with different denoisers and values of σ . We show the results of using BM3D and TV as the denoiser and use the rotation mixing matrix with an angle $\theta = 30^\circ$. Fig. 3 depicts the separation evaluations as a function of the number of iterations.

One can remark that the choice of σ affects the performance of the algorithm significantly. For TV, $\sigma = 0.3 \times 10^{-2}$ gives us a convergence in the first 25 iterations, while for $\sigma = 0.3 \times 10^{-3}$, another 35 iterations are needed. Similar behaviour is observed over all the proposed denoisers.

In the following, for all the tested denoisers, we use the σ that leads to their best performance respectively.

C. Separation performance for different mixing conditions

In the determined scenario, the linear mixing process can be seen as a rotation of the two sources. Indeed, with the whitening processing of the mixtures, the mixing matrix can be reduced to a rotation matrix as follows:

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (6)$$

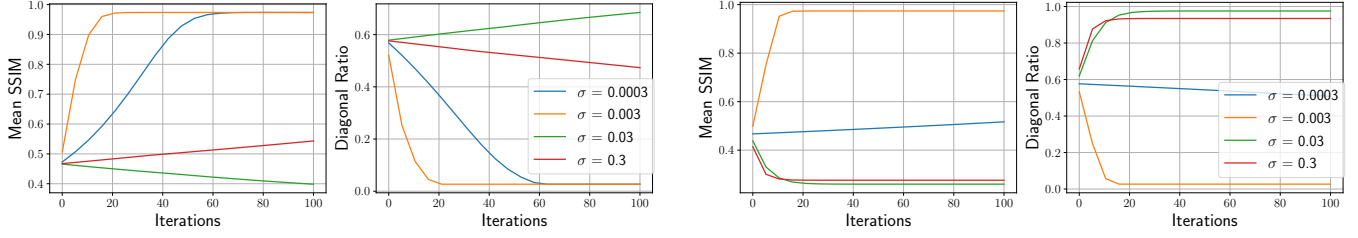


Fig. 3: Mean SSIM and DR for TV (two leftmost) and BM3D (two rightmost) denoising approach using different σ

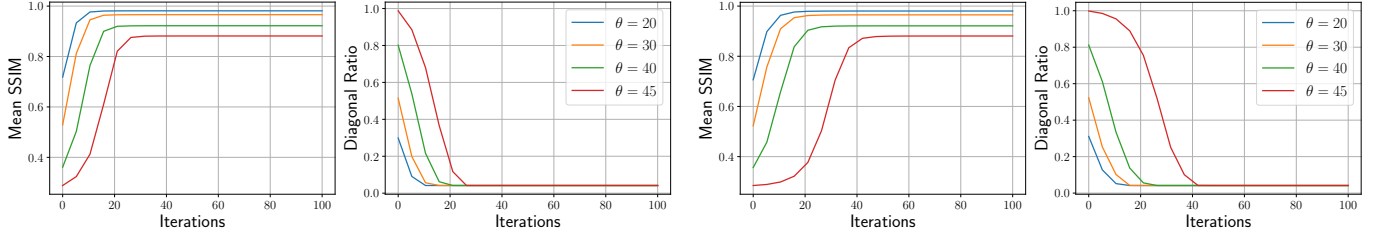


Fig. 4: Mean SSIM and DR for Double Dictionary Learning (two leftmost) and TV denoising (two rightmost) methods using different θ

where θ is the rotation angle and a large rotation angle indicates a "deep" mixture. In this subsection, we test the proposed approach with TV, BM3D, and Non-local mean denoiser for different mixing conditions with different θ . More precisely, we consider $0^\circ \leq \theta \leq 45^\circ$. Note that testing the proposed approach with a θ outside this range do not provide additional information due to trigonometric properties. Fig. 4 shows the separation results as a function of the iteration.

One can remark that, for large θ , more iterations are needed to reach the convergence for all the tested denoisers. It is also interesting to notice the difference between the mean SSIM for different mixing conditions which is consistent with the expectation that higher θ yields hard-to-separate mixtures. Such difference is not visible in Diagonal Ratio (DR) and this shows that the mean SSIM is a more sensitive measure and illustrates the importance of using multiple evaluation methods.

D. Separation for document image restoration

In this subsection, we compare the performance of several denoising techniques within the plug-and-play framework for old image document separations. The mixed images are obtained by using the rotation matrix with $\theta = 45^\circ$ as the case which needs more iterations to reach convergence. In this case, FastICA does not perform well (Mean SSIM $\leq 0.89 \times 10^{-3}$). Fig. 5 depicts the performance of the proposed methods. It can be noticed that Double Dictionary Learning and the Non-local approach are clearly the two methods that yield the fastest convergence. Fig. 1 demonstrate the results obtained by Dictionary Learning. However, Non-local denoising takes a considerable amount of time compared to Double Dictionary Learning, making the latter preferable in terms of computa-

tional time. Moreover, Double Dictionary Learning performs better than Single Dictionary Learning, emphasizing this way the chiral properties of document images.

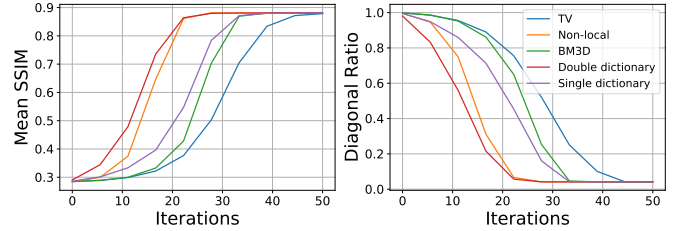


Fig. 5: Separation performance of the proposed approach with TV, Non-local, BM3D, Single and Double dictionary learning.

V. CONCLUSION AND FUTURE WORK

In this paper, we demonstrate that existing denoising methods can be adapted to blind image separation problems. The use of TV minimization, BM3D, and Non-local Mean denoisers lead to promising results in the context of old documents. Moreover, dictionary learning methods provide faster computational time while maintaining a high quality of separation because of the chiral effects of document images.

We plan to expand this framework for color images as this poses a challenge on itself. Furthermore, besides the experiments on synthetic images, the next step would be to use real-life old document images with show-through and bleed-through effect. Appropriate separation quality assessment has to be taken into consideration for this purpose.

REFERENCES

- [1] Gaurav Sharma, "Show-through cancellation in scans of duplex printed documents," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 736–754, 2001.
- [2] Antonio Boccuto, Ivan Gerace, and Valentina Giorgetti, "A blind source separation technique for document restoration," *SIAM Journal on Imaging Sciences*, vol. 12, no. 2, pp. 1135–1162, 2019.
- [3] Anna Tonazzini, Pasquale Savino, and Emanuele Salerno, "A non-stationary density model to separate overlapped texts in degraded documents," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 155–164, 2015.
- [4] Francesca Martinelli, Emanuele Salerno, Ivan Gerace, and Anna Tonazzini, "Nonlinear model and constrained ML for removing back-to-front interferences from recto-verso documents," *Pattern Recognition*, vol. 45, no. 1, pp. 596–605, 2012.
- [5] Reza Farrahi Moghaddam and Mohamed Cheriet, "Low quality document image modeling and enhancement," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 11, no. 4, pp. 183–201, 2009.
- [6] Mariana SC Almeida and Luís B Almeida, "Wavelet-based separation of nonlinear show-through and bleed-through image mixtures," *Neurocomputing*, vol. 72, no. 1-3, pp. 57–70, 2008.
- [7] Anna Tonazzini, Ivan Gerace, and Francesca Martinelli, "Multichannel blind separation and deconvolution of images for document analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 912–925, 2009.
- [8] Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.
- [9] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent component analysis*, vol. 46, John Wiley & Sons, 2004.
- [10] Anna Tonazzini, Luigi Bedini, and Emanuele Salerno, "Independent component analysis for document restoration," *Document Analysis and Recognition*, vol. 7, no. 1, pp. 17–27, 2004.
- [11] Wided Souidène, Abdeldjalil Aïssa-El-Bey, Karim Abed-Meraim, and Azeddine Beghdadi, "Blind image separation using sparse representation," in *2007 IEEE International Conference on Image Processing*. IEEE, 2007, vol. 3, pp. III–125.
- [12] Michael Zibulevsky and Barak A Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.
- [13] Jérôme Bobin, J-L Starck, Jalal Fadili, and Yassir Moudden, "Sparsity and morphological diversity in blind source separation," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2662–2674, 2007.
- [14] Fangchen Feng and Matthieu Kowalski, "Revisiting sparse ICA from a synthesis point of view: Blind source separation for over and underdetermined mixtures," *Signal Processing*, vol. 152, pp. 165–177, 2018.
- [15] Alexander M Bronstein, Michael M Bronstein, Michael Zibulevsky, and Yehoshua Y Zeevi, "Sparse ICA for blind separation of transmitted and reflected images," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 84–91, 2005.
- [16] Zois Boukouvalas, Yuri Levin-Schwartz, Vince D Calhoun, and Tülay Adalı, "Sparsity and independence: Balancing two objectives in optimization for source separation with application to fMRI analysis," *Journal of the Franklin Institute*, vol. 355, no. 4, pp. 1873–1887, 2018.
- [17] Farnood Merrikh-Bayat, Massoud Babaie-Zadeh, and Christian Jutten, "Using non-negative matrix factorization for removing show-through," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 482–489.
- [18] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, vol. 2, pp. 60–65.
- [19] Wided Souidene, Azeddine Beghdadi, and Karim Abed-Meraim, "Image denoising in the transformed domain using non local neighborhoods," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. IEEE, 2006, vol. 2, pp. II–II.
- [20] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [21] Quoc Bao Do, Azeddine Beghdadi, and Marie Luong, "Combination of closest space and closest structure to ameliorate non-local means method," in *2011 IEEE Symposium On Computational Intelligence For Multimedia, Signal And Vision Processing*. IEEE, 2011, pp. 134–141.
- [22] Paul Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.
- [23] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.
- [24] Stanley H Chan, Xiran Wang, and Omar A Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2016.
- [25] Shunsuke Ono, "Primal-dual plug-and-play image restoration," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1108–1112, 2017.
- [26] Antonin Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1-2, pp. 89–97, 2004.
- [27] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [28] Xiaochen Zhao, Guangyu Zhou, Wei Dai, and Wenwu Wang, "Blind source separation based on dictionary learning: a singularity-aware approach," in *Blind Source Separation*, pp. 39–59. Springer, 2014.
- [29] Zhiqiu Lin, Jin Sun, Abe Davis, and Noah Snavely, "Visual chirality," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12295–12303.
- [30] Róisín Rowley-Brooke, François Pitié, and Anil Kokaram, "A ground truth bleed-through document image database," in *International Conference on Theory and Practice of Digital Libraries*. Springer, 2012, pp. 185–196.
- [31] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.