TRIDENT: A Trimodel with Sparse Dictionary Encoding Technique applied to Ringing Stripe Artifact Removal from Medical X-ray Images

Matthieu Chancel
GIPSA-Lab
Univ. Grenoble Alpes, CNRS, Grenoble INP
Grenoble, France
matthieu.chancel@grenoble-inp.fr

Hacheme Ayasso
GIPSA-Lab
Univ. Grenoble Alpes, CNRS, Grenoble INP
Grenoble, France
hacheme.ayasso@grenoble-inp.fr

Michel Desvignes
GIPSA-Lab
Univ. Grenoble Alpes, CNRS, Grenoble INP
Grenoble, France
michel.desvignes@grenoble-inp.fr

Jean-Michel Vignolle
Image Group
TRIXELL
Moirans, France
jean-michel.vignolle@trixell-thalesgroup.com

Eric Lespessailles
Department of Rheumatology
Centre Hospitalier Universitaire Régional d'Orléans
Orléans, France
eric.lespessailles@chu-orleans.fr

***Abstract*—We present, in this work, a new approach for artifact extraction in digital radiography. We propose to decompose the observed image into two components: exclusively clinical and joint. The latter component is then represented by two sparse dictionaries to account for the artifact and the residual clinical content. Both dictionaries are learned to maximize the performance of the method. Application on ringing stripe artifact extraction over a large dataset of medical X-ray images, show better performance for our method in comparison to unimodel and bimodel approaches.**

***Index Terms*—X-ray, Sparse Dictionary Learning (SDL), Joint Modeling.**

I. Introduction

Under extreme condition, X-ray radiography can suffer from several cosmetic artifacts (deterministic or stochastic) that impair proper analysis of their medical content [1], [2] such as ringing stripe artifacts (Fig1).

A first approach to remove the artifacts consists in using a single model (unimodel) for the artifact for its extraction [2]–[4]. These methods are relatively fast. However, they lack a proper model for the clinical content. Thus, they can remove vital clinical information because of shared features with the artifact.

A second approach proposes a bimodel for the clinical content and the artifact as in [5] and [6]. The later works propose joint sparse dictionary representation. This allows a better estimation of the artifact without degrading the clinical content. However, a relatively big dictionary is necessary for the clinical content which is computationally intensive.

The aim of this work is to introduce a **TRI**model with sparse **D**ictionary **EN**coding **T**echnique: **TRIDENT**, where the observed image is represented by two components. The first, the Exclusively Clinical Component (ECC), is represented by a basic model that accounts for features only present in the medical content. The second, Joint Component (JC), is composed of a mixture between the artifact and the residual of the clinical content. The joint component is represented by a joint sparse dictionary model. The advantage of this approach is that the dictionary of the residual clinical content is smaller and the estimation runs faster.

We apply TRIDENT to the extraction of the ringing stripe artifacts where several sequences of bright and dark lines add to the image at random positions (Fig.1). Their intensities vary across the columns in a random oscillating pattern (ringing).

II. Observation Model

The stripe artifacts $\mathcal{A} \in \mathbb{R}^n$ are considered additive to the clinical content $\mathcal{C} \in \mathbb{R}^n$. Thus, the image $\mathcal{Y} \in \mathbb{R}^n$ observation model reads:

$$\mathcal{Y} = \mathcal{C} + \mathcal{A} \tag{1}$$

Where n is the number of pixels. It is generally large ($n \sim 10^6$ pixels) which leads to computationally intensive methods. Therefore, we adopt in this work a patch-based approach where we divide each image $\mathcal{Y}$ into several patches $y_i \in \mathbb{R}^{n_p}$:

$$y_i = c_i + a_i \tag{2}$$

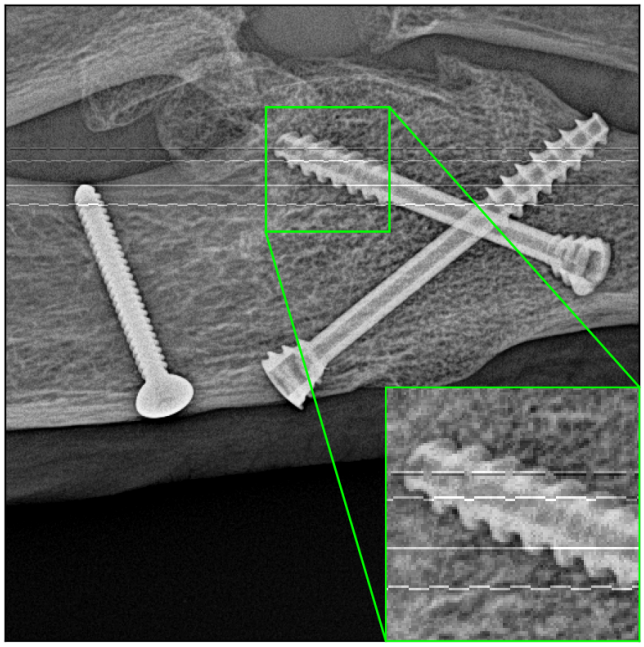
With $n_p = p_r \times p_c$ is the number of pixels in a patch, p_r is the number of rows and p_c is the number of columns.

A. Unimodel

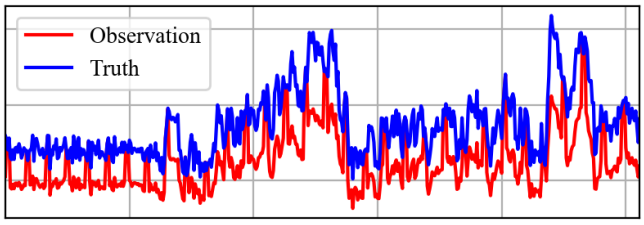
Filtering approaches such as [2]–[4] focus on estimating $\mathcal{A}$ based on prior knowledge only about the artifact, without any model for the clinical content. Then, a notch filter is designed to eliminate the frequency generated by the artifact:

$$\tilde{\mathcal{C}} = \mathrm{filt}_n(\mathcal{Y}) \tag{3}$$

However, the filter removes the medical content frequencies in the same interval.



(a)



(b)

Fig. 1. (a): Radiography of a foot with different types of ringing stripe artifacts (b): Profile over a row with a ringing stripe artifact and the ground truth.

B. Bimodel sparse dictionary representation

Unimodel approach is insufficient, since it lacks a model for the clinical content. Therefore, several works, as [5] and [6], propose a joint sparse dictionary representation for c_i and a_i:

$$y_i = D_C x_{c_i} + D_A x_{a_i} + \epsilon = D x_i + \epsilon \tag{4}$$

With $D_C \in \mathbb{R}^{n_p \times k_C}$ and $D_A \in \mathbb{R}^{n_p \times k_A}$, the dictionaries for the clinical content and the artifact respectively. $x_{c_i} \in \mathbb{R}^{k_C}$ and $x_{a_i} \in \mathbb{R}^{k_A}$ are the corresponding weight vectors. These can be combined in one dictionary with $D = \left[D_C \mid D_A\right] \in \mathbb{R}^{n_p \times (k_C + k_A)}$ the concatenation of D_C and D_A and $x_i = \left[x_{c_i} \mid x_{a_i}\right]^T \in \mathbb{R}^{k_C + k_A}$. ϵ stands for the representation errors.

x_i is supposed sparse and only a few elements are nonzero. Thus, solving equation (4) is done using sparse representation algorithms such as pseudo-ℓ_0 norm algorithms [7], [8] or ℓ_1 norm [9] or [10] imposing a limit of nonzero elements $L_0 \in \mathbb{N}^*$ in x_i that is:

$$\hat{x}_i = \arg\min_{x_i} \|Dx_i - y_i\|_2^2 \ \text{ s.t } \ \|x_i\|_0 \leq L_0 \tag{5}$$

C. Trimodel with exclusive clinical component and sparse dictionary representation for the joint component

The model proposed in (4) tries to represent all the clinical features in D_C which results in a big dictionary. We propose to simplify the later model by separating the clinical content into two parts: $\mathcal{C}_e$ which contains exclusive clinical features and $\mathcal{C}_r$ the residual component. We rewrite the observation model for a patch i:

$$y_i = c_{e_i} + c_{r_i} + a_i \tag{6}$$

Then, we represent c_{r_i} and a_i by two sparse dictionary models

$$y_i = c_{e_i} + D_{C_r} x_{c_{r_i}} + D_A x_{a_i} + \epsilon = c_{e_i} + D_r x_i + \epsilon \tag{7}$$

The principal motivation is that we want to discriminate the artifact only for ambiguous areas and remove from the clinical dictionaries all features that are unambiguous with respect to the artifact. This leads to smaller clinical dictionaries and less decomposition error because it is done only over the common component.

The choice for the estimation of the $\mathcal{C}_e$ is the key of the quality of the artifact correction. For better extraction, $\mathcal{C}_e$ should represent the clinical information without capturing the artifact.

III. Proposed Methodology

A. Exclusive Clinical Content Estimation

We view the exclusive clinical component (ECC) $\mathcal{C}_e$ estimation as a rough approximation of the clinical content. We handle it in a filtering framework as in a unimodel approach. The filter should be designed without being porous to the artifact. It must remove most of the clinical content. Therefore, it is designed as a tradeoff between the quantity of clinical features capture and its non-porous properties to the artifact.

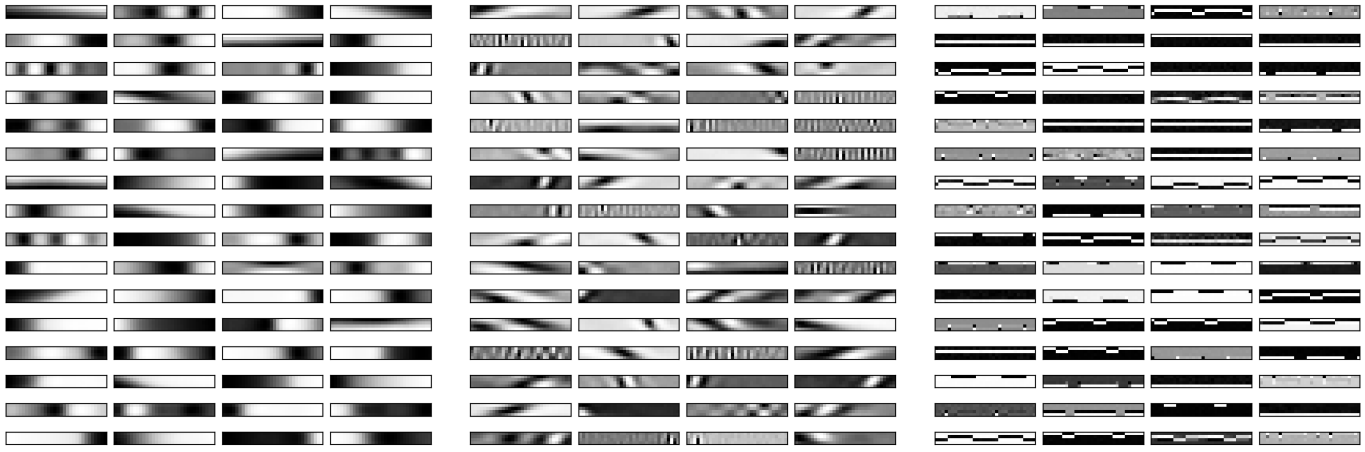
In this article, we choose the moving median to be the best tradeoff. It has the ability to capture a lot of complex clinical features (sharp edges as well as low-frequency features) and needs small computation time.

The dimension of the median kernel must be defined with respect to the artifact features. In the case of the ringing stripe artifact, we choose a rectangular window with more rows than columns.

B. Sparse Dictionary Learning

The residual clinical dictionary D_{C_r} and the artifact dictionary D_A are a crucial component of our model. They can be obtained by design or learning. However, learned sparse dictionaries have shown their superiority compared to fixed dictionaries due to the variety of the clinical content and the artifacts [11], [12]. Several algorithms allow learning sparse dictionaries [13], [14], [15]. We adopt an alternating minimization learning scheme for D_{C_r} and D_A.

Let $D_l \in \mathbb{R}^{n_p \times k}$ be the dictionary of k atoms to learn over a database stored in $Y \in \mathbb{R}^{n_p \times S}$ (each y_i is a sample from the database stored as a vector in Y and S the number of samples).



(a) (b) (c)

Fig. 2. (a): Clinical dictionary learned without the removal of the ECC, (b): Clinical dictionary learned with the removal of ECC, (c): Artifact learned dictionary (only 64/256 atoms showed for each dictionary).

By fixing $X^{(t)}$ as the current sparse solution minimizing the error between Y and $D_l X$ at iteration t, we write:

$$D_l^{(t+1)} \triangleq \underset{D_l \in \mathcal{S}}{\operatorname{argmin}} \|Y - D_l X^{(t)}\|_F^2 \tag{8}$$

With $D_l^{(t+1)}$ the updated dictionary, $\| \bullet \|_F$ denoting the Frobenius norm, $\mathcal{S}$ a constraint set.

We use (8) for both artifact D_A and the residual clinical content dictionary learning D_{C_r}. When learning D_{C_r}, the learning data is preprocessed by subtracting $\mathcal{C}_e$ from the training samples.

C. Proposed Algorithm

Using the image formation model defined in equation (7), we propose to estimate and remove the artifact $\mathcal{A}$. First, $\mathcal{C}_e$ is subtracted from the image. Then, $\hat{x}_{c_{r_i}}$ and $\hat{x}_{a_i}$ are jointly computed using (5) and the artifact estimation is $\tilde{a}_i = D_A \hat{x}_{a_i}$. Finally, the clinical content is obtained by subtracting the artifact (i.e. $\tilde{\mathcal{C}} = \mathcal{Y} - \tilde{A}$). Algorithm 1 presents TRIDENT in detail.

Algorithm 1 TRIDENT Algorithm

Require: $\mathcal{Y}$ (input image), $D_r = \left[D_{C_r} \mid D_A\right]$ (dictionaries), L_0 (sparsity limit)

1: $\tilde{\mathcal{A}} \leftarrow \vec{0}$

2: Estimate $\mathcal{C}_e$ from $\mathcal{Y}$: $\mathcal{C}_e \leftarrow \text{filt}(\mathcal{Y})$

3: Remove the $\mathcal{C}_e$ from $\mathcal{Y}$: $\mathcal{Y}_r \leftarrow \mathcal{Y} - \mathcal{C}_e$

4: **for** $i = 1$ to P **do**

5: Extract the i^{th} patch from $\mathcal{Y}_r$:

$$y_i \leftarrow R_i \mathcal{Y}_r$$

6: Solve image representation model (7):

$$\hat{x}_i = \left[\hat{x}_{c_{r_i}} \mid \hat{x}_{a_i}\right]^T = \arg\min_x \|D_r x_i - y_i\|_2^2 \text{ s.t } \|x_i\|_0 \leq L_0$$

7: Estimate the artifact and restore it in 2D :

$$\tilde{a}_i = D_A \hat{x}_{a_i}$$

$$\tilde{\mathcal{A}} \leftarrow \tilde{\mathcal{A}} + R_i^{-1} \tilde{a}_i$$

8: **end for**

9: Correct the input image from the artifact:

$$\tilde{\mathcal{C}} = \mathcal{Y} - \tilde{\mathcal{A}}$$

10: **return** $\tilde{\mathcal{C}}$ (estimated clinical content over the image)

IV. Application

A. Overview

We evaluate our method on several types of ringing stripe artifact composed of 4 intensity profiles (presented in Fig.1). Then $\mathcal{Y}$ is obtained by adding the artifact to a real X-ray image extracted from a dataset of feet and hands provided by the Centre Hospitalier Régional d'Orléans (under the patient cohort described in Acknowledgment).

We compared our proposition to two different methodologies adapted to remove this artifact since we are unaware of any existing solution in the literature. We compared to a unimodel adapted from [4] and a bimodel sparse dictionary method with learned artifact dictionary [5], [6] denoted "Unimodel" and "Bimodel" respectively.

We extract horizontal patches $p_r = 4, p_c = 32$ with a half overlap from the image to account for the strong horizontal aspect of the artifact. We use the moving median filter with 11 rows and 3 columns window to obtain $\mathcal{C}_e$.

We learned 2 clinical dictionaries D_C and D_{C_r} over 1 million samples randomly extracted from the aforementioned dataset with both the same parameters: $k_C = 256$ components and $L_{0_C} = 3$ maximum nonzero components to use for representation. The full clinical content dictionary D_C (Fig.2.a) learns most of the clinical features but lacks precision: only low frequency features learned and unambiguous with the artifact (low vertical features for example).

On the other hand, the residual dictionary D_{C_r} (Fig.2.b)

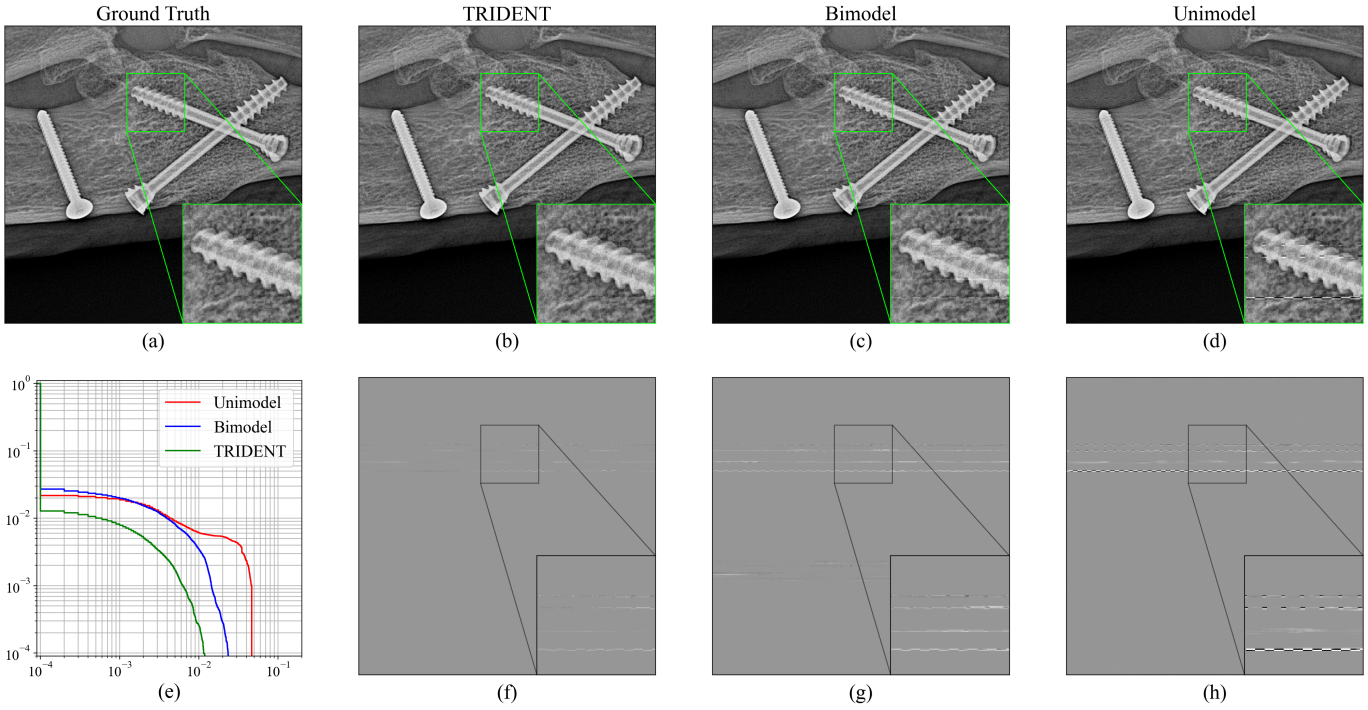


Fig. 3. Comparison of each method (arranged by columns) applied to a ringing stripe artifact removal from an X-ray image of a foot with medical screws. First row compares the output correction (zoom in the bottom right corner) to the ground truth (a), second row the clinical content error (zoom in the bottom right corner) and (e) the normalized decumulative histogram of error.

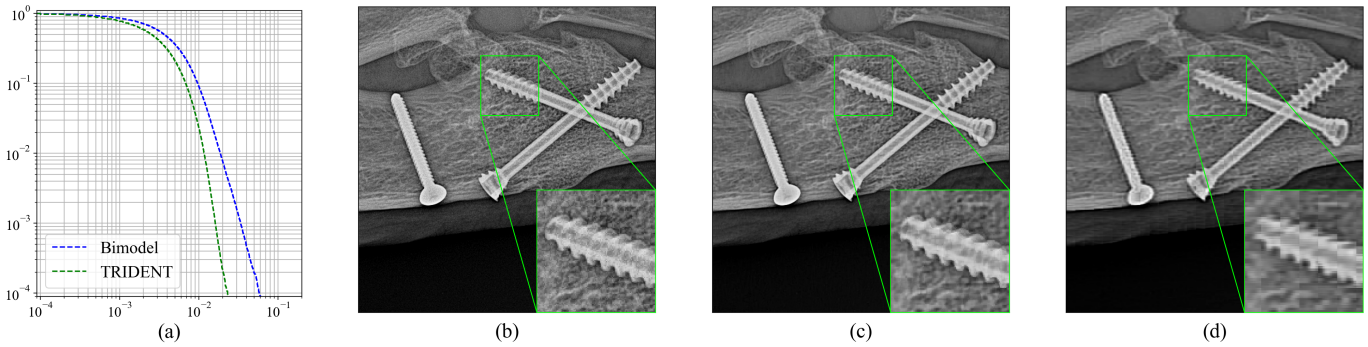


Fig. 4. Clinical representation estimation with dictionaries. (a): Decumulative histogram of error of the clinical content between the dictionary representation and the ground truth, (b): Ground Truth $\mathcal{C}$, (c): With ECC removal ($\hat{\mathcal{C}}_r + \mathcal{C}_e$). (d): Without ECC removal ($\hat{\mathcal{C}}$).

has higher frequency features, ambiguous with the artifact. All of the unambiguous features have been removed and are represented by $\mathcal{C}_e$.

We also learned an artifact dictionary D_A of $k_A = 256$ components over ringing stripes with a maximum of nonzero components to use fixed at $L_{0_a} = 3$ and presented in Fig.2.c. The training samples are retrieved at different locations since we aim to represent each translation possible of the artifact. Orthogonal Matching Pursuit [8] is used to estimate the sparse code x_i during the inference phase with $L_0 = L_{0_c} + L_{0_a} = 3 + 3$.

To improve the results, we used an ad hoc artifact detection function based on the fact that this kind of artifact appears over the width of the image. We measured the normalized profile sum of the artifact representation $\tilde{\mathcal{A}}$ for each row and made a binary decision for the detection of the artifact for a row with respect to a threshold. This method has been applied to each method (unimodel, bimodel and TRIDENT) with the same parameters.

B. Qualitative Evaluation

We study qualitatively an artifact correction example in Fig.3. Our method allows a better extraction of the artifact (Fig.3.b & Fig.3.f) with minimal underestimation. The bimodel approach (Fig.3.c & Fig.3.g) underestimate the artifact since some of it is captured by the clinical model. However, the unimodel is unable to capture the complexity of the artifact because of the simplicity of its model. Most of the varying features of the artifact have been missed (Fig.3.d & Fig.3.h) creating dashed pieces of the residual artifact.

Furthermore, we compare the final artifact correction for the bimodel method ($\hat{\mathcal{C}}$) and TRIDENT ($\hat{\mathcal{C}}_r + \mathcal{C}_e$) in Fig.4. Thanks

to the ECC separation, TRIDENT outperforms the bimodel reconstruction, thus reducing the risk of the overestimation of the artifact over the clinical content. Nevertheless, this estimation is less accurate than the correction $\tilde{\mathcal{C}}$ (Algorithm 1) as pointed out by the error histogram of the clinical estimation (Fig.4.a) vs. (Fig.3.e).

C. Quantitive Evaluation

The normalized decumulative histograms of error comparison between the different methods Fig.3.e shows that TRIDENT produces fewer errors. The other methods small errors are comparable. However, the unimodel generates bigger errors since it lacks a proper model for the artifact and clinical content.

We also applied TRIDENT over 952 clinical images (same dataset) of feet and hands with ringing artifacts at random levels and, positions over the images. The PSNR and a SafeGuarding Measure (SFM), that measures the percentage of erroneous pixels, are evaluated:

$$\mathrm{PSNR} = -10\log_{10}\left(\frac{\|\mathcal{C}-\tilde{\mathcal{C}}\|_2^2}{n}\right) \quad ,\mathrm{SFM} = \frac{\|\tilde{\mathcal{C}}-\mathcal{C}\|_0}{n} \quad (9)$$

with $\|\bullet\|_2$ denoting the ℓ_2 norm and $\|\bullet\|_0$ the ℓ_0 one.

Computing the mean metrics over the whole dataset (Table I) demonstrates that our method outperforms all the other methods. TRIDENT offers higher PSNR and lower SFM (safer). Even if the bimodel method can estimate the artifact in general, it fails in complex scenarios (screws in Fig.3.c). It also has the highest SFM results because the clinical dictionary is less efficient, the artifact dictionary has taken a major part of the representation resulting in an overestimation in some cases (Fig.3.g).

To improve the performances of this method, a large number of atoms must be added in the clinical dictionary leading to much higher computation time diminishing its appeal. Moreover, adding more atoms does not guarantee a better correction because more atoms increase the risk of learning shared features with the artifact. With TRIDENT, this is avoided by removing the exclusively clinical component.

	TRIDENT	Bimodel	Unimodel
PSNR (dB) ↗	**70.3**	59.8	52.1
SFM (%) ↘	**0.75**	1.7	1.2

TABLE I
PSNR AND SFM MEASUREMENTS

V. CONCLUSIONS

We presented a trimodel approach for artifact extraction in X-ray images, where the observed image is divided into exclusively clinical component and a joint one. Its application to ringing stripes demonstrates its superiority to unimodel and bimodel approaches.

Future work will focus on accelerating the resolution of the joint model. Furthermore, we will study a better representation for the ECC which depends on the artifact features.

ACKNOWLEDGMENT

The patient cohort was created during a clinical trial in the frame of the "Programme Hospitalier de Recherches Cliniques" (PHRC Research Program IDRCB 2008- A01422-53) funded by the French Ministry of Health. The original goal of this research was to compare the diagnostic and metrological performance of conventional radiography versus high-resolution digital radiographs for detecting joint space narrowing and erosions in RA. All patients fulfilled the 1987 American College of Rheumatology revised criteria for RA. Authorization to conduct the study was obtained from the ethics committee (CPP Tours-2008-A01422-53). All patients gave their written consent.

REFERENCES

[1] A. I. Walz-Flannigan, K. J. Brossoit, D. J. Magnuson, and B. A. Schueler, "Pictorial Review of Digital Radiography Artifacts," *RadioGraphics*, vol. 38, no. 3, pp. 833–846, May 2018.

[2] L. L. Barski and X. Wang, "Characterization, detection, and supression of stationary grids in digital projection radiography imagery," S. K. Mun and Y. Kim, Eds., San Diego, CA, May 1999, p. 502.

[3] T. Maruyama and H. Yamamoto, "Elimination of gridlines by using nonlinear filter in mammographic image," *IET Image Processing*, vol. 5, no. 5, p. 457, 2011.

[4] R. Sasada, M. Yamada, S. Hara, H. Takeo, and K. Shimura, "Stationary grid pattern removal using 2D technique for moire-free radiographic image display," R. L. Galloway, Jr., Ed., San Diego, CA, May 2003, p. 688.

[5] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, and W. Chen, "Artifact Suppressed Dictionary Learning for Low-Dose CT Image Processing," *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2271–2292, Dec. 2014.

[6] F. Cotte, M. Desvignes, H. Ayasso, and J.-M. Vignolle, "A sparse dictionary representation approach for anti-scattering grid artifact removal in X-ray images," *Biomedical Signal Processing and Control*, vol. 86, p. 105247, Sep. 2023.

[7] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[8] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov. 1993, pp. 40–44 vol.1.

[9] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, Apr. 2004.

[11] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for Sparse Representation Modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[12] M. Elad and M. Aharon, "Image Denoising Via Learned Dictionaries and Sparse representation," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)*, vol. 1. New York, NY, USA: IEEE, 2006, pp. 895–900.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal Quebec Canada: ACM, Jun. 2009, pp. 689–696.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[15] S. K. Sahoo and A. Makur, "Dictionary Training for Sparse Representation as Generalization of K-Means Clustering," *IEEE Signal Processing Letters*, vol. 20, no. 6, pp. 587–590, Jun. 2013.