

Manuscrit de thèse en vue de l'obtention du diplôme de doctorat

Thèse CIFRE issue d'un partenariat entre :
l'école CentraleSupélec
l'université Paris Diderot
SAFRAN

Représentations pour la détection d'anomalies
Application aux données vibratoires des moteurs d'avions

Mina ABDEL-SAYED

11 mars 2018

Table des matières

Table des matières	VI
Résumé	1
Abstract	1
Introduction	3
Le contexte industriel	3
Problématique	4
Les approches préconisées	6
Contributions	8
Structure du mémoire	8
 I L'analyse vibratoire et détection de nouveautés	 9
1 L'analyse et les données vibratoires	13
1.1 Introduction	13
1.2 Les moteurs d'avions	13
1.2.1 Les caractéristiques des moteurs	13
1.2.2 Les vibrations du moteur	15
1.2.3 L'acquisition des mesures vibratoires et des vitesses de rotation	16
1.3 Conversion des signaux temporels en spectrogrammes	16
1.3.1 Intérêt de cette conversion	16
1.3.2 La transformation du signal temporel en spectrogramme	18
1.3.3 Gains et limites de cette représentation	19
1.4 Construction de la base de données	21
1.4.1 L'annotation manuelle des experts	21
1.4.2 Extraction automatique des zones anormales sur les données textuelles . .	23
1.4.3 La base de données enrichie	24
1.5 Étude des spectrogrammes par patch	25
1.5.1 Localisation des signatures inusuelles sur le spectrogramme	25
1.5.2 Subdivision du spectrogramme en patches	26
1.5.3 Labélisation ponctuelle du patch - enrichissement de la base de données .	29
1.5.4 La grande variabilité des signatures inusuelles	30
1.6 L'état de l'art de l'analyse vibratoire	31
1.6.1 L'état de l'art provenant de la littérature	31
1.6.2 Les algorithmes d'analyse vibratoire de Safran Aircraft Engines	34
1.7 Une première approche de détection	36
1.7.1 Représentation des patches par leurs histogrammes d'intensités vibratoires	37
1.7.2 Représentation des histogrammes dans un espace réduit	39
1.8 Conclusions	39

2	La détection de nouveautés	41
2.1	Définition	41
2.2	État de l'art de la détection de nouveautés	42
2.2.1	Les approches probabilistes	43
2.2.2	Les approches basées sur les distances	44
2.2.3	Les approches basées sur la reconstruction des données	44
2.2.4	Les approches basées sur la caractérisation des limites des données normales	45
2.2.5	Les approches basées sur la théorie de l'information	46
2.3	La détection de nouveautés appliquée aux données vibratoires	46
2.3.1	Application aux données temporelles et fréquentielles	46
2.3.2	Application sur les harmoniques du signal	47
2.3.3	Application aux spectrogrammes	48
2.4	Caractérisation de la base de données construite	49
2.4.1	Répartition des données en sous-ensembles	49
2.4.2	Visualisation des résultats	51
II	Les approches de représentation globale par dictionnaire	53
3	Représentation par dictionnaire fixe - les curvelets	57
3.1	Introduction	57
3.1.1	La représentation par dictionnaire	57
3.1.2	Les dictionnaires non-adaptatifs	59
3.2	La transformée en curvelet	60
3.2.1	La transformée en ridgelet	60
3.2.2	Les ridgelets multi-échelles	61
3.2.3	La construction de la transformée en curvelet	62
3.2.4	Le dictionnaire des curvelets	64
3.3	Application des curvelets aux spectrogrammes vibratoires	64
3.3.1	Caractérisation des raies vibratoires à partir des curvelets	64
3.3.2	Comparaison des représentations en curvelet	66
3.4	Normalité définie dans le dictionnaire des curvelets	68
3.4.1	Le modèle de normalité	68
3.4.2	Normalité définie par optimisation avec contraintes de parcimonie	72
3.4.3	Comparaison des supports normaux	73
3.5	Le score de détection d'anomalies sur le patch	75
3.5.1	Les scores de détection	75
3.5.2	Le score de normalité	76
3.5.3	Résultats sur la base de test	78
3.6	L'étude des résidus ponctuels	80
3.6.1	Les résidus du modèle de normalité \mathcal{D}_{Supp}^c	80
3.6.2	La détection d'anomalies	81
3.6.3	Calibration des paramètres des modèles sur Ω_{Val}	83
3.6.4	Résultats sur la base de test Ω_{Test}	89
3.7	Conclusions	93
4	Représentation par dictionnaire data-driven-NMF	95
4.1	Les dictionnaires adaptatifs/data-driven	95
4.1.1	Définition mathématique du problème	95
4.1.2	Les méthodes de résolution	96
4.1.3	Quelques exemples de dictionnaires data-driven (adaptatifs)	97
4.1.4	Les dictionnaires appris pour caractériser les spectrogrammes	98

4.2	Non-Negative Matrix Factorization (NMF)	99
4.2.1	Formulation mathématique	99
4.2.2	Résolution de la problématique	99
4.3	Le modèle de normalité défini à partir de la NMF	100
4.3.1	L'apprentissage du dictionnaire	100
4.3.2	Définition du rang du dictionnaire	101
4.3.3	Représentation de la normalité à partir du dictionnaire de la NMF	102
4.4	Détection d'anomalies sur les patches	105
4.4.1	Les scores de détection	105
4.4.2	Résultats sur la base de test	108
4.5	Les erreurs ponctuelles issues du dictionnaire de la NMF	109
4.5.1	Les résidus de la NMF	109
4.5.2	La détection des points inusuels	110
4.5.3	Calibration des paramètres	112
4.5.4	Résultats sur la base de test Ω_{Test}	115
4.5.5	Complémentarité des approches adaptatives et non-adaptatives	116
4.6	Conclusions	119

III Analyse ponctuelle des spectrogrammes

121

5 Analyse ponctuelle indépendante

125

5.1	Introduction	125
5.1.1	Considération ponctuelle des points des spectrogrammes	125
5.1.2	La base de données	126
5.1.3	Les modèles de normalité	127
5.2	Modélisation paramétrique de la distribution de normalité	127
5.2.1	Le modèle de normalité	127
5.2.2	Le score de détection	128
5.2.3	Calibration des seuils de détection sur la base de validation Ω_{Val}	129
5.2.4	Résultats sur la base de test Ω_{Test}	130
5.3	Estimation non paramétrique de la densité par noyau	135
5.3.1	Formulation	135
5.3.2	L'estimation de la densité par noyau gamma	136
5.3.3	Estimation de l'échelle du noyau	137
5.4	Distribution de normalité estimée par les noyaux gaussiens	139
5.4.1	Le modèle de normalité	139
5.4.2	Le score de détection	140
5.4.3	Calibration des seuils	142
5.4.4	Résultats sur la base de test Ω_{Test}	143
5.5	Distribution de normalité estimée par les noyaux gamma	147
5.5.1	Le modèle de normalité	147
5.5.2	Le score de détection	148
5.5.3	Calibration des seuils de détection	150
5.5.4	Résultats sur la base de test Ω_{Test}	151
5.6	Conclusions	153

6 Analyse ponctuelle conditionnelle au voisinage

155

6.1	Considération des points voisins	155
6.1.1	Dépendance par rapport au voisinage des points	155
6.1.2	Modèle de normalité par rapport au voisinage	156
6.2	Estimation de densité conditionnelle par noyau	157

6.2.1	L'estimation de densité par noyau dans un cadre multidimensionnel	157
6.2.2	Calibration de la matrice d'échelle	158
6.2.3	L'estimation de la densité conditionnelle par noyau	159
6.3	Le modèle de normalité défini à partir du voisinage d'ordre 1	160
6.3.1	La structure du voisinage	160
6.3.2	Le modèle de normalité	161
6.3.3	Le score de détection	162
6.3.4	Calibration du modèle	165
6.3.5	Résultats sur la base de test Ω_{Test}	167
6.4	Normalité défini en fonction de la direction du voisinage	169
6.4.1	Le voisinage directionnel	169
6.4.2	Le modèle de normalité	170
6.4.3	Le score de détection	171
6.4.4	La caractérisation des signatures inusuelles	172
6.4.5	Calibration du seuil de détection	173
6.4.6	Résultats sur la base de test Ω_{Test}	173
6.5	Fusion des différentes approches	180
6.5.1	Comparaison des approches par dictionnaire et ponctuelles	180
6.5.2	Fusion des approches par dictionnaire et ponctuelles	180
6.6	Conclusions	184
Conclusions et perspectives		187
	Conclusions	187
	Perspectives	191
Bibliographie		193
Table des figures		200
Liste des tableaux		207
Annexes		209
A Les tests multiples		209

Liste des symboles

f	Fréquence
N_2	Régime de l'arbre haute pression du moteur
N_1	Régime de l'arbre basse pression du moteur
S^i	Spectrogramme vibratoire du moteur i
S_{f,N_2}^i	Intensité vibratoire à la fréquence f et au régime N_2 du moteur i
\mathcal{K}	Subdivision du spectrogramme en patches
\mathcal{K}_j	Élément j de la subdivision \mathcal{K} du spectrogramme en patches
$Z_{\mathcal{K}_j}^i$	Patch correspondant à la l'élément j de la subdivision \mathcal{K} extrait du spectrogramme i
$\hat{Z}_{\mathcal{K}_j}^i$	Estimation normale du patch correspondant à la l'élément j de la subdivision \mathcal{K} extrait du spectrogramme i
Ω_{App}^j	Base d'apprentissage du patch j de la subdivision \mathcal{K}
Ω_{Val}^j	Base de validation du patch j de la subdivision \mathcal{K}
Ω_{Test}^j	Base de test du patch j de la subdivision \mathcal{K}
Ω_{App}^{f,N_2}	Base d'apprentissage des points aux coordonnées (f, N_2)
Ω_{Val}^{f,N_2}	Base de validation des points aux coordonnées (f, N_2)
Ω_{Test}^{f,N_2}	Base de test des points aux coordonnées (f, N_2)
$Y_{Z_{\mathcal{K}_j}}^i$	Vérité terrain de la classe normale ou atypique pour le spectrogramme i du patch Z_j provenant de l'élément j de la subdivision \mathcal{K}
$\hat{Y}_{Z_{\mathcal{K}_j}}^i$	La classe normale ou atypique estimée pour le spectrogramme i du patch Z_j provenant de l'élément j de la subdivision \mathcal{K}
Y_{f,N_2}^i	Vérité terrain sur la classe normale ou atypique du point de coordonnées (f, N_2) pour le spectrogramme i
\tilde{Y}_{f,N_2}^i	La classe, normale ou atypique, estimée sans filtrage par voisinage du point de coordonnées (f, N_2) pour le spectrogramme i
\hat{Y}_{f,N_2}^i	La classe, normale ou atypique, estimée avec filtrage par voisinage du point de coordonnées (f, N_2) pour le spectrogramme i
\mathcal{V}_{f,N_2}	Voisinage du point de coordonnées (f, N_2)
\mathcal{V}_{f,N_2}^k	Voisinage du point de coordonnées (f, N_2) dans la direction k
\mathcal{D}^C	Dictionnaire des curvelets
\mathcal{D}^{NMF}	Dictionnaire de la NMF

$SuppZ_{\mathcal{K}_j}^i$	Atomes des curvelets activés pour le patch $Z_{\mathcal{K}_j}^i$
$Supp^*$	Atomes des curvelets activés pour un pourcentage des données d'un patch de la base d'apprentissage
$R(Z_{\mathcal{K}_j}^i)$	Résidus de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$R^+(Z_{\mathcal{K}_j}^i)$	Résidus positifs de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$R^-(Z_{\mathcal{K}_j}^i)$	Résidus négatifs de reconstruction du patch $Z_{\mathcal{K}_j}^i$
$\mathcal{N}(\mu, \sigma)$	Loi et densité gaussienne de paramètre (μ, σ)
f_X	densité de la variable aléatoire X
$f_{\mathcal{N}}$	Densité de la loi gaussienne
$F_{\mathcal{N}}$	Fonction de répartition de la loi gaussienne
$K^{\mathcal{N}}$	Noyau gaussien
K^{Γ}	Noyau gamma
h	Echelle du noyau en une dimension
H	Matrice d'échelle du noyau multidimensionnel
\mathcal{F}	Transformée de Fourier
$\mathbf{1}$	Fonction indicatrice
$\ x\ _0$	norme 0, $\ x\ _0 = \text{card}(i : x_i \neq 0)$ avec $x = (x_1, \dots, x_n)$
$\ x\ _1$	norme 1, $\ x\ _1 = \sum x_i $ avec $x = (x_1, \dots, x_n)$
$\ x\ _2$	norme 2, $\ x\ _2 = \sqrt{\sum x_i ^2}$ avec $x = (x_1, \dots, x_n)$

Résumé

Les mesures de vibrations sont l'une des données les plus pertinentes pour détecter des anomalies sur les moteurs. Les vibrations sont acquises sur banc d'essai en phase d'accélération et de décélération pour assurer la fiabilité du moteur à la sortie de la chaîne de production. Ces données temporelles sont converties en spectrogrammes pour permettre aux experts d'effectuer une analyse visuelle de ces données et de détecter les différentes signatures atypiques. Les sources vibratoires correspondent à des raies sur les spectrogrammes. Dans cette thèse, nous avons mis en place un outil d'aide à la décision automatique pour analyser les spectrogrammes et détecter tout type de signatures atypiques, ces signatures ne proviennent pas nécessairement d'un endommagement du moteur. En premier lieu, nous avons construit une base de données numérique de spectrogrammes annotés. Il est important de noter que les signatures inusuelles sont variables en forme, intensité et position et se trouvent dans un faible nombre de données. Par conséquent, pour détecter ces signatures, nous caractérisons les comportements normaux des spectrogrammes, de manière analogue aux méthodes de détection de nouveautés, en représentant les patches des spectrogrammes sur des dictionnaires comme les curvelets et la Non-negative matrix factorization (NMF), ainsi qu'en estimant la distribution de chaque point du spectrogramme à partir de données normales dépendamment ou non de leur voisinage. La détection des points atypiques est réalisée par comparaison des données tests au modèle de normalité estimé sur des données d'apprentissage normales. La détection des points atypiques permet la détection des signatures inusuelles composées par ces points.

mots clés : apprentissage de dictionnaire, curvelets, estimation de densité par noyau, détection de nouveautés, détection d'anomalies, vibrations

Abstract

Vibration measurements are one of the most relevant data for detecting anomalies in engines. Vibrations are recorded on a test bench during acceleration and deceleration phases to ensure the reliability of every flight engine at the end of the production line. These temporal signals are converted into spectrograms for experts to perform visual analysis of these data and detect any unusual signature. Vibratory signatures correspond to lines on the spectrograms. In this thesis, we have developed a decision support system to automatically analyze these spectrograms and detect any type of unusual signatures, these signatures are not necessarily originated from a damage in the engine. Firstly, we have built a numerical spectrograms database with annotated zones, it is important to note that data containing these unusual signatures are sparse and that these signatures are quite variable in shape, intensity and position. Consequently, to detect them, like in the novelty detection process, we characterize the normal behavior of the spectrograms by representing patches of the spectrograms in dictionaries such as the curvelets and the Non-negative matrix factorization (NMF) and by estimating the distribution of every points of the spectrograms with normal data depending or not of the neighborhood. The detection of the

unusual points is performed by comparing test data to the model of normality estimated on learning normal data. The detection of the unusual points allows the detection of the unusual signatures composed by these points.

keywords : dictionary learning, curvelets, kernel density estimation, novelty detection, anomaly detection, vibrations