Classification d'images à l'aide d'un codage par motifs fréquents

Image classification using a frequent item sets coding

A. Cornuéjols[1] J. Mary[2] M. Sebag[2]

[1] Institut d'Informatique d'Entreprise et Laboratoire de Recherche en Informatique (L.R.I.)
[2] Laboratoire de Recherche en Informatique (L.R.I.)

L.R.I., bât. 490, Université de Paris-Sud, Orsay, F-91405 Orsay Cedex
(antoine,mary,sebag)@lri.fr

Résumé

Les applications de reconnaissance d'images se caractérisent par des données décrites à l'aide d'un très grand nombre d'attributs alors que les données disponibles sont souvent en nombre limité en comparaison. C'est pourquoi on a recours à des méthodes de prétraitement des données permettant d'identifier un recodage des entrées mieux adapté à la tâche. Ce papier passe en revue les grandes techniques de représentation des données en grande dimension et présente une nouvelle méthode consistant à chercher des motifs fréquents dans les données et à les utiliser pour coder les entrées. Nous illustrons son application à la classification de scènes naturelles et à la reconnaissance de caractères manuscrits.

Mots Clef

Classification, codage, images, motifs fréquents, .

Abstract

Images recognition implies a high dimensional input space and, often, a limited amount of learning data. This is why preprocessing techniques are generally used in order to find a better suited representation. In this paper, after reviewing several data representation techniques, we present a new coding method where low coverage frequent item sets discovered in the training sample are used to encode the data. This method has been used for the classification of natural scenes and of hand-written digits.

Keywords

Classification, coding, Frequent Item Sets, images.

1 Introduction

Les applications telles que la reconnaissance d'images ou la classification de puces à ADN se caractérisent par des espaces d'entrée de grandes dimensions et, souvent, par des échantillons d'apprentissage de taille très limitée en comparaison. Dans ces conditions, les données sont difficiles à analyser directement car elles ne sont pas décrites de manière à faire apparaître les régularités sous-jacentes pertinentes qui existent si le phénomène mesuré correspond à un sous-espace très limité de l'espace de description initial. On a alors généralement recours à des techniques de pré-traitement des données afin de chercher à identifier une redescription des données mieux adaptée à la tâche.

Parmi ces méthodes, les plus employées concernent des redescriptions utilisant des combinaisons linéaires des descripteurs initiaux. Deux raisons en particulier motivent ce choix. D'une part, le problème d'optimisation correspondant à ces méthodes est plus simple à poser et corrélativement ces méthodes sont relativement aisées à mettre en œuvre (e.g. inversion de matrices). D'autre part, dans le cas de la reconnaissance d'images, il semble que le système visuel primaire fonctionne en codant les entrées comme des superpositions (linéaires) de structures simples (e.g. arêtes, lignes, textures) [4].

Une méthode classique de représentation de signal est naturellement l'analyse de Fourier. L'analyse en ondelettes qui la généralise est de plus en plus employée dans une grande variété d'applications. Ces méthodes ont l'avantage de possèder de solides fondations mathématiques et sont l'objet d'implémentations informatiques performantes. En revanche, elles ne prennent pas en compte les caractéristiques particulières des données analysées.

Par contraste, les méthodes dépendantes des données cherchent à extraire les propriétés statistiques du domaine d'application. Ces représentations sont construites à partir d'un échantillon d'apprentissage en optimisant des mesures qui quantifient les propriétés jugées pertinentes pour l'application. Cette classe de méthodes inclue l'analyse en composantes princi-

pales (ACP), l'analyse en composantes indépendantes (ACI) et la factorisation non négative de matrices (FNM).

Brièvement, l'*analyse en composantes principales* identifie les axes principaux d'inertie des données et permet une redescription par projection sur ces axes. Cette méthode conduit à une compression de l'information, mais ne permet pas de trouver les sous-structures présentes dans les données. L'*analyse en composantes indépendantes* [2, 5] suppose que les données mesurées résultent de la superposition linéaire d'un nombre donné de sources statistiquement indépendantes. Elle cherche à reconstruire ces signaux de base en les supposant non gaussiens. Lorsque les « sources » sont nombreuses et assimilées à un vocabulaire de formes latentes, le codage des données ainsi obtenu devient *clairsemé*, c'est-à-dire que chaque donnée est codée grâce à un nombre réduit de formes latentes. Les études récentes tendent à montrer que c'est le type de codage réalisé dans l'aire primaire visuelle (aire V1) [7]. Il est donc sans doute bien adapté au traitement d'images. La *factorisation non négative de matrice* cherche également un dictionnaire de formes latentes, mais impose de plus que les données soient décrites par des superposition linéaires de ces formes ne faisant intervenir que des coefficients positifs. Il en résulte que les formes latentes trouvées correspondent en général à des composantes identifiables (voir par exemple l'analyse d'images de visages [6]).

Si ces techniques sont séduisantes pour le traitement d'images dans lequel on peut chercher à caractériser celles-ci par les formes latentes présentes, elles ne sont malheureusement pas directement applicables aux données décrites en grande dimension et sont en fait utilisées en découpant les images en imagettes de dimensions réduites (8×8 ou 12×12). Dans ce papier, nous présentons une nouvelle méthode de codage de données, le codage clairsemé par motifs fréquents (*Frequent Item Sets for Independant Component Analysis : FISICA*), qui est particulièrement adaptée à l'analyse de données en grandes dimensions. Nous illustrons son fonctionnement sur deux applications : la reconnaissance de scènes naturelles et la reconnaissance de chiffres manuscrits.

2 La méthode FISICA

Si l'approche par l'analyse en composantes indépendantes d'un codage clairsemé est impossible, une approche directe est-elle envisageable ?

En faisant l'hypothèse que les données résultent d'une somme de formes latentes, et que ces formes pour être intéressantes doivent figurer suffisamment souvent dans les données, le rapprochement avec la technique de recherche de motifs fréquents s'impose. Les données étant décrites par un ensemble de descripteurs (attributs-valeur), on cherche les conjonctions d'attributs-valeur présentes dans un certain pourcentage des données. Grâce à certaines contraintes, on peut guider la recherche de tels motifs de manière à favoriser la découverte d'un ensemble de primitives (les motifs) permettant un codage clairsemé des données. (Voir la section 2.1 ci-dessous).

À ces contraintes peuvent s'ajouter des critères supplémentaires liées au domaine d'application permettant de sélectionner les motifs fréquents les plus intéressants. Par exemple, dans le domaine de l'analyse d'images, on pourra favoriser la recherche de fonctions de base ou motifs correspondant à des régions connexes, ou à des lignes (voir plus bas, les résultats expérimentaux).

Les *fonctions de base* cherchées sont des conjonctions d'attributs-valeur, on parlera aussi d'atomes. Dans le cas de l'analyse d'images, il s'agira de collections de pixels, chacun de ceux-ci étant associé à un niveau de gris donné. Une fonction de base est ainsi une fonction boolénne prenant la valeur `vrai` si la collection d'atomes (pixel = valeur) correspondante est satisfaite dans l'image étudiée. On dira que le *support* d'une fonction de base est de ε% si cette fonction prend la valeur `vrai` dans ε% des images testées. Réciproquement, on appellera *code* d'une forme d'entrée (ici une image), l'ensemble des fonctions de base vérifiant cette forme.

2.1 Propriétés des fonctions de base recherchées

Dans ce nouveau cadre, les propriétés désirées du système de codage se traduisent comme suit :

1. **Représentativité**. Chaque fonction de base a un support supérieur à ε%. Elle est donc suffisamment représentée dans la base d'exemples pour être utile.
2. **Parcimonie**. Peu de fonctions de base sont vérifiées par un exemple (e.g. image). On vérifie ainsi l'une des propriétés du codage clairsemé.
3. **Suffisance**. Tout exemple rend vrai un nombre minimal de fonctions de base.
4. **Orthogonalité**. Pour chaque paire de fonctions de base, l'intersection des exemples qui rendent `vrai` l'une et l'autre est réduite. Les exemples sont donc décrits par des codes différents.

Nous avons adapté la méthode de recherche de motifs fréquents dans une base de données pour chercher un codage tendant à vérifier les propriétés ci-dessus.

La recherche de motifs fréquents de taux de couverture faible dans des données décrites par de nombreux attributs ne peut s'effectuer sans précautions. C'est pourquoi nous présentons rapidement l'algorithme développé à cet effet.

2.2 Échec d'une approche naïve

Une approche consiste à utiliser un algorithme existant de recherche de motifs fréquents, tel qu'Apriori [1], pour chercher tous les motifs ayant un certain taux de couverture, puis ensuite à sélectionner parmi eux ceux qui satisfont les contraintes soulignées dans la section 2.1. Mais cette approche se révèle impraticable pour la reconnaissance d'images car le nombre de motifs fréquents croît exponentiellement en fonction de la taille des motifs comme le montre la table ci-dessous pour des images de taille 32x32 en 64 niveaux de gris (où m signifie mille, M million et MM milliard) :

1	2	3	4	5	6
2m	110m	3,8M	80M	1,15MM	12,5MM

Il faut donc renoncer à une méthode de recherche exhaustive des motifs.

2.3 Approche randomisée

Nous avons développé une méthode de construction incrémentale de motifs fréquents par ajouts successifs d'atomes (ici, de pixels d'un certain niveau de gris) en les sélectionnant à chaque pas afin que le motif en construction satisfasse aux critères désirés. L'exploration des motifs fréquents est donc maintenant stochastique, guidée mais non exhaustive. Des essais successifs peuvent ainso produire des bases de motifs différentes.

Algorithme 1 Recherche itérative et stochastique de motifs fréquents.

Paramètres : taux de couverture ε%. Nombre de motifs cherchés = N.
Nombre de motifs trouvés = $n \leftarrow 0$.
while $n \leq N$ **do**
  Choix dans un exemple x_i encore peu couvert, d'un premier atome a_0 présent dans au moins ε% des exemples.
  $motif \leftarrow a_0$
  while Taux de couverture de $motif > \varepsilon$% **do**
    Tirer au hasard un atome a de x_i couvrant au moins ε% des exemples et peu utilisé dans les motifs existants et satisfaisant des contraintes additionnelles sémantiques (voir section 3).
    if $motif + a$ couvre au moins ε% des exemples **then**
      $motif \leftarrow motif + a$
    end if
  end while
end while

2.4 Exploitation pour l'apprentissage supervisé

Une fois N fonctions de base trouvées sur un ensemble d'apprentissage, chaque exemple est recodé, devenant un vecteur de N booléens prenant la valeur `vrai` ou `faux` selon que la fonction de base correspondante couvre l'exemple ou non. [1].
Dans le nouvel espace d'exemples ainsi construit, il est possible d'utiliser n'importe quelle méthode d'apprentissage supervisé. Dans les expériences rapportées ici, nous avons utilisé une méthode de classification par plus proche voisin. Les exemples d'apprentissage utilisés pour la recherche de fonctions de base sont également employés comme exemples étiquetés servant à la classification des exemples testés.

3 Application à la reconnaissance d'images

La méthode développée a été testée sur des tâches de classification d'images de scènes naturelles et de chiffres manuscrits. Elle implique deux phases : d'abord une étape de détermination d'une base de fonctions de base permettant de redécrire les données, ensuite l'emploi du système de codage ainsi obtenu pour classer de nouvelles formes.

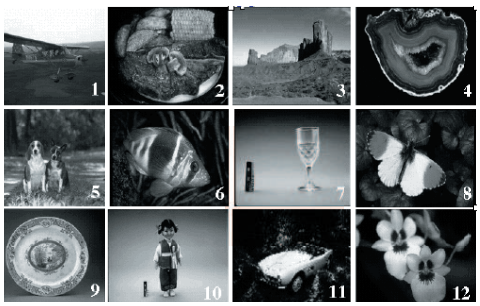

Fig. 1 – Échantillon d'images utilisées dans cette étude. Noms des classes : avions (1), plats (2), Utah (3), minéraux (4), chiens (5), poissons (6), verres (7), papillons (8), porcelaines (9), figurines (10), voitures (11), fleurs (12). (Cette figure est reprise de [3]).

Dans le cas de la reconnaissance de scènes naturelles, le problème consiste à apprendre à reconnaître des images de scènes naturelles classées en 12 catégories (voir la figure 1). Ces images proviennent de la base COREL (`http://www.corel.com/gallery_line/`). Les images sont redécrites par 128×128 pixels en 128 niveaux de gris. Pour ces expériences, la base utilisée comportait 1082 images réparties également entre les 12 classes. Nous considérerons dans la suite que chaque pixel est un attribut pouvant prendre une valeur parmi 128. La dimension de l'espace d'entrée est donc dans ce cas de $32768 = 128 \times 128$.
Pour l'application étudiée, nous avons fixé à 1000 le nombre de fonctions de base recherchées. Plusieurs

[1] Nous avons aussi utilisé une formule d'appariement plus souple utilisant une fonction sigmoïde à valeur dans $[0, 1]$ qui tient compte du nombre d'atomes de la fonction de base qui couvrent l'exemple, et donnant un recodage dans $[0, 1]^N$ au lieu de $\{0, 1\}^N$

bases ont été obtenues en faisant varier les paramètres suivants :

- *Taux de couverture* : 1%, 2%, 5% et 10%
- *Critère sémantique additionnel*. Nous avons introduit des contraintes supplémentaires sur la construction des fonctions de base afin de tester des équivalences possibles avec d'autres types de codages classiques en traitement d'images. Quatre conditions ont été testées :
  1. Aucune contrainte.
  2. Les fonctions de base doivent correspondre à des *régions connexes* sur l'image : un nouveau pixel n'est ajouté à la fonction de base courante que si il est contigu à un pixel déjà sélectionné.
  3. Les fonctions de base doivent correspondre à des *lignes* de l'image (régions de dimension 1). L'idée ici est de voir si l'on peut forcer le système de codage à retenir des contours dans l'image.
  4. Les fonctions de base doivent correspondre à des *lignes raisonnables* de l'image, c'est-à-dire plus contraintes dans les changements de directions possibles. Cette contrainte a été imposée lorsqu'il s'est avéré que la précédente produisait des « vermiceaux » remplissant des régions et non pas des lignes.

Environ la moitié des images de la base initiale de 1082 images, soit 500, ont été utilisées pour le calcul des fonctions de base. (Note : notre algorithme calcule une base de 1000 motifs en quelques minutes sur un PC équipé d'un Pentium II à 266 Mhz et 384 Mo de RAM). Les figures 2 et 3 illustrent le type de fonctions de base obtenues pour certaines conditions expérimentales. Des résultats plus complets sont disponibles sur le site `http://www.lri.fr/~antoine/Research/FISICA/egc-03.html`.
L'histogramme présenté dans la figure 4 (à gauche) permet de contrôler l'orthogonalité des fonctions de base obtenues. Ces fonctions de base sont orthogonales lorsqu'elles sont rarement vérifiées par les mêmes images. La figure montre que les différentes bases obtenues pour des conditions différentes peuvent effectivement être considérées comme orthogonales. Inversement, l'histogramme de la figure 4 (à droite) indique le nombre de fonctions de base qui sont vérifiées par les images. On constate que ce nombre varie autour d'une dizaine, ce qui traduit bien que le codage obtenu est clairsemé.

3.1 Les résultats en classification

Les performances en classification ont été calculées sur les 582 images non utilisées pour déterminer la base des fonctions de base. Toutes les images sont recodées en utilisant les fonctions de base obtenues, et s'expriment donc sous la forme d'un vecteur de 1000

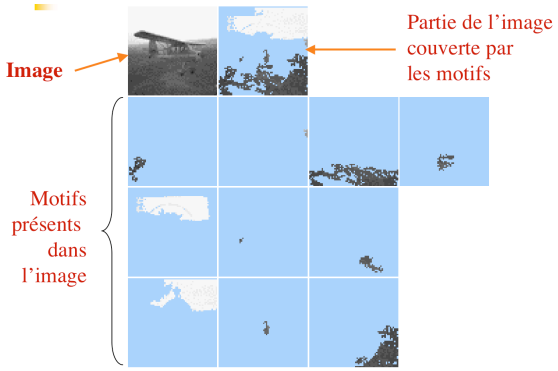

FIG. 5 – Représentation d'une image (ici un avion) à l'aide des motifs fréquents sélectionnés. Il est clair que le codage obtenu ici ne permet pas de reconstruire l'image d'origine et ne vérifie donc pas la propriété d'approximation des méthodes classiques de codage.

valeurs booléennes (en fait, dans certaines expériences, cette valeur booléenne était remplacée par une mesure plus continue d'appariement de l'image avec une fonction de base). Les 500 images employées pour la détermination des fonctions de base sont également utilisées comme base d'exemples étiquetées. Les images à classer sont alors étiquetées en utilisant une méthode de plus proche voisin. Dans les expériences rapportées ici, la distance utilisée est la distance L_1.
La table 3.1 fournit les résultats obtenus avec une base de 1000 fonctions de base de taux de couverture de 5% soumis à la contrainte de connexité. Quoique les différents nombres puissent varier sensiblement, on observe en général que les résultats obtenus dans une grande variété de conditions sont assez similaires en moyenne. Ils sont très sensiblement supérieurs à ceux rapportés dans [3] utilisant un réseau de neurones à bases radiales. Des résultats plus complets sont accessibles à l'url :
`http://www.lri.fr/~antoine/Research/FISICA/corel.html`.

4 Conclusion

Ce papier présente une méthode de précodage des données par l'utilisation de fonctions de base correspondant à des motifs fréquents (de faible taux de couverture) trouvés dans les données d'apprentissage. Ce codage est inhabituel : il n'est pas défini *a priori*, mais au contraire dépend des données ; il ne permet pas de reconstruire les données codées et n'a donc pas de capacité d'approximation ; finalement, l'orthogonalité des fonctions de base est définie par rapport aux données. Ce codage, obtenu de manière non supervisée, s'est révélé très performant, beaucoup plus que des méthodes dédiées, dans une tâche réputée difficile de reconnaissance de scènes naturelles. Il semble en revanche donner des résultats plus médiocres pour la reconnaissance

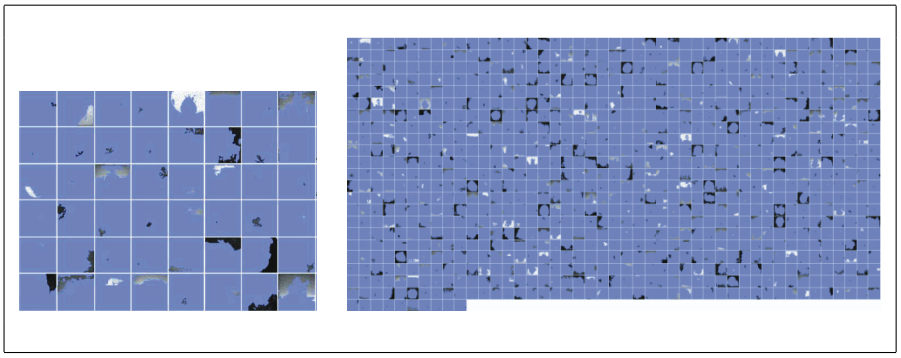

FIG. 2 – À gauche, le détail de quelques unes des fonctions de base obtenues sur des images 64 × 64 en 16 niveaux de gris avec un taux de couverture de 1%, et en cherchant des régions connexes de l'image. Dans les images accessibles sur le site internet, le fond bleu correspond aux zones qui ne font pas partie des fonctions de base, tandis que les fonctions de base sont figurées par des pixels de niveaux de gris variés. Ici, le gris moyen correspond au fond. À droite, figure une base de 1000 fonctions de base.

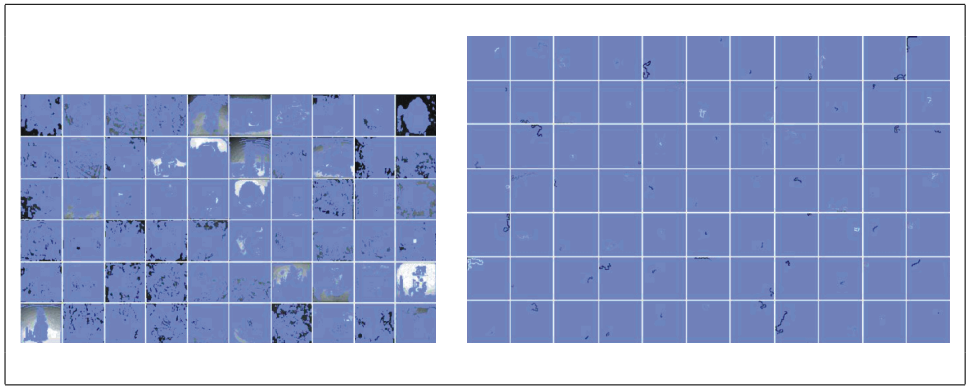

FIG. 3 – À gauche, des exemples de fonctions de base obtenues avec un taux de couverture de 1%, et sans contrainte. À droite, des exemples de fonctions de base obtenues avec un taux de couverture de 1%, et sous contrainte de linéarité raisonnable. L'examen des motifs trouvés montre qu'ils ne correspondent pas à des contours des images de la base d'exemples.

de chiffres manuscrits (de l'ordre de 80% de reconnaissance au lieu d'environ 90% pour l'état de l'art sur ces données). Il reste donc à explorer les possibilités, et les limites, de cette nouvelle méthode de prétraitement et de codage des données.

Remerciements

Ce travail a été en partie réalisé grâce au soutien du contrat BQR FISICA de l'Université de Paris-Sud. Sébastien Jouteau, Jean-Sylvain Liénard et Philippe Tarroux ont largement participé à une phase préalable de cette recherche.

Références

[1] R. Agrawal et R. Srikant, Fast algorithms for mining association rules, *Very Large Data Bases*, Santiago, Chile, September, 1994, pp.487-499.

[2] J-F. Cardoso, Blind signal separation : statistical principles, *Proc. of the IEEE*, vol.9, n° 10, 1998, pp.2009-2025.

[3] N. Denquive et Ph. Tarroux, Catégorisation de scènes visuelles, *Technique et Science Informatique*, vol. X, n° X, 2002, pp.1-18.

[4] D. Field, What is the goal of sensory coding ?, *Neural Computation*, vol. 6, 1994, pp.559-601.

[5] Hyvarinen, Karhunen et Oja, *Independent component analysis*, John Wiley and Sons, 2001.

[6] D. Lee et S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature*, vol. 401, 21 October 1999, pp.788-791.

[7] B. Olshausen et D. Field, Sparse coding with an overcomplete bais set : A strategy employed by V1 ?, *Vision Research*, vol. 37, n° 23, 1996, pp.3311-3325.

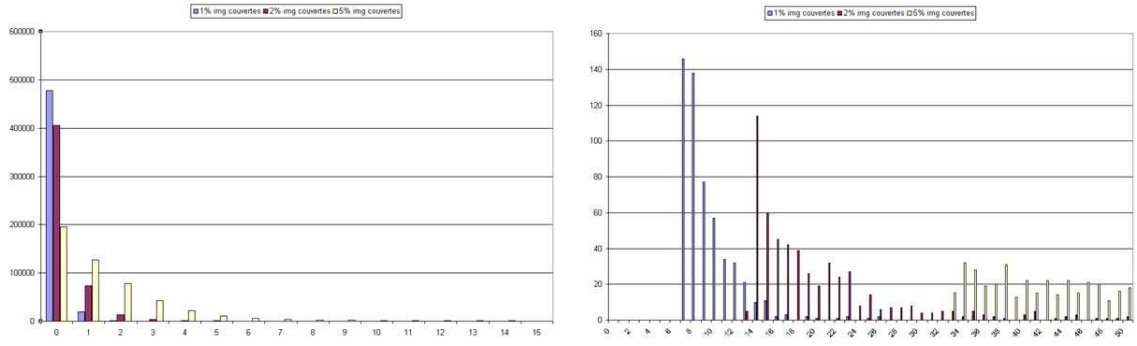

FIG. 4 – À gauche : Histogramme représentant le nombre y de couples de fonctions (en ordonnée) ayant x images en commun (en abscisse). Ces fonctions ont été calculées à partir d'images de taille 64x64 en 16 niveaux de gris et sous la contrainte de connexité. Les résultats sont présentés pour des fonctions de base de taux de couverture 1%, 2% et 5%. Les fonctions de base obtenues pour $\varepsilon = 1\%$ sont les plus orthogonales entre elles. À droite : Histogramme représentant le nombre y d'images (en ordonnée) activant x motifs (fonctions de base) en abscisse. Ces fonctions ont été calculées à partir d'images de taille 64x64 en 16 niveaux de gris et sous la contrainte de connexité. Les résultats sont présentés pour des fonctions de base de taux de couverture 1%, 2% et 5%. Plus le taux de couverture est élevé, plus chaque exemple est couvert en moyenne par un nombre élevé de fonctions. On peut ainsi régler la parcimonie de la représentation et donc son caractère clairsemé.

	Av	Pl	Ut	Mi	Ch	Po	Ve	Pa	Por	Fi	Vo	Fl
Avi	**67%**	2%	-	-	2%	2%	10%	10%	4%	2%	-	-
Pla	-	**21%**	-	2%	7%	19%	10%	12%	5%	-	19%	5%
Uta	17%	-	**33%**	-	7%	-	-	3%	10%	10%	13%	7%
Min	-	-	-	**100%**	-	-	-	-	-	-	-	-
Chi	26%	5%	7%	-	**14%**	9%	12%	9%	5%	2%	12%	-
Poi	5%	13%	3%	8%	-	**13%**	18%	21%	-	3%	10%	8%
Ver	2%	2%	-	-	10%	7%	**43%**	-	21%	5%	7%	2%
Pap	6%	6%	-	-	2%	14%	14%	**35%**	6%	-	12%	4%
Por	2%	2%	-	-	-	2%	-	12%	**70%**	10%	-	2%
Fig	-	-	-	-	-	-	6%	-	24%	**70%**	-	-
Voi	21%	6%	-	-	4%	4%	8%	4%	4%	29%	**19%**	-
Fle	2%	9%	-	-	-	9%	21%	14%	-	-	16%	**28%**

TAB. 1 – Matrice de confusion obtenue avec des fonctions de base de taux de couverture $\epsilon = 5\%$ sous contrainte de connexité, en utilisant une formulle d'appariemment continu entre les images et les fonctions de base.