Analyse factorielle des correspondances hiérarchique pour la fouille d'images

Nguyen-Khang Pham*, Annie Morin**
François Poulet** Patrick Gros***

*Université de Cantho
1 Ly Tu Trong, Cantho, Vietnam
pnkhang@cit.ctu.edu.vn
http://www.cit.ctu.edu.vn
**IRISA, Université de Rennes 1
Campus de Beaulieu, 35042 RENNES Cedex, France
prénom.nom@irisa.fr
http://www.irisa.fr
***INRIA Rennes Bretagne Atlantique
Campus de Beaulieu, 35042 RENNES Cedex, France
prénom.nom@inria.fr

Résumé. Nous proposons un outil graphique interactif qui permet de visualiser et d'extraire des connaissances à partir des résultats de l'Analyse Factorielle des Correspondances (AFC) sur les images. L'AFC est une technique descriptive développée pour analyser des tableaux de contingence. L'AFC est originellement utilisée dans l'Analyse des Données Textuelles (ADT) où le corpus est représenté par un tableau de contingence croisant des documents et des mots. Dans la fouille d'images, nous définissons d'abord les « mots visuels » dans les images (analogues aux mots textuels). Ces mots visuels sont construits à partir des descripteurs locaux SIFT (Scale Invariant Feature Transform) dans l'image. Ensuite, nous appliquons l'AFC sur le tableau de contingence obtenu. Notre outil (appelé HCAViz) analyse ce tableau de contingence de façon récursive et aide l'utilisateur à interpréter et interagir avec les résultats de l'AFC. D'abord, les résultats de la première AFC sur les images sont visualisés. L'utilisateur sélectionne ensuite un groupe d'images et fait une deuxième AFC sur le nouveau tableau de contingence. Ce processus peut continuer jusqu'à ce qu'un thème « pur » se dévoile. Ceci permet de découvrir une arborescence des thèmes dans une collection d'images. Une application sur la base Caltech-4 illustre l'intérêt de HCAViz dans la fouille d'images.

1 Introduction

La fouille de données (Fayyad et al., 1996) vise à extraire des connaissances utiles cachées à partir de grandes bases de données. Cette utilité est très liée au but des utilisateurs, c'est-à-dire que c'est l'utilisateur qui peut déterminer si les connaissances résultantes répondent à

son but. Il est donc préférable que les outils pour la fouille de données soient interactifs et permettent la participation des utilisateurs. L'idée ici est d'augmenter la participation humaine à travers des techniques de visualisation interactive dans l'environnement de fouille de données. Au cours de la dernière décennie, un grand nombre de méthodes de visualisation développées dans différents domaines ont été utilisées dans l'exploration des données et le processus d'extraction des connaissances (Fayyad et al., 2001; Keim, 2002). Les méthodes de visualisation sont utilisées pour la sélection des données (phase de pré-traitement) et la visualisation des résultats (phase de post-traitement). Certaines méthodes récentes d'extraction visuelle de données (Ankerst et al., 2000; Do et Poulet, 2004) essaient de faire participer plus intensivement l'utilisateur dans le processus de fouille de données par l'utilisation de la visualisation. Cette coopération peut apporter des avantages comme l'utilisation des connaissances du domaine au cours de la construction du modèle, l'amélioration de la confiance et de la compréhensibilité des modèles obtenus et l'utilisation des capacités de reconnaissance des formes de l'humain dans la phase de construction et exploration du modèle (Poulet, 2004).

En analyse de données textuelles, l'outil Bi-Qnomis (Kerbaol et al., 2006) développé par M. Kerbaol permet de visualiser les résultats et de trouver des thèmes pertinents dans un corpus textuel traité par une AFC. Les nuages de points (documents et mots) sont projetés sur un plan factoriel. Puis pour chaque axe, on fait une liste des mots ayant une forte contribution (correspondant à une métaclé). Ces mots sont affichés à gauche et le titre des documents bien représentés sur ce plan sont affichés en haut. L'utilisateur peut cliquer sur un titre pour voir en détail un document. En examinant des métaclés et des documents un expert aura ensuite le résumé du contenu de ces documents. Bi-Qnomis fournit aussi une visualisation des métaclés et leurs mots en utilisant un arbre hyperbolique. Par contre, pour les images on rencontre des difficultés car il n'y a pas de « vrais » mots au sens littéral dans les images. On utilise donc des « mots visuels » à la place des mots textuels et les images comme les documents. Nous nous intéressons ici à l'adaptation de l'AFC pour les images et à l'interprétation de ses résultats à l'aide d'indicateurs pertinents avec un outil de visualisation dans lequel l'utilisateur peut explorer les résultats de façon interactive pour mieux les comprendre. Notre outil (appelé HCAViz) analyse ce tableau de contingence de façon récursive et aide l'utilisateur à interpréter et interagir avec les résultats de l'AFC. D'abord, les résultats de la première AFC sur les images sont visualisés. L'utilisateur sélectionne ensuite un groupe d'images et fait une deuxième AFC sur le nouveau tableau de contingence. Ce processus peut continuer jusqu'à ce qu'un thème « pur » se dévoile. Ceci permet de découvrir une arborescence des thèmes dans une collection d'images.

L'article est organisé de la façon suivante : nous décrivons brièvement la méthode de l'AFC et son adaptation sur les images dans la section 2. La section 3 présente l'extraction interactive des connaissances avec l'AFC hiérarchique, avant la conclusion et les perspectives de ce travail.

2 Analyse factorielle des correspondances et représentation des images

2.1 Analyse factorielle des correspondances

L'AFC est une méthode exploratoire classique pour l'analyse de tableaux de contingence. Elle a été proposée par J. P. Benzécri (Benzécri, 1973) dans le contexte de la linguistique, c'est-à-dire pour l'analyse de données textuelles. La première étude a été réalisée sur les tragédies de Racine. L'AFC sur un tableau croisant des mots et des documents permet de répondre aux questions suivantes : y a-t-il des proximités entre certains mots ? Y a-t-il des proximités entre certains documents ? Y a-t-il des liens entre certains mots et certains documents ? L'AFC comme la plupart des méthodes factorielles utilise la décomposition en valeurs singulières (ou décomposition en valeurs propres et vecteurs propres) d'une matrice particulière et permet la visualisation des mots et des documents dans un espace de dimension réduit. Cet espace de dimension réduit a la particularité d'avoir un nuage de points projetés (mots et/ou documents) d'inertie maximale. Par ailleurs, l'AFC fournit des indicateurs pertinents pour l'interprétation des axes comme la contribution d'un mot ou d'un document à l'inertie de l'axe ou la qualité d'un mot et/ou d'un document sur un axe (Morin, 2004).

2.2 Représentation des images

Puisque l'AFC ne traite que des tableaux de contingence, pour l'adaptation aux images, nous devons représenter les images sous forme d'un tableau de contingence. Nous utilisons les « *résumés des descripteurs locaux* »pour représenter les images. Le tableau de contingence est obtenu en croisant les *images* et les *mots visuels*. Après obtention du tableau de contingence, l'application de l'AFC aux images est analogue aux textes. Dans ce cas, nous avons les *images* comme *documents* et les *mots visuels* comme *mots*. Les mots dans les images, appelés *mots visuels*, sont calculés à partir des descripteurs locaux pour constituer un vocabulaire de N *mots visuels*. La construction des mots visuels se fait donc en deux étapes : (i) calcul des descripteurs locaux pour un ensemble d'images, (ii) classification des descripteurs calculés en N *clusters*. Chaque *cluster* sera, par définition, un *mot visuel*.

Description locale par des points d'intérêt

Le calcul des descripteurs locaux dans une image consiste d'abord à détecter des points d'intérêt et à caractériser ces points. La figure 1 montre quelques points d'intérêt détectés par un détecteur Hessien-Affine (Mikolajczyk et Schmid, 2002). La description des points d'intérêt se fait sur la région autour de ces points. Nous avons utilisé le descripteur SIFT (Lowe, 2004) pour caractériser les points d'intérêt car il est le plus utilisé des descripteurs proposés dans la littérature. Chaque descripteur SIFT est un vecteur à 128 dimensions.

Classification des descripteurs locaux

La seconde étape de la construction des mots visuels consiste à former des mots visuels à partir des descripteurs locaux obtenus à l'étape précédente. La plupart des travaux effectuent un algorithme de k-moyennes sur les descripteurs locaux et prennent les centres de gravité de

(a) Image originale (b) Image avec points d'intérêts

FIG. 1 – *Points d'intérêt détectés par un détecteur Hessian-Affine. Un descripteur local sera calculé sur la région délimité par une ellipse.*

chaque *cluster* comme *mots visuels* (Sivic et al., 2005; Bosch et al., 2006). Chaque descripteur est ensuite affecté au *cluster* le plus proche et il est étiqueté par le *mot visuel* associé au *cluster* correspondant. Il s'agit d'une quantification vectorielle des descripteurs locaux.

Construction du tableau de contingence

Après avoir construit le vocabulaire visuel et affecté les descripteurs aux mots visuels correspondant, une image est caractérisée par la distribution des mots visuels (la fréquence des mots visuels qui apparaissent dans l'image). On obtient ainsi le tableau de contingence pour les images.

D'autres codages sont envisageables comme par exemple les *pixmots*. Les *pixmots* ont été introduits par M. Kerbaol et utilisés par F. Le Guillarm pour un projet informatique d'encodage d'images noirs et blancs (Le Guillarm, 2007).

3 Fouille d'images avec HCAViz

Les résultats de l'AFC sur l'analyse de données textuelles montre que les axes factoriels de l'AFC (déterminés par des groupes de mots, appelés les *métaclés*) correspondent à des concepts plus généraux que les mots individuels et appréhendent les *thèmes* dans un corpus (Morin, 2004; Kerbaol et al., 2006). L'application de l'AFC sur un tableau de contingence qui croise des images et des mots visuels permet de trouver des thèmes à un niveau sémantique plus élevé que les mots visuels individuels.

3.1 Projection sur le plan factoriel

Après avoir fait une AFC sur les images, on les projette sur les plans factoriels en s'aidant des indicateurs usuels de l'AFC : la contribution à l'inertie et la qualité de représentation. La figure 2 présente la projection des images de la base Caltech-4 (Sivic et al., 2005) sur le premier plan factoriel constitué par les deux axes 1 et 2.

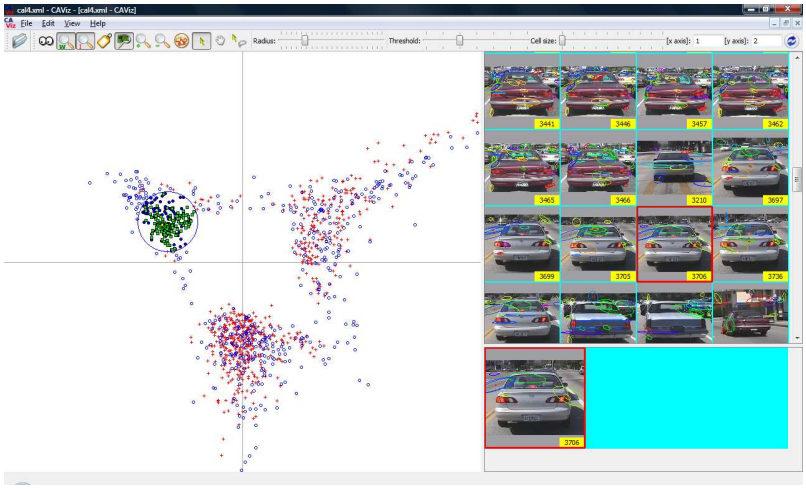


FIG. 2 – *Projection des images et mots visuels sur le plan factoriel constitué par les axes 1–2. Les images et mots visuels sélectionnés (points dans le cercle bleu) sont visualisés dans la partie droite de l'écran. Les mots visuels (sous forme des ellipses) sont superposés sur les images.*

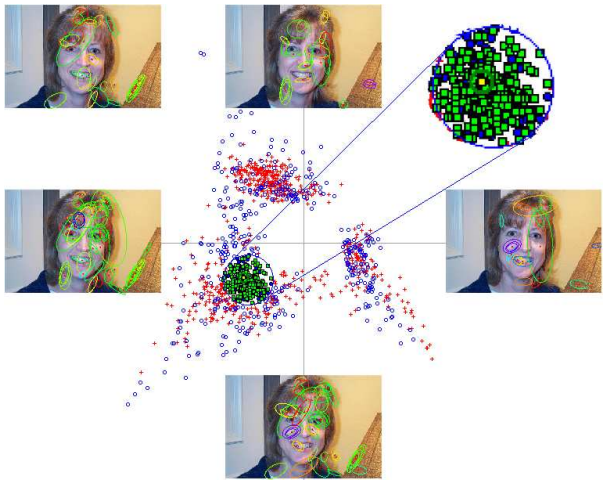


FIG. 3 – *Visualisation des métaclés*

L'écran est divisé en deux parties : à gauche, nous projetons les nuages de points (images et/ou mots visuels) et la partie droite est réservée à l'affichage des images sélectionnées. Un point-image s'affiche en rouge, un point-mot visuel est de couleur bleue. L'utilisateur peut sélectionner une image ou un groupe d'images en pointant sur l'image intéressante. Toutes les images se trouvant dans un voisinage de rayon r de l'image intéressante sont également affichées à droite. Ceci nous donne tout de suite une vue générale du contenu de ces images.

La couleur des points correspondant aux images sélectionnées sera modifiée de rouge en vert. Les mots visuels (sous forme d'une ellipse) seront superposés sur les images sélectionnées (voir la figure 2).

Pour se concentrer sur des images et/ou des mots intéressants, nous n'affichons, dans le plan, que les images et/ou les mots ayant une forte contribution, généralement 2 ou 3 fois la contribution moyenne. L'inertie totale sur un axe est égale à la valeur propre associée à cet axe. Le seuil est donc facile à déterminer. M. Kerbaol appelle métaclés les groupes des mots dont la contribution est très élevée sur un axe. Nous avons donc 2 métaclés par axe, une positive et une négative. Les mots visuels appartenant à chaque métaclé seront affichés sur l'image correspondante. Dans la figure 3, nous affichons les métaclés et leurs mots visuels. Les images se trouvant à chaque extrémité des axes correspondent à des métaclés. L'image en haut à gauche est superposée aux mots visuels proches d'elle. Il est facile de vérifier que la plupart de ces mots se trouvent à la fois dans la métaclé à gauche et la métaclé en bas.

3.2 Visualisation et extraction des connaissances

Après avoir visualisé les nuages de points sur le plan, on veut chercher les thèmes bien représentés dans le sous espace. C'est plus compliqué que dans le cas de textes car l'interprétation des mots visuels reste encore plus difficile. Cependant, en regardant sur un groupe d'images projetées, on peut voir ce qu'est le thème principal tandis qu'avec les textes on doit lire tout le texte pour découvrir le thème sous-jacent. Avec HCAViz, l'information concernant une image est également extraite de façon interactive en sélectionnant l'image en question. Il y a 2 indicateurs pertinents pour interpréter le résultat de l'AFC : la qualité de représentation d'une part et la contribution à l'inertie d'autre part. Ces informations nous aideront à certaines tâches ultérieures. La figure 4 montre un exemple de l'extraction interactive de ces informations. L'image dans cet exemple est bien représentée par les axes 2, 25, 3 et 5 et contribue beaucoup aux axes 2, 25, 16 et 23.

Un des avantages de l'AFC est que l'on peut visualiser des mots visuels et des images dans un même plan. Les mots visuels proches d'une image caractérisent bien cette image. Cette excellente propriété nous permet de concentrer sur un sous groupe d'images et de mots visuels et de faire l'AFC hiérarchique. Cela permet de « *zoomer* » sur le groupe d'images choisies. Les sous thèmes seront donc découverts. Il est aisé de visualiser les mots caractérisant bien une ou un groupe d'images (correspondant à un thème) en sélectionnant l'image et affichant les mots proches de cette image. La figure 5 montre les mots qui représentent bien le thème « visage ».

Sur un plan factoriel, un groupe de points proches les uns des autres peut définir un thème dans ce plan factoriel. Néanmoins, pour pouvoir découvrir des thèmes plus fins, il faut chercher les axes (le plan) qui représentent ces thèmes. HCAViz permet d'afficher les informations sur la qualité de représentation des images sur les axes et fournit un couplage des vues qui permet de voir la qualité de représentation d'un groupe d'images.

3.2.1 Qualité de représentation des images

Rappelons que la qualité de représentation d'un point i (correspondant à l'image i) sur l'axe α est le cosinus carré de l'angle entre l'axe α et le vecteur joignant le centre du nuage au point i. Si le cosinus carré d'une image sur un axe est grand, cet axe la représente bien. On utilise ce critère pour chercher les axes représentant bien les images et les mots. Les nuages de

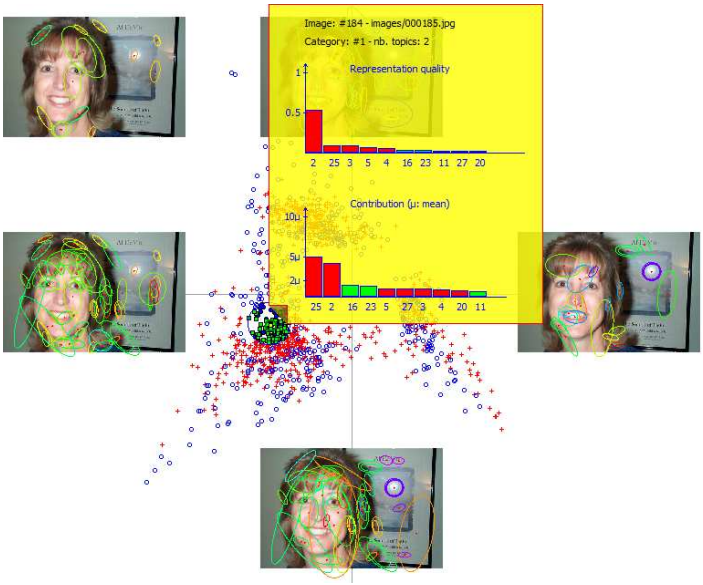


FIG. 4 – *Extraction interactive de l'information : histogramme de qualité de représentation et histogramme de contribution de l'image en examen (rouge : positif, bleu : négatif)*

FIG. 5 – *Les mots visuels caractérisant le thème « visage »*

points sont projetés sur le premier plan factoriel (par ex. les axes 1 et 2). On extrait la qualité de représentation des images sur les axes en regardant les premiers facteurs dans l'histogramme de qualité de représentation. Ces informations nous donnent un guide pour chercher le bon plan factoriel qui représente bien un groupe d'images.

3.2.2 Couplage des vues

CAViz affiche les nuages de points dans la partie gauche et les images sélectionnées dans la partie droite. En cliquant sur une image à droite, le point correspondant à gauche est sélectionné et ses informations sont aussi affichées. Ce couplage des vues nous permet de sélectionner facilement les images intéressantes. La perception visuelle humaine est un bon outil pour la reconnaissance de formes. Nous sélectionnons les images similaires à droite et regardons la qualité de représentation de ces images. Si on trouve que ce sont les mêmes axes qui représentent bien les images sélectionnées à droite, on prend alors ces axes pour projeter les points.

3.3 Découverte de thèmes

Nous présentons ici une étude de cas sur la découverte de thèmes dans la base Caltech-4 (Sivic et al., 2005). Cette base contient $4\,090$ images réparties en 5 catégories. Nous avons récupéré le tableau de contingence calculé par Sivic et al. (2005). Le vocabulaire se compose de 2,224 mots visuels.

Nous appliquons d'abord l'AFC sur les images. Ensuite, les points sont projetés dans le premier plan et on pointe sur un groupe à gauche (correspondant aux voitures). Les images sont donc affichées à droite. On commence à sélectionner des images à droite en cliquant et regardant la qualité de représentation. On constate par exemple qu'il y a des images bien représentées par les axes 6 (négatif) et 7 (positif) et d'autres images bien représentées par les axes 11 (négatif) et 12 (positif) (cf. Figure 6).

On projette ensuite les points selon ces axes. On trouve alors des thèmes plus fins qui contiennent des images très similaires. En projetant sur les 2 axes 6 et 7, on sélectionne un groupe de points en haut (correspondant à la contribution négative sur l'axe 6 et positive sur le 7) et on voit les images affichées à droite : ces images sont très similaires (ce sont les voitures blanches, cf. Figure 7). De même manière, sur les 2 axes 11 et 12, on découvre encore un autre thème pour les voitures (les voitures grises).

3.4 AFC hiérarchique par visualisation

Nous avons montré dans la section précédente, une méthode pour la découverte des thèmes d'images à l'aide d'un des indicateurs de l'AFC : la qualité de représentation. Cette qualité est un critère pour choisir le bon plan à projeter. Nous présentons maintenant une autre méthode, qui peut accompagner la méthode précédente. La méthode se base sur une propriété intéressante de l'AFC : la représentation simultanée des lignes (images) et des colonnes (mots visuels).

Les images et les mots visuels sont représentés sur un même plan factoriel. L'association des images et des mots visuels est interprétée par la distance (dans le plan) entre eux. Les mots visuels autour d'un groupe d'images caractérisent bien ce groupe. Nous pouvons donc nous concentrer sur un groupe d'images et leurs mots visuels associés (tous sont sélectionnés interactivement dans le plan) et nous faisons une AFC. Cela permet de « *zoomer* » sur ce groupe. Par conséquent, les sous thèmes peuvent se découvrir. Ce procédure peut être répété sur les sous thèmes pour découvrir encore des « *sous sous thèmes* » et ainsi de suite.

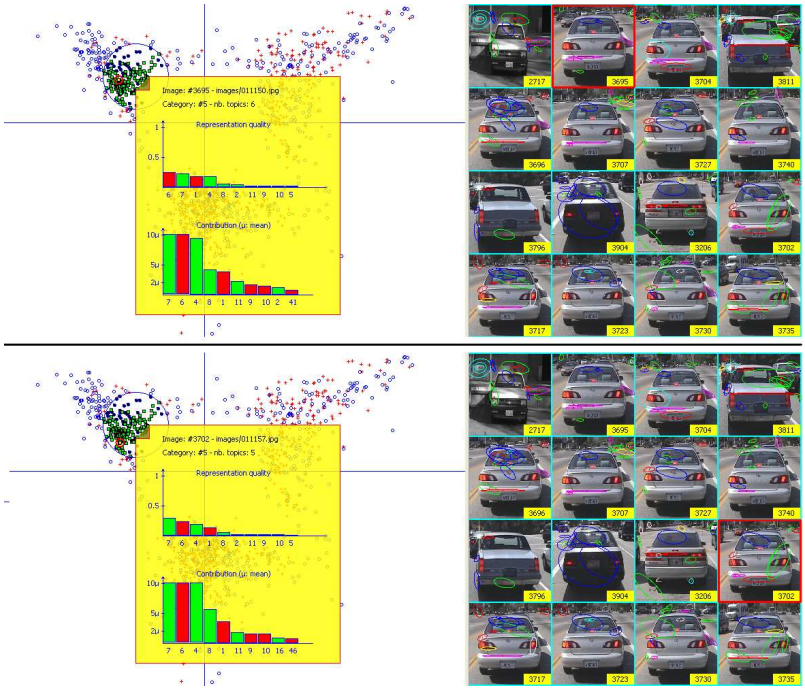
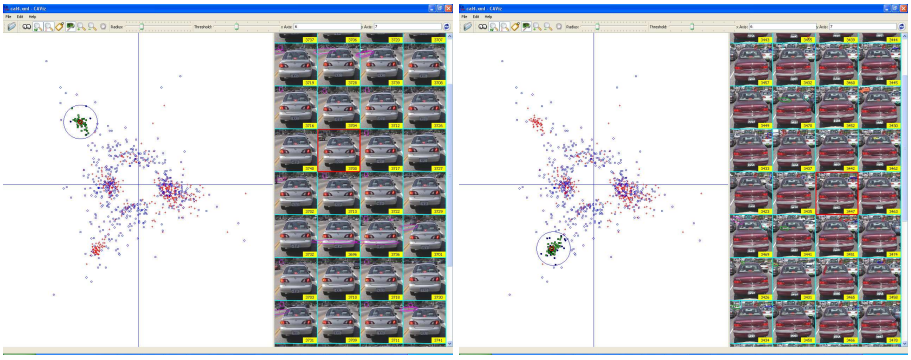


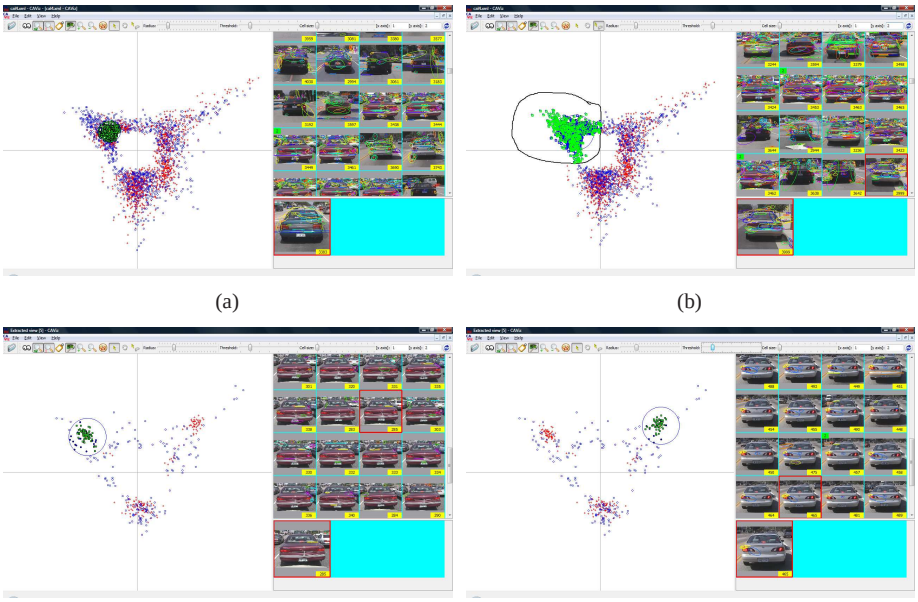
FIG. 6 – *Extraction des informations pour trouver de bons axes : deux images bien représentées par les 2 axes 6 (rouge : négatif) et 7 (vert : positif)*



(a) (b)

FIG. 7 – *Découverte des thèmes en projetant sur le bon plan factoriel : deux sous groupes de « voitures » sur le plan 6–7*

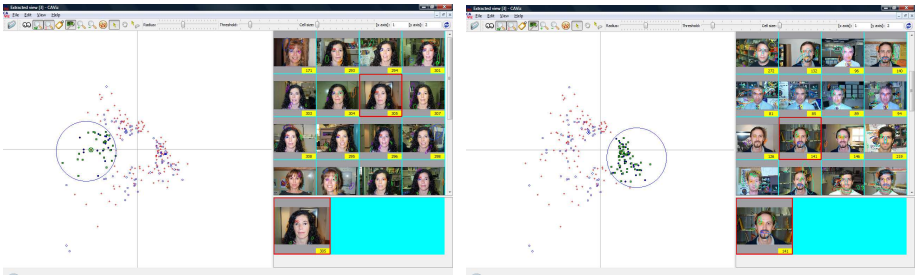
Se concentrer sur un seul groupe permet de trouver d'autres axes factoriels (d'autres thèmes seront donc trouvés) qui sont cachés à cause de l'influence des axes trouvés par la première AFC.



(a) (b)

(c) (d)

FIG. 8 – *Découvert des thème par AFC hiérarchique : (a) thème « voiture », (b) sélection des images et de mots visuels pour AFC, (c) thème « voiture rouge » et (d) thème « voiture blanche »*



(a) (b)

FIG. 9 – *Découvert des thème par AFC hiérarchique : (a) thème « visage », (b) thème « visage des femmes » et (b) thème « visage des hommes »*

Après avoir projeté les images dans le plan 1–2, on sélectionne un groupe d'images à gauche (le thème « voiture »), et on fait la deuxième AFC. Dans le premier plan de la 2e AFC, les images s'organisent en 3 groupes correspondant à 3 sous thèmes dont deux sont dans la figure 8.

De la même manière, deux sous thèmes « visage des hommes » et « visage des femmes » sont découverts dans le thème « visage » (voir la figure 9).

4 Conclusion et travaux futurs

Nous avons présenté un outil graphique interactif, appelé HCAViz, qui permet de visualiser et d'extraire des informations à partir des résultats de l'AFC sur les images afin de mieux comprendre les données et d'interpréter les résultats. Nous présentons aussi une application sur la découverte des thèmes d'images dans la base Caltech-4. Les thèmes obtenus dans cette étude ont montré l'intérêt de HCAViz pour l'interprétation de l'AFC en utilisant ses indicateurs comme la qualité de représentation et la contribution aux axes factoriels. Ils nous permettent de découvrir, par exemple, les sous groupes non répertoriés initialement dans la catégorie « voitures » qui correspondent à deux différentes couleurs et/ou type des voitures.

Références

Ankerst, M., M. Ester, et H.-P. Kriegel (2000). Towards an effective cooperation of the user and the computer for classification. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, (KDD'00)*, Boston, Massachusetts, États-Unis, pp. 179–188.

Benzécri, J. P. (1973). *L'analyse des correspondances*. Paris : Dunod.

Bosch, A., A. Zisserman, et X. Munoz (2006). Scene classification via pLSA. In *Proceedings of the European Conference on Computer Vision (ECCV'06)*, Graz, Autriche, pp. 517–530.

Do, T.-N. et F. Poulet (2004). Enhancing SVM with visualization. In E. Suzuki et S. Arikawa (Eds.), *Discovery Science*, Volume 3245 of *LNCS*, pp. 183–194. Springer Berlin / Heidelberg.

Fayyad, U., P.-S. G., et P. Smyth (1996). From data mining to knowledge discovery in databases. *AI Magazine 17*(3), 37–54.

Fayyad, U., G. G. Grinstein, et A. Wierse (Eds.) (2001). *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco, Californie, États-Unis : Morgan Kaufmann Publishers Inc.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics 8*(1), 1–8.

Kerbaol, M., J. Y. Bansard, et J. L. Coatrieux (2006). An analysis of IEEE publications. *IEEE Engineering in Medicine and Biology Magazine 25*(2), 6–9.

Le Guillarm, F. (2007). Encodage et reconstruction d'images grâce à l'analyse des correspondances. Rapport de projet XL, IRISA/ESIEA Laval.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*(2), 91–110.

Mikolajczyk, K. et C. Schmid (2002). An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision (ECCV'02)*, Copenhague, Danemark, pp. 128–142. Springer Verlag.

Morin, A. (2004). Intensive use of correspondence analysis for information retrieval. In *Proceedings of the 26th International Conference on Information Technology Interfaces (ITI'04)*, Cavtat, Croatie, pp. 255–258.

Poulet, F. (2004). SVM and graphical algorithms : A cooperative approach. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04)*, Brighton, Royaume-Uni, pp. 499–502.

Sivic, J., B. C. Russell, A. A. Efros, A. Zisserman, et W. Freeman (2005). Discovering objects and their location in images. In *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05)*, Pékin, Chine, pp. 370–377.

Summary

We propose an interactive graphical tool, HCAViz, which allows to visualize and to extract knowledge from images using Hierarchical Correspondence Analysis. Correspondence Analysis (CA) is originally for analysis of contingency tables. In order to adapt CA on images, we first define the "visual words" in images. These words are constructed from local descriptors (SIFT, Scale Invariant Feature Transform) in images. HCAViz analyses recursively a such contingency table. First, the results of the first CA are visualized. Then the user selects a group of images and performs the second CA on the new contingency table. This procedure can continue until a "pure" topic is discovered. An application to the Caltech-4 database shows the interest of HCAViz for image mining.