

Fusion des approches visuelles et contextuelles pour l'annotation des images médicales

Filip Florea^{*,**}, Valeriu Cornea^{*}, Alexandrina Rogozan^{*}
Abdelaziz Bensrhair^{*} Stefan Darmoni^{*,**}

^{*}Laboratoire LITIS - EA 4051, INSA de Rouen, France
filip.florea@insa-rouen.fr,
www.litislab.eu/

^{**}Equipe CISMef, CHU de Rouen, France
stefan.darmoni@chu-rouen.fr,
www.chu-rouen.fr/cismef/

Résumé. Dans le contexte de la recherche d'information sur Internet, nous proposons une architecture d'annotation automatique des images médicales, extraites à partir des documents de santé en ligne. Notre système est conçu pour extraire des informations médicales spécifiques (i.e. modalité médicale, région anatomique) à partir du contenu et du contexte des images. Nous proposons une architecture de fusion des approches contenu/contexte adaptée aux images médicales. L'approche orientée sur le contenu des images, consiste à annoter des images inconnues par la catégorisation des représentations visuelles compactes. Nous utilisons en même temps le contexte des images (les régions textuelles) ainsi que des ontologies médicales spécialement adaptées aux informations recherchées. Finalement, nous démontrons qu'en fusionnant les décisions des deux approches, nous améliorons les performances globales du système d'annotation.

1 Introduction

A l'heure actuelle, l'Internet est devenu une des sources d'information les plus importantes dans des nombreux domaines, comme celui de la santé. Afin de faciliter l'accès aux informations médicales disponibles en ligne, l'élaboration de nouveaux instruments et méthodes de recherche s'avère nécessaire. Le projet CISMef¹ (Catalogue et Index des Sites Médicaux Francophones) Darmoni et al. (2000) est le catalogue de santé lancé par le CHU de Rouen en 1995. L'objectif du catalogue est de décrire et de classer les principales ressources (documents sur le Web) de santé en français pour aider les utilisateurs dans leur recherche d'information médicale de qualité disponibles en ligne.

Des efforts considérables ont été engagés par l'équipe CISMef afin de développer des architectures d'indexation automatique, et des avancements significatifs ont été présentés Névéol et al. (2006). Cependant, l'indexation automatique a des limites et un des principaux problèmes reste la difficulté d'indexation des médias non textuels, comme *les images*.

¹<http://www.cismef.org>

2 Les travaux précédents

Il existe deux approches principales pour rechercher des images : en utilisant le *contenu* ou en utilisant le *contexte* de l'image (les régions textuelles associées aux images). Initialement, les images étaient indexées à l'aide des index reposant sur des mots clés Frankewitsch et Prokosch (2001). Au cours des années, des méthodes basées sur le contenu visuel des images ont été proposées pour l'annotation, l'indexation et la recherche des images médicales non annotées Lehmann et al. (2003) Müller et al. (2003). Dernièrement, nous avons remarqué un intérêt croissant pour les architectures qui proposent des descriptions d'images combinant les deux approches Deselaers et al. (2005) Besancon et Millet (2006).

Dans cet article, nous proposons une architecture combinant le contenu et le contexte pour l'annotation des images médicales, le but étant d'extraire au mieux des informations bien définies. Etant placés dans le contexte réel d'un catalogue en ligne, nous avons créé une application capable de traiter tous les aspects techniques de l'extraction d'information à partir de différents formats (cryptées, non structurées). En même temps, nous avons cherché à minimiser la sensibilité de notre système aux variations de contenu et de qualité spécifiques à l'Internet.

3 L'architecture du module MedIC

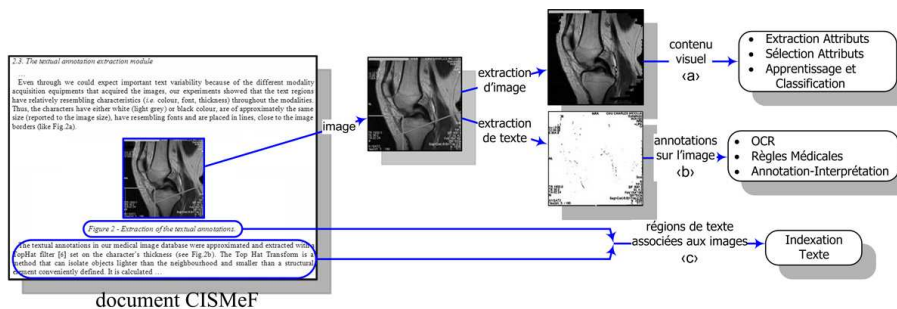


FIG. 1 – Le module MedIC

Dans le contexte du projet d'indexation automatique des documents de santé développé par l'équipe CISMef, le module MedIC (Medical Image Categorization) a comme tâche l'annotation des images médicales. Le but du module est de localiser, d'extraire et d'annoter les images médicales à partir d'un document donné. Le MedIC a été conçu pour rechercher plusieurs types d'informations médicales : la modalité médicale, la région anatomique, l'angle de vue d'acquisition et la pathologie. Ces informations permettront aux utilisateurs du catalogue CISMef, de formuler des requêtes orientées vers l'image, telles que <Trouve-moi les documents contenant des images ANGIOGRAPHIQUES (modalité) présentant un EMBOLISME (pathologie) PULMONAIRE (région anatomique)>.

L'architecture de module MedIC est présentée dans la figure 1. Pour extraire l'information médicale, nous traitons : (a) le contenu visuel des images Florea et al. (2006), (b) les annotations marquées directement sur l'image Florea et al. (2005), et (c) les régions textuelles associées aux images.

Même si la source ⟨b⟩ a été prouvée comme étant précise, elle est rarement utilisable en ligne à cause du manque d’annotations marquées sur la majorité des images publiées sur l’Internet. Dans cet article, nous évaluons la pertinence de l’information extraite à partir des sources ⟨a⟩ et ⟨c⟩ et le gain de performance obtenu en combinant les deux approches.

4 Les images et les régions textuelles associées

Les ressources médicales (i.e. documents) contiennent des quantités considérables d’informations relatives aux images. Habituellement, il y a deux régions textuelles associées aux images : *la légende* - courte et placée près de l’image et *le paragraphe* plus long et plus détaillé.

Pour les expériences d’extraction/annotation des images que nous présentons dans cet article, nous avons créé une base de 657 enregistrements, extraits automatiquement à partir des documents classés par CISMeF. Les images représentant les six principales modalités médicales : l’angiographie, l’échographie, l’imagerie à résonance magnétique, la radiographie standard, la tomographie par l’ordinateur (scanner) et la scintigraphie. Chaque modalité est liée à une hiérarchie de régions anatomiques et sous anatomiques. Pour nous permettre l’évaluation automatique des performances d’annotation, chaque enregistrement est manuellement annoté avec la modalité médicale et la région anatomique capturée par l’image.

5 L’annotation basée sur le contenu visuel (V)

La catégorisation des images médicales basée sur le contenu d’image, peut être un outil d’annotation très performant, dans le contexte de la recherche d’images dans des bases non annotées. Notre approche de catégorisation est basée sur la classification supervisée des représentations numériques des images, reposant sur des attributs de texture (e.g. *matrices de co-occurrence*, *dimension fractale*, les réponses aux *filtres de Gabor*, et autres) et statistiques. Une analyse en composantes principales (PCA) est employée pour réduire la dimensionnalité de l’espace des attributs. Finalement, un classifieur SVM (Support Vector Machines) est employé par MedIC, pour la projection des données de test dans les catégories correspondantes. Plus de détails sur l’approche « visuelle », sont disponibles dans Florea et al. (2006).

Les performances d’annotation de la modalité sur les 657 enregistrements de notre base sont présentées dans le tableau 1. Pour chaque information extraite (modalité et région anatomique), nous présentons les résultats dans la forme des précision (p), rappel (r) et f-mesure (f_m), des mesures souvent utilisées en « recherche d’informations ».

	p (%)	r (%)	f_m (%)
V - modality	75.97	51.76	60.76
V - anatomic region	45.52	30.26	32.76

TAB. 1 – *Décision visuelle (V)*

Les résultats d’annotation que nous avons obtenus en utilisant le contenu des images sont plus faibles que prévu. En utilisant la même architecture, nous avons obtenu des résultats bien meilleurs au cours de la campagne d’évaluation CLEF 2006 Florea et al. (2006). Les résultats

inférieurs sont dus principalement au fait que, pour les expérimentations présentées dans cet article, nous avons fait l'apprentissage SVM sur une base d'images de qualité sensiblement différente (meilleure) que notre base de 657 images de test. Généralement, avec des images extraites à partir des documents en ligne nous obtenons des résultats inférieurs, à cause de la faible résolution et compression élevée de ces images.

6 L'annotation basée sur des informations contextuelles (T)

Cette deuxième approche vise à traiter l'information portée par les régions textuelles associées aux images (*légendes* et *paragraphes*), pour extraire les mêmes annotations que pour l'approche visuelle : les modalités médicales et les régions anatomiques.

La première étape est de modéliser l'information à extraire sous forme de dictionnaires. Pour chaque information, des dictionnaires DELA ont été créés, basés sur la terminologie MeSH² et des termes et synonymes CISMef. Les dictionnaires devraient nous permettre de localiser les termes MeSH, sous les diverses formes qu'ils peuvent prendre en langage naturel. Les dictionnaires ont été manipulés en utilisant l'environnement linguistique INTEX/NOOJ³.

Les termes obtenus suite à l'application des dictionnaires, sont utilisés pour l'annotation des images, en employant une décision par vote majoritaire. Nous avons implémenté plusieurs stratégies de traitement des deux zones textuelles que nous disposons :

Légendes et paragraphes traitées ensemble $T(L+P)$ - les légendes et les paragraphes sont considérés comme ayant la même importance.

Priorité aux légendes $T(L \text{ en priorité})$ - nous traitons la légende en priorité, car souvent, la légende contient des informations plus précises et succinctes que le paragraphe.

Approche voisinage (pour les régions anatomiques) $T(L \text{ en priorité} + \text{voisins})$ - Les confusions les plus courantes concernent les régions anatomiques voisines (e.g. bras/coude, avant-bras/main). Par conséquent, nous avons ajouté des conditions supplémentaires, en définissant une table de voisinage. Même si cette stratégie de voisinage semble adaptée, les résultats concrets sont moins satisfaisants.

modalité				régions anatomiques			
approche	p(%)	r(%)	f _m (%)	approche	p(%)	r(%)	f _m (%)
$T(L+P)$	95.13	81.34	87.52	$T(L+P)$	76.56	42.88	50.93
$T(L \text{ en priorité})$	96.62	81.79	88.49	$T(L \text{ en priorité})$	73.86	45.23	53.13
				$T(L \text{ en priorité} + \text{voisins})$	73.45	45.23	53.03

TAB. 2 – Résultats de décisions modalité et régions anatomiques (T)

Les résultats pour l'annotation des modalités et des régions anatomiques sont présents dans le tableau 2. Une analyse détaillée nous a montré que les principaux responsables des erreurs sont : les légendes communes aux plusieurs images, les extractions erronées des couples image-texte ou les erreurs grammaticales dans le texte original.

Il est important de remarquer que même si cette approche est incapable de proposer une décision pour toutes les images (faible rappel), elle est très précise (surtout pour les modalités).

²<http://www.nlm.nih.gov/mesh/>

³<http://www.nooj4nlp.net/>

7 Fusion des décisions (F)

Dans cette section, nous allons évaluer le gain de performance obtenu après la combinaison des sources $\langle a \rangle$ et $\langle c \rangle$. Les résultats obtenus sont présentés dans le tableau 3.

Approche prioritaire texte $F(T \text{ en priorité})$ - la décision textuelle est traitée en priorité (avec aussi des légendes prioritaires), pour exploiter la bonne précision de cette approche. Pour pondérer les informations, nous avons utilisé les rangs cumulés de chaque décision pour les deux sources.

Approche équilibré $F(T+V)$ - nous utilisons les mêmes critères de rangs cumulés. Par rapport à l'approche précédente, nous notons des faibles améliorations.

Approche équilibrée, avec la modalité décidée $F(T+V)_{\text{mod}}$. Motivées par les bons résultats de la décision sur la modalité, nous avons proposé une troisième approche où nous extrayons les annotations sur les régions anatomiques en utilisant la décision sur la modalité. Cependant, dans la pratique, cette approche est moins efficace.

Les approches de fusion (F) que nous avons essayées présentent toutes des améliorations de performance comparées aux méthodes reposant sur le contenu visuel (V) et sur le texte associé aux images (T), considérées séparément. Même avec les résultats plus faibles de la décision visuelle, la f-mesure globale, après la fusion, affiche une amélioration significative. Combinant les deux décisions, nous perdons généralement un petit pourcentage sur la précision moyenne (comparée à la décision textuelle), mais nous obtenons des taux de rappel et f-mesure bien plus élevés.

modalité				régions anatomiques			
approche	p (%)	r (%)	f _m (%)	approche	p (%)	r (%)	f _m (%)
T(L en priorité)	96.62	81.79	88.49	T(L en priorité)	73.86	45.23	53.13
v	75.97	51.76	60.76	v	45.52	30.26	32.76
F(T en priorité)	94.98	90.21	92.48	F(T en priorité)	67.4	66.2	65.75
F(T+V)	95.19	90.14	92.54	F(T+V)	69.74	67.78	66.97
				F(T+V) _{mod}	67.24	65.67	65.37

TAB. 3 – Résultats fusion de décisions modalité et régions anatomiques (F)

8 Conclusion

Le but du module MedIC est de permettre l'extraction et l'annotation automatique des images médicales extraites à partir des documents de santé complexes. L'objectif est de fournir des annotations médicales précises pour l'indexation des documents et de leurs images attachées.

La catégorisation des représentations visuelles peut fournir des annotations précises pour des images médicales. En même temps, elle est dépendante de l'existence des images d'apprentissage déjà annotées et très sensible aux variations de qualité des images. L'approche textuelle que nous présentons dans cet article peut fournir des annotations médicales précises pour les images, mais, à son tour, elle est dépendante des dictionnaires linguistiques.

En perspective, nous allons développer cette architecture pour traiter des informations autres que les modalités et les régions anatomiques. L'architecture que nous avons présentée dans cet article est extensible en définissant des dictionnaires additionnels ou des classes d'apprentissage visuelles supplémentaires.

L'architecture présentée a été conçue pour être intégrée dans le module d'indexation automatique des documents du catalogue CISMéF, pour offrir aux utilisateurs un meilleur (et plus complet) outil de recherche d'information médicale sur Internet.

Références

- Besancon, R. et C. Millet (2006). Using text and image retrieval systems. lic2m experiments at imageclef 2006. *LNCS CLEF Workshop*, in press.
- Darmoni, S., J. Leroy, B. Thirion, F. Baudic, M. Douyére, et J. Piot (2000). Cismef : a structured health resource guide. *Meth Inf Med* 39(1), 30–35.
- Deselaers, T., T. Weyand, D. Keysers, W. Macherey, et H. Ney (2005). Fire in imageclef2005 Combining content-based image ret with text info. ret. *LNCS CLEF Workshop*, in press.
- Florea, F., A. Rogozan, A. Bensrhair, J.-N. Dacher, et S. Darmoni (2005). Modality categorisation by textual annotations interpretation. In *Medical Inforamtics Europe*, 1270–1275.
- Florea, F., A. Rogozan, V. Cornea, A. Bensrhair, et S. Darmoni (2006). MedIC/CISMéF at imageCLEF 2006. In *LNCS CLEF Workshop*, in press.
- Frankewitsch, T. et U. Prokosch (2001). Navigation in medical internet image databases. *Medical Informatics* 26(1), 1–15.
- Lehmann, T. M., M. O. Güld, C. Thies, B. Fischer, M. Keysers, D. Kohnen, H. Schubert, et B. B. Wein (2003). Content-based image retrieval in medical applications for picture archiving and communication systems. In *Medical Imaging*, 5033, 440–451.
- Müller, H., A. Rosset, J.-P. Vallée, et A. Geissbuhler (2003). Integrating content-based visual access methods into a medical case database. In *Medical Inforamtics Europe*, 480–485.
- Névéol, A., A. Rogozan, et S. J. Darmoni (2006). Automatic indexing of online health resources for a french quality controlled gateway. *Info. Proc. and Manag.* 42(3), 695 – 709.

Summary

In this paper we propose an automatic annotation architecture, for medical images extracted from on line documents. Our system is designed to extract specific medical information (i.e medical modality, anatomical region) using both the content and the context of the images. The content/context fusion architecture we present is specifically adapted to medical images. The image content, in the form of compact visual representations, is used for image categorization and annotation. At the same time the image context (textual regions) is interpreted using adapted medical ontologies. We show that by combining the two approaches, we improve the overall performances of the annotation system.