Fusion d'informations pour l'indexation de photos

Saïd Kharbouche [*] — **Michel Plu**[*] — **Patrick Vannoorenberghe** [**]

[*] *France Telecom Orange Labs*
2 Avenue Pierre Marzin, 22300 Lannion
{said.kharbouche, michel.plu}@orange-ftgroup.com

[**] *Laboratoire de Télédetection à Haute Résolution, Université Toulouse III*
118, route de Narbonne, 31062 Toulouse cedex 4
Patrick.Vannoorenberghe@cict.fr

RÉSUMÉ. *Cet article présente une méthode d'indexation de photos appliquée à la reconnaissance de personnes dans des photos personnelles afin de permettre à un utilisateur de les retrouver à partir de requêtes correspondant à des identifiants de personnes. Cette méthode utilise la fusion d'index* **-FUSINDEX-** *issus de l'analyse de la photo elle-même et de l'analyse des commentaires textuels et oraux qui lui ont été associés. Ces analyses sont effectuées par différents moteurs d'indexations dédiés qui sont intégrés au sein d'un système global d'indexation et de recherche de photos. Le cadre théorique utilisé par la méthode de fusion définie est celui des fonctions de croyance afin de gérer au mieux l'incertitude et l'imprécision des moteurs.*

ABSTRACT. *This paper presents a method for image indexing based on recognition of people in image in order to allow the user to find images with queries corresponding to identifiers of people. This method uses the fusion of index* **-FUSINDEX-** *resulting from the analysis of the photo itself and from the analysis of textual and oral comments associated to it. These analyses are carried out by various indexing engines which are integrated within a global system for image indexing and retrieval. The theoretical framework used by the defined fusion method is based on belief functions theory in order to manage the imprecision and the uncertainty of the outputs of the integrated indexing engines outputs.*

MOTS-CLÉS : *Indexation et recherche d'images, fusion d'index multimodaux, fonctions de croyances, reconnaissance de visages.*

KEYWORDS: *Image indexing and retrieval, multimodality index fusion, belief functions, face recognition.*

CORIA 2007, pp. 425–440
Saint-Étienne, 28–30 mars 2007

1. Introduction

De nos jours, il est devenu facile pour une personne de prendre un nombre très important de photos, de les stocker et de les envoyer à d'autres personnes. Ainsi, chacun se retrouve rapidement à posséder un nombre colossal de photos personnelles qu'il a lui-même prises ou qui lui ont été envoyées par sa communauté qui elle aussi ne fait que s'agrandir. La gestion de ces bases importantes d'images devient alors un réel problème des nombreux usagers de ces technologies liées à la photo numérique. Pour le grand public, il existe de multiples façons d'organiser des albums photos, mais les plus courantes sont de les organiser selon les personnages, selon les dates, selon les lieux ou selon les événements (Naaman *et al.,* 2005). Si la date, les évènements associés et le lieu sont des méta-données qui sont ou pourront être facilement associées à l'image, il n'en est pas de même pour l'identification de personnes. De plus, cette tâche est particulièrement difficile dans un contexte d'usage grand public où les personnes pouvant être prises en photo peuvent être nombreuses et la qualité de prise de vue est souvent loin d'être parfaite. Pour traiter ce problème, des outils de détection et de reconnaissance de visages ont été développés. Mais dans les conditions d'usages qui nous intéressent, l'indexation de photos personnelles pour le grand public, ceux-ci n'ont pas encore de performances suffisantes et produisent des résultats encore trop imprécis et incertains. Une autre manière d'identifier les personnes dans une image est d'exploiter d'éventuelles annotations textuelles ou vocales associées aux photos. Ces annotations sont souvent créées lors de l'envoi ou du partage des photos avec d'autres. On peut les trouver sous forme de commentaires dans des blogs personnels ou des services en lignes de partage d'album de photos ou peuvent aussi être envoyées comme texte associé aux images lorsque des photos sont par exemple envoyées à partir d'un mobile en utilisant un MMS (MultiMedia Messaging Service). Dans ce dernier cas, l'annotation vocale sera alors bien plus facile à créer qu'une annotation écrite. Mais là encore de nombreux problèmes existent pour reconnaître la personne référencée dans une annotation le plus souvent par l'intermédiaire de son prénom. Même lorsque l'annotation est écrite, un prénom peut par exemple référencer plusieurs personnes connues de l'utilisateur qui portent le même prénom, un prénom peut aussi être utilisé sans indiquer qu'une personne est dans cette photo (par exemple dans des annotations comme "le bureau de Stéphanie", "Photo prise par Jérôme", "un ami de Michel", ...) , un prénom peut aussi se trouver dans un nom de famille, dans de nombreux lieux (par exemple Dominique, Nancy, Florence, Mont Saint Michel, Paris) ou dans de nombreux noms commun parfois mal orthographiés (par exemple Coquille Saint Jacques). D'autres problèmes s'ajoutent lorsque l'on traite des annotations vocales avec par exemple des voyelles muettes sur des prénoms féminins (par exemple Michelle, Paule, Frédérique) ou lorsque l'on cite des prénoms composés (Jean François, Marie Pierre, Pierre Henri, ...). Pour aborder ces problèmes, nous avons souhaité fusionner les résultats issus des deux approches, l'analyse de l'image et l'analyse des annotations. Cette fusion permet de fiabiliser les index retrouvés à partir de différentes sources ou au contraire d'ignorer les index en conflit avec d'autres propositions. Pour cela, nous avons intégré tout d'abord différents moteurs d'indexation au sein d'un système de gestion et de partage de métadonnées nommé Someone (Agosto *et al.,* 2003) utilisé

pour la gestion, l'indexation et le partage de photos personnelles puis nous avons défini une méthode de fusion des index issus des différents moteurs intégrés. Pour gérer au mieux l'incertitude et l'imprécision de ces index à fusionner, nous avons résolument opté pour la théorie des fonctions de croyance (Shafer, 1976). Cette théorie a été choisie pour ses méthodes adaptatives de fusion d'information et d'identification du conflit entre des sources de fiabilité variable. Elle constitue de plus un cadre suffisamment générique pour englober les autres théories des mesures de confiance de probabilités imprécises comme la théorie des possibilités. En plus, les mesures classiques de probabilités peuvent être vues comme des fonctions de croyances particulières (Vannoorenberghe, 2003) permettant ainsi d'établir un lien entre information statistique et incertitude.

Cet article s'organise de la façon suivante. La première section présente un état de l'art sur la fusion d'index pour l'indexation d'images. La section suivante précise le cadre formel de la méthode développée basé sur les fonctions de croyance. Ensuite, nous préciserons dans les deux sections suivantes comment les informations issues des différents moteurs sont transformées en croyances qu'une personne est dans la photo traitée, puis comment ces croyances sont fusionnées pour conduire à une proposition qu'une personne est dans une image avec une certaine mesure de pertinence. Nous illustrerons enfin cette méthode sur deux exemples et finirons par une conclusion.

2. Fusion d'informations pour l'indexation multimodale

L'idée de fusionner différentes informations dans le but d'améliorer la qualité d'une indexation d'un document n'est pas nouvelle. La plupart des méthodes de fusion peuvent être regroupées en trois catégories :

Vote majoritaire : Le vote a été utilisée dans (Philipp-Foliguet *et al.,* 2006) pour fusionner les deux modalités : l'image et le texte. Les avantages de cette technique de fusion sont sa facilité de mise en œuvre, son adaptativité aux nombres de sources et surtout sa rapidité. Néanmoins, travaillant sur des listes, elle ne permet pas de prendre en compte la pertinence de chacune des sources. La méthode n'est pas non plus capable de gérer les conflits qui peuvent intervenir entre sources, ce qui rend son application un peu limitée.

Combinaison linéaire : Cette technique de fusion, simple et rapide, est utilisée dans plusieurs systèmes d'indexation multimédia (Naaman *et al.,* 2005). Cependant, elle reste une méthode critiquable (Wu *et al.,* 2004) à cause de son manque de tolérance au bruit et ne gére pas de manière explicite les dépendances entre les sources.

Théorie des probabilités : Elle permet de fusionner les informations apportées par différentes sources mais la modèlisation probabiliste n'apparaît pas toujours suffisante ni très bien adaptée pour représenter l'incertitude et modéliser les méconnaissances (Dempster, 1967, Vannoorenberghe, 2003). De plus cette théorie ne permet pas la gestion des conflits entre les sources puisque aucune autre information que la probabilité *a posteriori* n'est disponible.

Théorie des possibilités : L'un des avantages de cette méthode de combinaison est sa facilité de mise en œuvre et sa capacité à gérer le conflit entre sources. Elle permet aussi en représentant l'imprécision et l'incertitude par deux fonctions de traiter des problèmes non modélisables par des approches probabilistes classiques (Dubois *et al.,* 1992). Cette théorie reste un cas particulier de la théorie des fonctions de croyance. Dans le contexte de l'indexation multimédia, cette technique a été utilisée pour l'indexation d'images basée sur le contenu (Azzam *et al.,* 2005).

Fusion par des techniques d'apprentissage : Les vecteurs des différentes modalités sont regroupés dans un seul vecteur qui sera utilisé comme données d'entrées d'un algorithme de discrimination du style séparateur à vaste marge (SVM : Support Vector Machines) (Boser *et al.,* 1992) ou de classification comme l'algorithme EM (Expectation-Maximization) (Dempster *et al.,* 1977). Les SVM sont appliqués avec une grande efficacité dans les systèmes de reconnaissance des formes et à l'indexation multimédia (Adams *et al.,* 2003). Un des points faibles de cette technique est le fait qu'elle ne supporte pas des données incomplètes. L'algorithme EM a été utilisé dans l'algorithme SFA (Canny, 2002) pour la fusion d'informations incomplètes. C'est effectivement un de ces avantages à traiter des données manquantes comme c'est le cas pour les images prises par des cameraphones (Davis *et al.,* 2006). Malheureusement, l'hypothèse d'une loi de probabilité ne permet pas de gérer l'incertitude des événements. De plus, la méthode se contraint à un modèle rigide qui ne permet pas de s'adapter facilement à l'intégration de nouvelles sources d'information. En outre, elle considère que chacune des sources possède la même fiabilité.

Fonctions de croyance : De nombreux travaux s'attachent à comparer le cadre formel des fonctions de croyance à plusieurs autres méthodes de fusion notamment aux méthodes probabilistes (dans le cadre Bayesien), à la théorie des possibilités et à la combinaison linéaire. Pour la fusion d'informations multimodales (audio et vidéo), les fonctions de croyance ont été jugées plus performantes (Aslandogan *et al.,* 2000) que la combinaison linéaire, les méthodes Bayesiennes et la théorie des possibilités.

3. Cadre formel pour la fusion

3.1. *Fonctions de croyance et outils associés*

Les travaux de *Dempster* sur les bornes inférieure et supérieure d'une famille de distributions de probabilités (Dempster, 1967) ont permis à *Shafer* d'asseoir les bases de la théorie des fonctions de croyance (Shafer, 1976). On ne présente ici que les concepts de cette théorie que nous avons utilisées. Soient une forme inconnue X et, un ensemble Δ de q hypothèses possibles pour X, appelé cadre de discernement défini par :

$$\Delta = \{\omega_1, \omega_2, \ldots, \omega_q\}. \qquad [1]$$

Soit une source S qui donne des informations (dans notre cas sous forme de distance ou de score) sur l'appartenance de X à un sous-ensemble ou à plusieurs sous-ensembles de Δ. Un jeu de masses élémentaires, noté $m^{\Delta}[S, X]$ défini de 2^{Δ} (partition de Δ) dans $[0,\ 1]$, qui vérifie :

$$\sum_{A \subseteq 2^{\Delta}} m^{\Delta}[S, X](A) = 1. \qquad [2]$$

La masse $m^{\Delta}[S, X](A)$ représente le degré de croyance attribué à la proposition A et qui n'a pas pu, compte tenu de l'état de la connaissance, être affectée à un sous-ensemble plus spécifique que A. Les sous-ensembles $A \subseteq \Delta$ tels que $m^{\Delta}[S, X](A) > 0$ sont appelés éléments focaux. A partir de $m^{\Delta}[S, X]$, la fonction de croyance $bel^{\Delta}[S, X]$ est une fonction de 2^{Δ} vers $[0,\ 1]$ définie par :

$$bel^{\Delta}[S, X](A) = \sum_{B/B \subseteq A} m^{\Delta}[S, X](B) \qquad \forall\ A \subseteq \Delta. \qquad [3]$$

où $bel^{\Delta}[S, X](A)$ quantifie la crédibilité du sous-ensemble $A \subseteq \Delta$.

Supposons que nous disposons de deux sources d'informations S_1 et S_2 pour identifier la variable X et les deux allocations de masses $m^{\Delta}[S_1, X]$ et $m^{\Delta}[S_2, X]$ associées définies toutes deux sur Δ. Ces deux fonctions peuvent être agrégées par l'opérateur de combinaison de Dempster noté $\oplus$ et défini pour tout $A \subseteq \Delta$:

$$m^{\Delta}[S_1, S_2, X](A) = \frac{\sum_{B \cap C = A} m^{\Delta}[S_1, X](B).m^{\Delta}[S_2, X](C)}{1 - \sum_{B \cap C = \emptyset} m^{\Delta}[S_1, X](B).m^{\Delta}[S_2, X](C)}. \qquad [4]$$

L'utilisation de la règle de *Dempster* est possible si les fonctions de masse $m^{\Delta}[S_1, X]$ et $m^{\Delta}[S_2, X]$ ne sont pas en conflit total où si $\sum_{B \cap C = \emptyset} m^{\Delta}[S_1, X](B).m^{\Delta}[S_2, X](C) \neq 1$.

L'affaiblissement par un coefficient $0 \leq \alpha \leq 1$ des informations issues d'une source S permet de transférer une partie de la croyance vers l'ensemble Δ. Ainsi, une fonction de masse affaiblie, notée $m^{\Delta}_{\alpha}[S, X]$, peut se déduire de $m^{\Delta}[S, X]$ par :

$$m^{\Delta}_{\alpha}[S, X](A) = \alpha\ m^{\Delta}[S, X](A) \qquad \forall A \subsetneq \Omega \qquad [5]$$

$$m^{\Delta}_{\alpha}[S, X](\Delta) = 1 - \alpha + \alpha\ m^{\Delta}[S, X](\Delta). \qquad [6]$$

Au niveau pignistique, une fonction de croyance unique, sorte de résumé exhaustif de l'information disponible au niveau crédal, est utilisée pour la prise de décision. En basant son raisonnement sur des arguments de rationalité développés dans le modèle des croyances transférables, Ph.Smets (Smets, 1994) propose de transformer une fonction de masse $m^{\Delta}[S, X]$ en une fonction de probabilité $BetP_{m^{\Delta}[S,X]}(.)$ définie sur Δ (appelée fonction de probabilité pignistique) qui se formalise pour tout ω_i par :

$$BetP_{m^{\Delta}[S,X]}(\omega_i) = \frac{1}{1 - m^{\Delta}[S, X](\emptyset)} \sum_{\omega_i \in A} \frac{m^{\Delta}[S, X](A)}{|A|} \qquad [7]$$

où $|A|$ représente la cardinalité du sous-ensemble $A \subseteq \Delta$. Dans cette transformation, la masse de croyance $m^{\Delta}[S, X](A)$ est uniformément distribuée parmi les éléments de A.

3.2. *Notations*

Par la suite, nous appelerons document une image à indexer et l'ensemble de ces annotations vocales et écrites. De façon à exploiter la complémentarité des différents médias de chaque document, le système **FUSINDEX** exploite les informations recueillies par différents moteurs d'indexation. Les informations utilisées pour l'indexation sont fournies de manière automatique par :

– un outil d'analyse automatique d'une image (Visani, 2005) lui-même composé d'un détecteur de visage, noté DV et d'un reconnaisseur de visage, noté RV ;

– un outil de reconnaissance du genre homme/femme d'une personne dans l'image, noté H/F ; cet outil a été simulé en respectant les performances d'outils existants (Cheng *et al.,* 2001) ;

– un outil de reconnaissance de prénoms dans une annotation vocale (Charlet *et al.,* 2005), noté RPa ;

– un outil de reconnaissance de prénoms dans une annotation écrite, noté RPt ; cet outil utilise un analyse syntaxique de langues naturelles (Smits *et al.,* 2006).

Les autres notations seront utilisées. La base des personnes connues, également appelée par la suite 'carnet d'adresses', est définie par :

$$B = \{P_1, P_2, \ldots, P_i, \ldots, P_n\} \quad [8]$$

où chaque personne P_i peut être représentée par le triplet $P_i = (V_i, N_i, F_i)$ où :

– $V_i = \{V_{i,1}, V_{i,2}, \ldots, V_{i,h(i)}\}$ l'ensemble des $h(i)$ visages de la personnes P_i ;

– N_i : le nom, le prénom ou le surnom de la personne P_i ;

– F_i : la probabilité que la personne P_i soit une femme ; cette information est extraite d'une base statistique du sexe des personnes auxquelles sont attribuées le prénom correspondant à P_i complétée par des informations fiables sur le sexe de la personne identifiée dans des images n'ayant qu'une personne et annotées explicitement par P_i ;

– i : l'identifiant de la personne P_i pour $i = 1, 2, \ldots, n$.

Les visages $V_{i,.}$ sont utilisés par le reconnaisseur de visages en terme d'apprentissage et sont extrait d'images ne contenant qu'une personne et explicitement annotées par P_i. Pour une image à indexer I_k, on suppose que DV a détecté z zones susceptibles être des visages (un visage par zone) :

$$Z = \{Z_1, Z_2, \ldots, Z_j, \ldots, Z_z\}. \quad [9]$$

Il n'y a pas des recouvrements entre les zones détectées, $Z_j \cap Z_{j'} = \emptyset$ si $j \neq j'$.

Pour identifier l'occupant de la zone Z_j, on définit le cadre de discernement Ω, de la façon suivante :

$$\Omega = \{P_1, P_2, \ldots, P_i, \ldots, P_n, Inconnu, Inconnue, *\} \quad [10]$$

où $Inconnu$ (respectivement $Inconnue$) représente une personne inconnue (c'est-à-dire une personne hors du carnet d'adresses) de sexe masculin (respectivement féminin) et où le symbole $\{*\}$ est utilisé pour matérialiser le fait que la zone ne contient pas de visage. Le cadre de discernement Ψ défini par :

$$\Psi = \{Per, \overline{Per}\} \quad [11]$$

va permettre de quantifier le degré de croyance associé à chaque proposition qui sera faite à l'utilisateur qu'une personne est dans une image. On appelle par la suite **métadonnée** cette proposition. L'hypothèse Per représente l'hypothèse de pertinence indiquant que la métadonnée est pertinente d'un point de vue de l'indexation du document I_k. L'hypothèse $\overline{Per}$ est le complémentaire de Per dans Ψ et représente la non pertinence. A partir de ces métadonnées proposées et leur fiabilité, l'utilisateur choisit celles à utiliser pour indexer l'image.

4. Processus de fusion

Dans le cadre de cette étude, les modèles choisis pour les allocations de masses sont fortement inspirés du modèle des distances de *Th. Denœux* (Denœux, 1995) qui évalue l'appartenance d'une variable inconnue X à une classe C par une fonction Φ de la distance $d(X, C)$ définie par :

$$\Phi(d(X, C)) = \tau \exp^{-\delta . d(X,C)^{\kappa}} \quad [12]$$

où τ, δ et κ sont des paramètres à déterminer. Cette distance, qui possède des propriétés intéressantes (décroissante, comprise entre 0 et 1, ...), a été utilisée avec succès dans plusieurs applications (Martin, 2005). Dans la phase de modélisation des paramètres, nous avons utilisé 200 images annotées avec une base de quinze personnes où les visages de chacune de ces personnes ont été appris par l'outil RV à partir de quatre images.

4.1. *Détecteur de visage :* DV

Soit une zone Z_j détectée par DV dans l'image du document I_k. On cherche à représenter l'incertitude inhérente à l'information fournie par le détecteur DV par une fonction de masse qui sera notée $m^{\Omega}[DV, Z_j]$ définie sur Ω. Le moteur DV donne un score $s(Z_j) \geq 0$ sur l'hypothèse qu'une zone donnée Z_j englobe effectivement un visage. Plus la valeur de ce score est élevée, plus la détection est bonne. Ces constats nous amènent à évaluer le degré de croyance associé au détecteur de visage DV dans

le cadre de discernement Ω. Ainsi, l'allocation de masse générée par le DV pour la zone Z_j peut être formulée par :

$$m^{\Omega}[DV, Z_j](B, Inconnu, Inconnue) = 1 - \tau_1 \exp^{-\delta_1 . s(Z_j)^{\kappa_1}}, \quad [13]$$

$$m^{\Omega}[DV, Z_j](*) = \tau_1 \exp^{-\delta_1 . s(Z_j)^{\kappa_1}} \quad [14]$$

avec τ_1, η_1 et κ_1 des paramètres positifs à optimiser avec la base d'apprentissage.

4.2. *Reconnaisseur de visage :* RV

Pour l'apprentissage du moteur RV, on dispose d'un ensemble d'images de visages V_i pour chaque personne P_i du carnet d'adresses. A partir de cette base d'apprentissage, le moteur fournit une distance $d_{i,j}$ entre la zone Z_j (détectée et localisée par le détecteur DV) et l'ensemble des visages V_i de la personne P_i (Visani, 2005). Plus la distance $d_{i,j}$ est petite, plus la reconnaissance de la personne P_i dans la zone Z_j est bonne. Le taux de bonne reconnaissance calculé pour le moteur RV est une fonction estimée de la forme :

$$\tau_2 \exp^{-\delta_2 . d_{i,j}} . \quad [15]$$

Le degré de croyance associé au fait que la personne P_i a été reconnue dans la zone Z_j sera calculé à partir d'un score, noté $s(Z_j, P_i)$, défini par :

$$s(Z_j, P_i) = \min(\tau_2 \exp^{-\delta_2 . d_{i,j}}, 1) \quad [16]$$

qui est normalisé et compris dans l'intervalle $[0, 1]$. Cette fonction de score, calculée pour toutes les personnes P_i, peut être interprétée comme une distribution de possibilité. Cette distribution peut être transformée en une fonction de masse dont les éléments focaux sont emboités (Dubois *et al.,* 1982). On obtient les éléments focaux :

$$m^{\Omega}[RV, Z_j](A_1), \ldots, m^{\Omega}[RV, Z_j](A_l), \ldots, m^{\Omega}[RV, Z_j](A_n) \quad [17]$$

$$m^{\Omega}[RV, Z_j](Inconnu \cup Inconnue \cup *) \quad [18]$$

avec $A_l \subseteq \{P_1, P_2, \ldots, P_n\}$.

4.3. *Reconnaisseur de genre homme/femme* H/F

L'outil d'indexation H/F consiste à identifier le genre homme/femme de la personne qui se trouve dans la zone Z_j afin de favoriser les personnes du carnet d'adresses B du même sexe. Pour quantifier l'incertitude inhérente au sexe de la personne, nous allons définir son degré de croyance dans un cadre de discernement, noté Θ, défini par :

$$\Theta = \{H, F\} \quad [19]$$

où H représente l'hypothèse d'un visage masculin et F un visage féminin. L'outil H/F a été simulé par un simulateur de taux de bonne réponse de 90%. Afin d'utiliser l'outil H/F dans le processus d'identification de la zone Z_j, il faut donc une étape préliminaire (qui se déroule hors-ligne) consistant à déterminer le genre H/F des personnes du carnet d'adresses. Cette étape, qui ne sera pas expliquée en détail dans cet article, détermine la probabilité qu'une personne connue P_i soit une femme F_i ou un homme $H_i = 1 - F_i$.

4.3.1. *Identification du genre H/F dans la zone Z_j*

L'outil d'indexation H/F donne une information sur le genre homme/femme de la personne dans une zone Z_j mais n'est pas capable de donner une information plus précise sur son identité. En effet, si par exemple l'outil émet l'hypothèse que la personne détectée est du sexe masculin, favoriser la reconnaissance de toutes les personnes de sexe masculin dans le carnet d'adresses augmente trop l'imprécision. C'est pourquoi l'approche que nous avons retenue est de défavoriser les personnes de sexe féminin du carnet d'adresses. Soient $d_{h,j}$ et $d_{f,j}$ les distances fournies par l'outil H/F représentant respectivement la distance entre la personne détectée dans la zone Z_j et une classe de visages d'hommes (respectivement de femmes). A partir de ces deux distances, on définit l'écart absolu, noté $d_{hf,j}$ par :

$$d_{hf,j} = |d_{h,j} - d_{f,j}|. \quad [20]$$

Nous avons choisi de modéliser le jeu de masses associé par :

$$m^{\Theta}[H/F, Z_j](A) = 1 - \tau_3 \exp^{-\delta_3 d_{hf,j}^{\kappa_3}} \quad [21]$$

$$m^{\Theta}[H/F, Z_j](\Theta) = 1 - m^{\Theta}[H/F, Z_j](A). \quad [22]$$

où $A \subseteq \Theta$ désigne le singleton $\{H\}$ ou $\{F\}$ dont la distance $d_{h,j}$ respectivement $d_{f,j}$ est la plus faible. On déduit de cette fonction la probabilité pignistique sur l'hypothèse $\{F\}$ pour la zone Z_j qui permet de quantifier la probabilité que la personne détectée dans la zone soit de sexe féminin.

Cette information permet, par passage du cadre Θ au cadre Ω, de construire le jeu de masses final associé à l'outil H/F pour chaque zone détectée de l'image à indexer qui sera noté $m^{\Omega}[H/F, Z_j]$.

4.4. *Fusion d'informations pour la zone Z_j*

Dans un premier temps, les informations pour chaque zone détectée utilisent uniquement les informations issues des outils d'indexation DV, RV et H/F pour identifier la personne qu'elle englobe. On fusionne ces trois sources d'informations pour obtenir une source d'information Image notée I'. Ainsi, $m^{\Omega}[I', Z_j]$ désigne le jeu de

masses issu de la fusion entre les trois fonctions relatives à ces moteurs par l'opérateur de combinaison de *Dempster*, calculé pour chaque zone Z_j par :

$$m^{\Omega}[I', Z_j] = m^{\Omega}[DV, Z_j] \oplus m^{\Omega}[RV, Z_j] \oplus m^{\Omega}[H/F, Z_j]. \quad [23]$$

Cette étape de fusion permet d'agréger les propositions générées par les trois moteurs capables d'extraire des informations de l'image sur chaque zone Z_j.

Dans un second temps, on prend en compte les informations issues des zones voisines pour diminuer l'incertitude pour chaque zone Z_j. L'idée consiste à utiliser le fait que si une personne P_i est dans la zone Z_j, elle ne peut pas se trouver dans une autre zone $Z_{j'}$ ($j' \neq j$). Les informations extraites des zones voisines vont être propagées entres les zones afin de diminuer l'ambiguïté sur certaines zones. Soit $m^{\Omega}_{\alpha_{Z_{j'}}}[Z_{j'}, Z_j]$ le jeu de masse extrait de la source d'information "zone $Z_{j'}$" avec $j' \in \{1, 2, \ldots, z\}$ et $j \neq j'$. Nous avons choisi de calculer ce jeu de masse de la façon suivante :

$$m^{\Omega}_{\alpha_{Z_{j'}}}[Z_{j'}, Z_j](\overline{P_i}) = \alpha_{Z_{j'}}\, bel^{\Omega}[I', Z_{j'}](P_i), \quad [24]$$

$$m^{\Omega}_{\alpha_{Z_{j'}}}[Z_{j'}, Z_j](\Omega) = 1 - m^{\Omega}_{\alpha_{Z_{j'}}}[Z_{j'}, Z_j](\overline{P_i}) \quad [25]$$

où $\alpha_{Z_{j'}}$ est un coefficient de fiabilité (cf. équation (5)) associé à la source d'information "zone $Z_{j'}$". Dans l'équation (24), $bel^{\Omega}[I', Z_{j'}](P_i)$ représente la croyance totale dans le fait que P_i soit dans la zone $Z_{j'}$ basée sur les informations extraites des sources DV, RV et H/F (cf. équation (23)). Pour limiter la propagation d'erreurs, les informations issues de la zone $Z_{j'}$ seront affaiblies en fonction :

– de la valeur du conflit interne $m^{\Omega}[I', Z_{j'}](\emptyset)$ (équation (23) avant normalisation)

– de la différence entre les probabilités pignistiques $BetP_{m^{\Omega}[I', Z_{j'}]}$ des deux personnes les mieux classées dans la zone $Z_{j'}$.

Ainsi, la fiabilité $\alpha_{Z_{j'}}$ est définie par :

$$\alpha_{Z_{j'}} = (1 - m^{\Omega}_{\textcircled{\scriptsize}}[I', Z_{j'}](\emptyset)).(BetP_{m^{\Omega}[I', Z_{j'}]}(P_{i'}) - BetP_{m^{\Omega}[I', Z_{j'}]}(P_{i''})).$$

où $P_{i'}$ et $P_{i''}$ sont les personnes les mieux classées dans l'ordre obtenu par le tri des probabilités pignistiques. Enfin, la source IZ rassemble toutes les zones de l'image sauf la zone Z_j elle-même ce qui conduit à :

$$m^{\Omega}[IZ, Z_j] = \bigoplus_{j' \in \{1,2,\ldots,z\}/j' \neq j} m^{\Omega}_{\alpha_{Z_{j'}}}[Z_{j'}, Z_j]. \quad [26]$$

On déduit enfin la fonction de croyance issue des outils d'indexation capables d'extraire l'information dans une zone à partir de l'image par sa fonction de masse :

$$m^{\Omega}[I, Z_j] = m^{\Omega}[I', Z_j] \oplus m^{\Omega}[IZ, Z_j]. \quad [27]$$

Cette fonction permet de quantifier la croyance dans le fait qu'une personne est effectivement présente au sein de la zone Z_j.

4.5. *Croyances issues de l'image I pour une métadonnée*

Les outils d'indexation DV, RV et H/F s'attachent à fournir des métadonnées pour chaque zone isolée Z_j de l'image. Cependant, notre objectif d'indexation consiste à générer des métadonnées pour l'image entière ce qui nécessite d'agréger les informations en provenance de chaque zone Z_j détectée. Il s'agit donc de passer de la croyance sur le fait qu'une personne est présente dans l'image (quantifiée par les fonctions de masse $m^{\Omega}[I, Z_j]$ pour chaque zone Z_j et définie dans le cadre de discernement Ω) à la pertinence que cette personne soit dans l'image représentée par une fonction de masse $m^{\Psi}[I, P_i]$ définie dans le cadre de discernement $\Psi = \{Per, \overline{Per}\}$. Ce jeu de masses $m^{\Psi}[I, P_i]$ est calculé à partir du maximum de la probabilité pignistique $BetP_{m^{\Omega}[I,Z_j]}$ déduite de chaque zone par :

$$m^{\Psi}[I, P_i](Per) = \max_j \{BetP_{m^{\Omega}[I,Z_j]}(P_i)\}, \quad [28]$$

$$m^{\Psi}[I, P_i](\Psi) = 1 - m^{\Psi}[I, P_i](Per). \quad [29]$$

Cette solution semble assez raisonnable puisqu'il suffit que la personne P_i ait été bien identifiée dans une zone pour dire que cette personne se trouve dans l'image.

4.6. *Analyse des commentaires*

Chacune des photos peut être commentée vocalement ou textuellement par son propriétaire et par les individus qui la partagent. Les informations fournies par les deux outils d'indexation qui traitent les fichiers sonores et texte (RPa et RPt) sont traitées et fusionner en utilisant directement le cadre de discernement Ψ, qui permet de quantifier la pertinence des métadonnées proposées. Ce choix est motivé par le fait qu'une photo peut avoir plusieurs commentaires vocaux et textuels dans lesquels on peut retrouver plusieurs prénoms détectés. Chaque commentaire est associé de manière globale à l'image, donc pour l'ensemble des zones Z_j. De plus, les commentaires n'étant pas forcément exhaustifs, ils n'ont pas le pouvoir de confirmer l'absence d'une personne dans l'image. Le fait qu'un prénom ne soit pas trouvé dans une annotation ne signifie pas que la personne associée soit absente de la photo. Les commentaires sont donc des sources qui ne peuvent avoir un effet que de renforcement dans la pertinence d'une métadonnée et ne sont pas en mesure de la contester. Ainsi leurs fonctions de masse respectives sont déduites en allouant de la masse aux deux éléments suivants : Ψ et Per. La manière précise de modéliser l'incertitude et l'ambiguïté ne sera pas détaillée dans cet article par soucis de brièveté, mais elle permet d'obtenir le jeu de masse noté $m^{\Psi}[Com, P_i]$ pour chaque P_i reconnu dans les commentaires. Ce jeu de masse permet de quantifier la pertinence de chacune des métadonnées selon les commentaires Com, avant d'être fusionné avec l'image.

4.7. *Fusion entre image I et commentaires Com*

Pour des raisons de non-exaustivité et de globalité des commentaires, la fusion d'informations entre l'image I et les commentaires Com ne permettront que de renforcer certaines méta-données. Ainsi, comme nous avons vu ci-dessus, leurs fonctions de masse respectives sont déduites en allouant de la masse à Per et Ψ. Par conséquent, le conflit entre l'image et les commentaires sera nul puisque $\Psi \cap Per = Per \neq \emptyset$. La fusion conduit au jeu de masse noté $m^{\Psi}[I, Com, P_i]$ obtenu par l'opérateur de Dempster. Enfin, les métadonnées seront triées et présentées à l'utilisateur selon leurs probabilités pignistiques respectives évaluées sur l'hypothèse Per :

$$BetP_{m^{\Psi}[I,Com,P_i]}(Per). \qquad [30]$$

5. Illustration sur des exemples

Dans cette section, on propose deux exemples qui permettent, dans un cas d'usage réel, de mettre en évidence les avantages du système de fusion d'index **FUSINDEX** avec une base de personnes défini par :
$B = \{OlivierT, OlivierC, Michel, PascalB, Irene, Muriel, Sonia, PascalF\}$.
Chaque personne de ce carnet d'adresses possède entre 3 et 6 images dans la base d'apprentissage du reconnaisseur de visages RV.

5.1. *Exemple 1*

Le document traité dans cet exemple est proposé à la figure 1. On dispose d'une

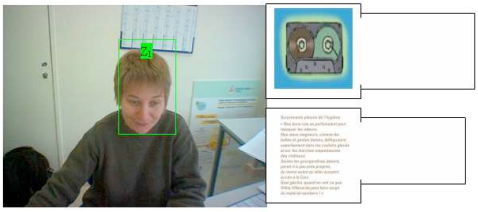

Figure 1. *Document de l'exemple 1 à indexer : image sans commentaire.*

image de $Muriel$ pour laquelle le détecteur de visage a détecté la zone avec un score $s(Z_1) = 35$. Les distances obtenues par le reconnaisseur de visages RV pour cet exemple sont telles que la distance minimale a été trouvée pour $OlivierC$ et que $Muriel$ n'a été classé qu'au second rang par le moteur RV. Le reconnaisseur de genre H/F fournit les distances suivantes : $d_{f,1} = 224439$ et $d_{h,1} = 360288$ permettant ainsi d'avoir une opinion sur le fait que le visage détecté est celui d'une femme. Nous

allons voir comment **FUSINDEX** permet de corriger cette erreur de reconnaissance du moteur RV.

Le tableau 1 présente les crédibilités et les probabilités pignistiques calculées par le système après la fusion des informations sur la zone détectée pour chaque métadonnée. On constate que dans la liste des métadonnées (ordonnées par probabilité pignis-

Prénoms	$bel^{\Psi}[I, Com, P_i](Per)$	$BetP_{m^{\Psi}[I,Com,P_i]}(Per)$
Muriel	**0.156**	**0.578**
Olivier C	**0.016**	**0.508**
Irène	0.124	0.562
Sonia	0.124	0.562
Michel	0.008	0.504
Pascal B	0.008	0.504
Olivier T	0.008	0.504
Pascal F	0.009	0.504

Tableau 1. *Résultats obtenus par le système après la fusion des informations.*

tique décroissante), $OlivierC$ n'occupe désormais plus la première place malgré la reconnaissance de cette personne par le moteur RV. Ceci provient de la contestation du reconnaisseur de genre H/F dans la zone détectée. **FUSINDEX** a ainsi permis à la métadonnée $Muriel$ d'occuper maintenant la première place dans cette liste de métadonnées ce qui était conforme au résultat attendu.

5.2. *Exemple 2*

Le document à indexer dans ce second exemple est proposé à la figure 2. On sup-

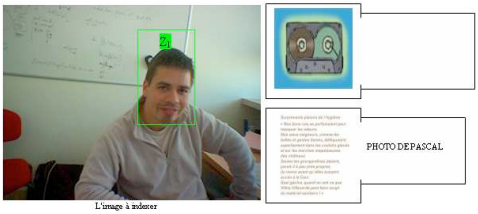

Figure 2. *Document de l'exemple 2 à indexer : image avec commentaire textuel.*

pose que le document est constitué d'une image de $PascalF$ associée à un commentaire textuel : "`photo de Pascal`". Le score obtenu par le moteur DV est $s(Z_1) = 203$ tandis que les distances obtenues par le reconnaisseur de visages RV pour cet exemple sont présentées dans le tableau 2. Le reconnaisseur de genre H/F

$d_{i,j}$	Olivier T	Olivier C	Michel	**Pascal B**
Z_1	1798729	779543	1280191	**1448802**
$d_{i,j}$	Irène	Muriel	Sonia	**Pascal F**
Z_1	2031266	1214059	1907242	**846101**

Tableau 2. *Distances obtenues par le reconnaisseur de visages RV pour l'exemple 2.*

fournit les distances suivantes : $d_{f,1} = 315641$ et $d_{h,1} = 273277$ permettant ainsi d'avoir une opinion sur le fait que le visage détecté est plutôt celui d'un homme. L'analyse du commentaire vocal RPt a donné les scores suivants : $RPt(PascalB) = 0.38$ et $RPt(PascalF) = 0.38$. Le système va tacher de représenter les incertitudes liées à ces informations.

Le tableau 3 donne la crédibilité et la probabilité pignistique à l'issu du processus de fusion $I \oplus Com$. Pour cette image, l'outil RPt a détecté le prénom $Pascal$ dans

Prénoms	$bel^{\Psi}[I, Com, P_i](Per)$	$BetP_{m^{\Psi}[I,Com,P_i]}(Per)$
Pascal F	**0.461**	**0.730**
Pascal B	**0.418**	**0.709**
Olivier C	0.148	0.574
Michel	0.057	0.528
Olivier T	0.053	0.526
Muriel	0.052	0.526
Irène	0.049	0.524
Sonia	0.049	0.524

Tableau 3. *Résultats obtenus par le système après la fusion des informations $I \oplus Com$.*

le commentaire textuel. Comme il y a deux personnes dans le carnet d'adresses qui se prénomment $Pascal$ ($PascalB$ et $PascalF$), le RPt propose ces deux personnes avec le même score. L'outil RV a reconnu $PascalB$ à la cinquième place et $PascalF$ en deuxième place. Là encore, l'étape de fusion permet de trouver la bonne personne $PascalF$ (voir tableau 3). Cet exemple montre comment la fusion entre I et Com a permis d'oter l'ambiguïté du commentaire textuel. D'autre part, le résultat donné par le système **FUSINDEX** n'a pas été perturbé par la mauvaise reconnaissance du moteur RV mais a tiré profit de l'ordre fourni par ce dernier pour trouver la bonne indexation.

6. Conclusion

Nous avons montré dans cet article comment l'on pouvait fusionner différentes informations issues de divers moteurs d'indexation pour gérer les problèmes de fiabilité de chacun d'entre eux. Une première évaluation que nous avons conduit a confirmé

l'intérêt de cette méthode. Cette évaluation a été faite avec un corpus réel de 453 images ne contenant qu'un seul visage avec leurs commentaires audio et un carnet d'adresses de 25 personnes. Avec des taux de bonne reconnaissance de 94% pour DV, de 92% pour H/F, de 63% pour RV, de 40% pour RPa, le fait de fusionner les trois outils DV, RV et H/F a amélioré le taux de bonne réponse à 77% et le fait de fusionner les quatre outils DV, RV, H/F et RPa a amélioré encore le taux de bonne réponse à 83%, une réponse étant dite bonne si la bonne personne se trouve parmi les trois premières métadonnées proposées. Ainsi le système **FUSINDEX** a apporté un gain important. Ce gain est essentiellement dû à la fusion des informations dont l'incertitude et l'imprécision ont été quantifiées par des fonctions de croyance. Cette stratégie a permis de réduire la propagation des erreurs et de lever certaines ambiguïtés causées par l'indexation de certains moteurs.

Il est également important de souligner que les stratégies et les algorithmes de fusion développés dans cet article permettent facilement d'intégrer d'autres outils d'indexation en s'inspirant des différentes méthodes utilisées. Ceci permet à notre système d'être extensible. Il suffit de trouver une méthode pour passer du cadre de discernement initial de ce nouvel outil au cadre quantifiant la pertinence. Nous envisageons aussi d'appliquer la même méthode pour la reconnaissance de personnes dans des vidéos télévisées en utilisant d'autres moteurs adaptés à ce media comme par exemple un outil de reconnaissance du locuteur ou d'inscriptions textuelles dans une image.

7. Bibliographie

Adams W., Iyengary G., Linz C., Naphade M., Neti C., Nock H., Smith J., « Semantic Indexing of Multimedia Content Using Visual, Audio and Text Cues », *Eurasip Journal on Applied Signal Processing*, vol. 2, p. 1-16, Feb, 2003.

Agosto L., Plu M., Bellec P., Vignollet L., « Someone : A Cooperative System for Personalized Information Exchange », *International Conference of Enterprise Information Systems*, Angers, France, 2003.

Aslandogan Y. A., Yu C., « Multiple Evidence Combination in Image Retrieval : Diogenes Searches for People on the Web », *Proceedings of the 23rd Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, p. 88-95, June, 2000.

Azzam I., Leung C., Horwood J., « A Fuzzy Expert System for Concept-Based Image Indexing and Retrieval », *Proceedings of the 11th International Multimedia Modelling Conference (MMM'05)*, IEEE Computer Society, p. 452-457, 2005.

Boser B., Guyon I., Vapnik V., « A Training Algorithm for Optimal Margin Classifiers », *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, ACM Press, New York, NY, USA, p. 144-152, 1992.

Canny J., « Collaborative Filtering with Privacy Via Factor Analysis », *ACM Conference on Research and Development in Information Retrieval, SIGIR'2002*, ACM, Tampere, Finland, August, 2002.

Charlet D., al, « Neologos : an optimized database for the development of new speech processing algorithms », *Proceedings of the 9th European Conference on Speech Communication and Technology INTERSPEECH'2005*, Lisbon, Portugal, p. 1549-1552, September, 2005.

Cheng Y. D., O'Toole A. J., Abdi H., « Classifying adults' and children's faces by sex : computational investigations of subcategorical feature encoding », *Cognitive science*, vol. 25, n° 5, p. 659-731, 2001.

Davis M., Smith M., Stentiford F., Bambidele A., Canny J., Good N., King S., Janakiraman R., « Using Context and Similarity for Face and Location Identification », *Proceedings of the IS& T/SPIE 18th Annual Symposium on Electronic Imaging Science and Technology Internet Imaging VII*, IS& T/SPIE Press, San Jose, California, 2006.

Dempster A., « Upper and Lower Probabilities Induced by a Multivalued Mapping », *Annals of Mathematical Statistics*, vol. AMS 38, p. 325-339, 1967.

Dempster A., Laird N., Rubin D., « Maximum Likelihood from Incomplete Data Via the EM Algorithm », *Royal Statistical Society*, vol. 39, n° 1, p. 1-38, November, 1977.

Denœux T., « A k-nearest neighbor classification rule based on Dempster-Shafer theory », *IEEE Transaction on Systems, Man and Cybernetics*, vol. 25, n° 5, p. 804-813, 1995.

Dubois D., Prade H., *On several representations of an uncertainty body of evidence*, M.M. Gupta and E. Sanchez, chapter Fuzzy Information and Decision Processes, p. 167-181, 1982.

Dubois D., Prade H., « Combinaison of Information in the Framework of Possibility Theory », *Data fusion in robotics and machine intelligence*p. 481-505, 1992.

Martin A., « Fusion D'information Haut Niveau Application À la Classification D'images Sonar », *Atelier : Fouille de Données Complexes, EGC'05*, Paris, France, Janvier, 2005.

Naaman M., Yeh R., Garcia-Molina H., Paepcke A., « Leveraging Context to Resolve Identity in Photo Albums », *Proceedings of the Fifth ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, Denver, Colorado, USA, June, 2005.

Philipp-Foliguet S., Logerot G., Constant P., Gosselin P., Lahanier C., « Multimedia Indexing and Fast Retrieval Based on a Vote System », *In IEEE International Conference on Multimedia & Expo*, Toronto, Ontario, Canada, July, 2006.

Shafer G., « A Mathematical Theory of Evidence », *Princeton University Press*, 1976.

Smets P., « Belief Induced by the Partial Knowledge of the Probabilities », *in* , D. Heckerman, , Al. (eds), *Uncertainty in Artificial Intelligence, UAI'94*, Morgan Kaufmann, San Mateo, p. 523-530, 1994.

Smits G., Plu M., Bellec P., « Personal Semantic Indexation of Images using Textual Annotations », *The First International Conference on Semantics and Digital Media Technology, SAMT2006*, Athens, Greece, 2006.

Vannoorenberghe P., « Un État de L'art sur Les Fonctions de Croyance Appliquées Au Traitement de L'information », *Revue I3*, vol. 3, n° 3(2), p. 9-45, 2003.

Visani M., Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages, Thèse, Institut National des Sciences Appliquées de Lyon, 2005.

Wu Y., Chang E., Chang K.-C., Smith J. R., « Optimal Multimodal Fusion for Multimedia Data Analysis », *Proceedings of the 12th Annual ACM International Conference on Multimedia*, ACM Press, New York, NY, USA, p. 572-579, October, 2004.