



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 15143

The contribution was presented at :
<https://www.irit.fr/EGC2013/>

To cite this version : Codreanu, Dana and Manzat, Ana-Maria and Sèdes, Florence
Vers une extraction contextuelle des métadonnées multimédias. (2013) In: 13eme
Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance
(EGC 2013), 28 January 2013 - 1 February 2013 (Toulouse, France).

Any correspondence concerning this service should be sent to the repository
administrator: staff-oatao@listes-diff.inp-toulouse.fr

Vers une extraction contextuelle des métadonnées multimédias

Dana Codreanu*, Ana-Maria Manzat*, Florence Sedes*

*Université de Toulouse – Université Paul Sabatier – IRIT UMR 5505
118 Route de Narbonne, 31062 Toulouse Cedex 9 - France
codreanu, manzat, sedes@email,
<http://www.irit.fr>

Résumé. La diversité des contenus multimédias a engendré une diversité croissante des algorithmes d'indexation. L'exécution de tous les algorithmes dont on dispose sur tous les contenus multimédias disponibles surcharge le système et extrait des informations qui peuvent ne pas être utilisées. Dans ce papier nous proposons une technique d'indexation différée, distribuée et adaptative dans un système d'information multimédia distribué. Les contenus multimédias sont analysés en deux pas : au moment de l'acquisition des contenus par des algorithmes (implicites) qui extraient des métadonnées génériques (e.g., présence de personne) et au moment de la requête par des algorithmes (explicites) qui analysent les contenus pour extraire des informations plus poussées (e.g., reconnaissance des personnes). La sélection des deux ensembles d'algorithmes (implicites et explicites) est réalisée en fonction des requêtes des utilisateurs et des performances des algorithmes dans différents contextes.

1 Introduction

La grande diversité des contenus multimédias a engendré le développement d'un grand nombre d'algorithmes d'indexation (d'extraction de métadonnées) de ces contenus (Berry et Castellanos (2008), Ma et Manjunath (1999), Pinquier et André-Obrecht (2006), Snoek et Worring (2005), Ces algorithmes analysent des caractéristiques différentes des contenus (descripteurs vidéo, audio et textuel comme la couleur, la forme des objets, la fréquence des mots etc.). Leur performances diffèrent en fonction des conditions d'acquisition et du contexte technique du système. L'exécution de tous ces algorithmes sur tous les contenus a comme conséquence la consommation de beaucoup de ressources et la génération des métadonnées qui seront peut être pas utilisées après, et donc la surcharge du système.

Dans ce papier, nous proposons une technique d'indexation différée et distribuée (Section 2) qui sélectionne les algorithmes d'indexation à exécuter en fonction des conditions d'acquisition des contenus (Section 2.1), des requêtes des utilisateurs (Section 2.2) et du contexte technique du système d'acquisition. Dans la Section 3 nous illustrons nos propositions dans le cadre d'un système de vidéo surveillance en donnant des exemples pour chacun des trois éléments pris en compte présentés dans la section antérieure. Nous finissons par des conclusions et perspectives (Section 4).

2 L'indexation différée des documents multimédias

Le choix des algorithmes d'indexation un problème essentiel dans la gestion des contenus multimédias. L'utilisation de manière systématique de tous les algorithmes d'indexation disponibles surcharge le système. Or, si le nombre d'algorithmes employés est trop restreint la performance du système peut être altérée à cause du manque des métadonnées. La stratégie d'indexation doit évoluer dans le temps en concordance avec le besoin de l'utilisateur. Les différents systèmes distribués de gestion des contenus multimédias (e.g., DISCO Boisson et al. (2008); CANDELA Pietarila et al. (2005); WebLab Giroux et al. (2008);) adoptent un ensemble figé d'algorithmes d'indexation, et ne prennent pas en compte l'évolution des besoins de l'utilisateur, ni le changement des conditions d'acquisition des contenus à indexer. Or, les algorithmes ont des performances différentes en fonction de : leur niveau de généralité, du contexte d'acquisition des contenus analysés : des conditions météo (e.g., pluie, nuages), des conditions de luminosité (e.g., nuit, jour), du lieu d'acquisition (à l'intérieur, à l'extérieur), du contexte technique du système d'acquisition : type des caméras (analogiques, numériques, couleurs/b/w), distance focale, type de scène observée par une caméra (e.g., gare, parc, route).

Brut et al. (2011) propose d'indexer les contenus en deux temps : à l'acquisition du contenu (i.e., l'indexation implicite) et à la demande, au moment de l'exécution d'une requête de l'utilisateur (i.e., l'indexation explicite). Nous proposons d'améliorer ce processus en prenant en compte des éléments du contexte d'acquisition, du contexte technique du système d'acquisition et l'historique des requêtes utilisateurs pour avoir une indexation adaptative et pertinente. Nous présentons ces trois aspects pris en compte dans l'algorithme de sélection des algorithmes d'indexation dans la section suivante.

2.1 La prise en compte des conditions d'acquisition des contenus

Si le système dispose de plusieurs algorithmes d'indexation qui extraient les mêmes informations du contenu multimédia (e.g., la présence des personnes), mais qui ont des performances différentes (e.g., un pour le jour ; un pour la nuit), pour des détecteurs de qualité, le système doit pouvoir choisir de façon continue l'algorithme d'indexation qui obtient les meilleurs résultats en fonction des conditions d'acquisition du contenu (e.g., luminosité).

Nous proposons d'utiliser des capteurs qui détectent le changement des conditions environnementales, qui ont un impact sur l'acquisition du contenu. Les informations acquises par ces capteurs sont stockées sur le même serveur que le contenu. Elles sont utilisées par le système pour sélectionner l'algorithme qui obtient les meilleures détecteurs dans les conditions spécifiées (e.g., pendant le jour un algorithme qui donne des détecteurs optimaux pendant la journée est utilisé, et un autre pendant la nuit qui donne des détecteurs optimaux la nuit). Cette prise en compte des changements de conditions d'acquisition permet d'améliorer les performances du système en termes de précision des détecteurs.

2.2 La prise en compte des requêtes des utilisateurs

Le fait de réaliser une partie de l'indexation au moment de la requête augmente le temps de réponse. Pour limiter ce temps nous proposons d'analyser l'historique des requêtes exécutées par les utilisateurs afin de déterminer quelles sont les caractéristiques multimédias les plus demandées pendant une période de temps, et utiliser les algorithmes qui les extraient pendant l'indexation implicite. Pour réaliser ce choix, nous associons un poids à chaque caractéristique multimédia, qui est incrémentée avec chaque requête qui la demande.

Nous proposons de considérer, de manière temporaire, une certaine caractéristique multimédia dans le processus d'indexation implicite (l'algorithme qui la détecte) à partir du moment où son poids atteint un certain seuil. Dans le cas où plusieurs algorithmes peuvent détecter une certaine caractéristique multimédia, le choix de l'algorithme plus adapté au contexte d'exécution est réalisé comme décrit dans la section 2.1. Au début de chaque période de temps (e.g., heure, jour, semaine), les poids de toutes les caractéristiques multimédias supplémentaires sont remis à zéro, et la liste des algorithmes implicites reinitialisée.

3 Application pour un système de vidéo surveillance

Le nombre de caméras de vidéo surveillance installées dans les villes par différentes organismes est de plus en plus grand et le contenu généré volumineux et divers.

Cette hétérogénéité du domaine de la vidéo surveillance en terme de : contextes d'acquisition, contexte technique des systèmes et de traitements nous ont déterminé de le choisir comme cas d'étude pour notre approche.

Pour exemplifier nos propositions, considérons l'exemple d'un système de vidéo surveillance d'une gare. Sur le serveur qui gère les contenus multimédias correspondants aux caméras vidéo situées dans le parking de la gare nous avons installé comme algorithmes d'indexation implicites des algorithmes qui détectent les caractéristiques suivantes : des personnes, des voitures, des plaques minéralogiques. Pour chacune de ces caractéristiques il y a installé deux algorithmes : un qui obtient des performances meilleures pendant le jour et l'autre pendant la nuit. Cette différence entre le jour et la nuit est déterminée par un capteur de luminosité. Dès que le serveur constate que la luminosité extérieure baisse en dessous d'un certain seuil il change les algorithmes d'indexation pour prendre en compte ce changement.

En prenant en compte le même système. Il est possible que pendant une semaine les requêtes des plusieurs utilisateurs concernant un "sac arraché" soient effectuées et en conséquence, sur un certain serveur distant l'algorithme qui détecte la caractéristique "sac arraché" soit exécutée de manière explicite. Par conséquent, le poids de la caractéristique multimédia "sac arraché" augmente. Lorsque son poids atteint le seuil considéré, l'algorithme qui détecte les "sac arraché" devient implicite. Au début d'une nouvelle semaine, le poids de cette caractéristique est réglé sur zéro, tandis que l'algorithme reste implicite.

Dans la même gare, supposons la même requête "sac arraché". Pour pouvoir localiser les caméras qui ont capturé les vidéos de l'agression la victime est demandée de donner des informations sur la localisation et l'heure de l'agression. Une fois l'ensemble de caméras concernées sélectionnées leur contexte technique est analysé : type de caméra, taux de compression (frame rate) etc. Ces éléments sont comparés avec les contextes d'évaluation des algorithmes disponibles qui extraient les fonctionnalités désirées ("sac arraché"). Les algorithmes qui donnent les meilleures performances dans un contexte proche de celui de notre requête sont sélectionnées.

4 Conclusion et perspectives

Ce papier présente une technique innovante d'extraction de métadonnées multimédias en deux temps : à l'acquisition du contenu (i.e., l'indexation implicite) et au moment de l'exécution d'une requête de l'utilisateur (i.e., l'indexation explicite). Pour les deux pas d'indexation les algorithmes sont sélectionnés en fonction de l'historique des requêtes utilisateur, du contexte d'acquisition des contenus et des du contexte technique du système.

Comme perspectives nous nous proposons d'améliorer la sélection des algorithmes d'indexation en prenant en compte plusieurs éléments du contexte d'acquisition et technique (les types de scènes, les codecs utilisés etc.). Dans notre approche nous considérons que les algorithmes d'indexation sont indépendants les uns des autres mais ce n'est pas toujours le cas. Parfois un algorithmes nécessite l'exécution d'un autre avant (e.g., pour réaliser une transcription d'un discours en texte on doit détecter la langue avant). Nous envisageons de définir des règles d'enchaînement d'algorithmes et d'évaluation de chaînes d'algorithmes et de les prendre en compte dans l'indexation implicite et explicite.

Références

- Berry, M. W. et M. Castellanos (Eds.) (2008). *Survey of Text Mining II : Clustering, Classification, and Retrieval*. London : Springer.
- Boisson, F., M. Crucianu, et D. Vodislav (2008). Publication framework for content-based search in heterogeneous distributed multimedia databases. Scientific rapport, CEDRIC.
- Brut, M., D. Codreanu, A.-M. Manzat, et F. Sèdes (2011). Adapting Indexation to the Content, Context and Queries Characteristics in Distributed Multimedia Systems. In *International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 118–125.
- Giroux, P., S. Brunessaux, S. Brunessaux, J. Doucy, G. Dupont, B. Grilheres, Y. Mombrun, et A. Saval (2008). Weblab : An integration infrastructure to ease the development of multimedia processing applications. In *Conf. on Software & Systems Engineering and their Applications*.
- Ma, W. Y. et B. S. Manjunath (1999). Netra : a toolbox for navigating large image databases. *Multimedia Systems* 7(3), 184–198.
- Pietarila, P., U. Westermann, S. Järvinen, J. Korva, J. Lahti, et H. Löthman (2005). CANDELA – storage, analysis, and retrieval of video content in distributed systems. In *The IEEE Conference on Multimedia and Expo*, pp. 1557–1560.
- Pinquier, J. et R. André-Obrecht (2006). Audio Indexing : Primary Components Retrieval-Robust Classification in Audio Documents. *Multimedia Tools and Applications* 30(3), 313–330.
- Snoek, C. G. M. et M. Worring (2005). Multimodal video indexing : A review of the state-of-the-art. *Multimedia Tools Appl.* 25(1), 5–35.

Summary

The diversity of multimedia contents led to a growing diversity of indexing algorithms. Executing all these algorithms on all available contents will overload the system and will produce metadata that might never be used. In this paper we present an indexing technique that is executed in two steps: at the content acquisition time using algorithms (implicit) that extract generic metadata (e.g., person detection) and at the user query time using algorithms (explicit) that analyze the content in order to extract more detailed information (e.g., person recognition). The selection of the two algorithms sets is done based on users queries and on algorithms performances in different technical and acquisition contexts.