

Organizing Personal Photo Collections Using Both Contextual Metadata and Content Analysis

Betül AYDIN

Laboratoire d'Informatique de Grenoble, équipe STEAMER
681, rue de la Passerelle, 38402 Saint Martin d'Hères, France
betul.aydin@image.fr

Abstract. With the increasing use of mobile phones, the amount of personal photo collections has increased tremendously. These photos are not kept organized because of the time consuming process of manual annotation. Efficient automatic photo annotation methods are necessary for the purpose of accurate photo retrieval. In order to annotate these photos, we can distinguish between two types of metadata: contextual metadata that are related to the moment and situation when the photo was taken, and metadata that are related to the content of the photos. Context relevant metadata can be created at shot time by using the sensors embedded in the mobile phone, or subsequently, by using Web services in order to enrich these metadata. In this way, it is possible to annotate a photo with the date, the hour, the location, but also with the weather conditions of the time of the capture (by activating a weather forecast Web service). However, even though these metadata are sufficiently rich and provide new opportunities for basic image retrieval (e.g. summer 2010, afternoon in London, Luxembourg garden in winter, photos of Anna at warm season...), they are primarily contextual annotations (they give information about what is *around* at the moment of capture) and often do not fully describe what is *in* the photo. In order to describe the content of the photo, we need a complex set of methods to further analyze the image (e.g. buildings, faces, object recognition, indoor/outdoor distinction). To identify objects in the image, low-level features (e.g. color histograms, textures) have to be transformed into high-level information (semantics). Analyzing low-level features is a popular process today, but it does not offer a sufficiently rich description of the image (this is called the “semantic-gap” problem). Our work consists of designing and developing an automated photo annotation method that combines the contextual metadata and the information extracted by the content analysis. Our goal is not only to build a set of annotations from simply combining all the contextual metadata with all the content information, but also to build incrementally and interactively these two sets. Thus, we seek to guide the extraction of content information by using contextual information and to extend, if necessary, the contextual information to assist the process of content analysis when encountering ambiguity.

Keywords: photo annotation, mobile devices, contextual metadata, content awareness, semantic Web

Organisation de collections de photos personnelles à l'aide de métadonnées issues à la fois du contexte de prise et de l'analyse du contenu

Betül AYDIN

Laboratoire d'Informatique de Grenoble, équipe STEAMER
681, rue de la Passerelle, 38402 Saint Martin d'Hères, France
betul.aydin@image.fr

Résumé. Avec l'utilisation croissante des téléphones mobiles, le nombre de collections de photos personnelles a augmenté de façon remarquable. Ces photos sont conservées le plus souvent sans être organisées, car ce processus, lorsqu'il est réalisé de manière manuelle, est une activité chronophage. L'indexation de ces photos est un besoin crucial pour rendre la recherche d'image par le contenu (CBIR) efficace. Afin d'annoter ces photos, on peut distinguer deux types de métadonnées : les métadonnées contextuelles, relatives à l'environnement lors de la prise, et les métadonnées relatives au contenu des photos. La création de métadonnées contextuelles pertinentes peut être réalisée, en situation, à l'aide de capteurs mobiles, ou ultérieurement, en activant des services Web afin

d'enrichir les métadonnées brutes captées. De cette manière, il est par exemple possible d'annoter une photo avec la date, l'heure, et le lieu (ce sont des informations captées), mais aussi avec les conditions météorologiques du moment de la prise (informations obtenues dans ce cas par l'activation d'un service web météorologique). Cependant, même si ces métadonnées sont assez riches et offrent des possibilités de requêtes originales et intéressantes (été 2010, après-midi à Londres, jardin de Luxembourg en hiver, photos en présence de Anna par temps chaud, etc.), elles sont avant tout contextuelles (elles renseignent sur ce qu'il se passe *autour* de la photo) et ne décrivent souvent pas entièrement ce qui est *dans* la photo. S'agissant du contenu des photos, il existe un ensemble de méthodes souvent complexes pour mener une analyse approfondie de l'image (reconnaissance des bâtiments, des visages, des objets, distinction de l'intérieur/extérieur, etc.). Pour identifier des objets sur une image, il faut transformer l'information de bas niveau (histogrammes de couleurs, textures, formes, etc.) en une information de haut niveau (sémantique). Analyser des images de bas niveau est maintenant devenu un processus bien maîtrisé, mais qui n'offre pas une description suffisamment riche de l'image (c'est ce qu'on appelle le fossé sémantique ou *semantic gap* en anglais). Notre travail consiste à concevoir et développer une méthode d'annotation automatique des photos qui combine des métadonnées contextuelles et des informations extraites par une analyse du contenu. Notre objectif n'est pas seulement de construire un ensemble d'annotations à partir de la simple union de l'ensemble des métadonnées contextuelles, et de l'ensemble des informations de contenu, mais bien de construire de manière incrémentale et interactive ces deux ensembles. Ainsi, on cherchera à la fois à orienter l'extraction d'informations de contenu à partir des informations contextuelles, et à étendre, au besoin, l'ensemble des informations contextuelles afin d'aider le processus d'analyse du contenu lors d'une indétermination.

Mots-clés: annotation de photos, les appareils mobiles, métadonnée contextuelle, contenu visuel des images, web sémantique