

Thesis: Search-based Automatic Image Annotation Using Geotagged Community Photos

Author: Hatem Mousselly Sergieh

Abstract

In the Web 2.0 era, platforms for sharing and collaboratively annotating images with keywords, called tags, became very popular. Tags are a powerful means for organizing and retrieving photos. However, manual tagging is time consuming. Recently, the sheer amount of user-tagged photos available on the Web encouraged researchers to explore new techniques for automatic image annotation. The idea is to annotate an unlabeled image by propagating the labels of community photos that are visually similar to it.

Most recently, an ever increasing amount of community photos is also associated with location information, i.e., geotagged. In this thesis, we aim at exploiting the location context and propose an approach for automatically annotating geotagged photos. Our objective is to address the main limitations of state-of-the-art approaches in terms of the quality of the produced tags and the speed of the complete annotation process. In order to achieve these goals, we, first, deal with the problem of collecting images with the associated metadata from online repositories. Accordingly, we introduce a strategy for data crawling that takes advantage of location information and the social relationships among the contributors of the photos. To improve the quality of the collected user-tags, we present a method for resolving their ambiguity based on tag relatedness information. In this respect, we propose an approach for representing tags as probability distributions based on the algorithm of Laplacian score feature selection. Furthermore, we propose a new metric for calculating the distance between tag probability distributions by extending Jensen-Shannon Divergence to account for statistical fluctuations. In order to efficiently identify the visual neighbors, the thesis introduces two extensions to the state-of-the-art image matching algorithm, known as Speeded Up Robust Features (SURF). To speed up the matching, we present a solution for reducing the number of compared SURF descriptors based on classification techniques, while the accuracy of SURF is improved through an efficient method for iterative image matching. Furthermore, we propose a statistical model for ranking the mined annotations according to their relevance to the target image. This is achieved by combining multi-modal information in a statistical framework based on Bayes' rule.

Finally, the effectiveness of each of mentioned contributions as well as the complete automatic annotation process are evaluated experimentally.

Keywords: Image Annotation, SURF, Tagging, Geotagging, Flickr, Folksonomy.

Résumé

La technologie Web 2.0 a donné lieu à un large éventail de plates-formes de partage de photos. Il est désormais possible d'annoter des images de manière collaborative, au moyen de mots-clés; ce qui permet une gestion et une recherche efficace de ces images.

Toutefois, l'annotation manuelle est laborieuse et chronophage. Au cours des dernières années, le nombre grandissant de photos annotées accessibles sur le Web a permis d'expérimenter de nouvelles méthodes d'annotation automatique d'images. L'idée est d'identifier, dans le cas d'une photo non annotée, un ensemble d'images visuellement similaires et, a fortiori, leurs mots-clés, fournis par la communauté. Il existe actuellement un nombre considérable de photos associées à des informations de localisation, c'est-à-dire géo-localisées. Nous exploiterons, dans le cadre de cette thèse, ces informations et proposerons une nouvelle approche pour l'annotation automatique d'images géo-localisées. Notre objectif est de répondre aux principales limites des approches de l'état de l'art, particulièrement concernant la qualité des annotations produites ainsi que la rapidité du processus d'annotation.

Tout d'abord, nous présenterons une méthode de collecte de données annotées à partir du Web, en se basant sur la localisation des photos et les liens sociaux entre leurs auteurs. Par la suite, nous proposerons une nouvelle approche afin de résoudre l'ambiguïté propre aux tags d'utilisateurs, le tout afin d'assurer la qualité des annotations. L'approche démontre l'efficacité de l'algorithme de recherche de caractéristiques discriminantes, dit de Laplace, dans le but d'améliorer la représentation de l'annotation. En outre, une nouvelle mesure de distance entre mots-clés sera présentée, qui étend la divergence de Jensen-Shannon en tenant compte des fluctuations statistiques. Dans le but d'identifier efficacement les images visuellement proches, la thèse étend sur deux points l'algorithme d'état de l'art en comparaison d'images, appelé SURF (Speeded-Up Robust Features). Premièrement, nous présenterons une solution pour filtrer les points-clés SURF les plus significatifs, au moyen de techniques de classification, ce qui accélère l'exécution de l'algorithme. Deuxièmement, la précision du SURF sera améliorée, grâce à une comparaison itérative des images. Nous proposerons un modèle statistique pour classer les annotations récupérées selon leur pertinence du point de vue de l'image-cible. Ce modèle combine différents critères, il est centré sur la règle de Bayes. Enfin, l'efficacité de l'approche d'annotation ainsi que celle des contributions individuelles sera démontrée expérimentalement.

Mots-clés: Annotation d'images, SURF, Tagging, Code Géographique, Flickr, Folksonomie

Zusammenfassung

Seit der Einführung von Web 2.0 steigt die Popularität von Plattformen, auf denen Bilder geteilt und durch die Gemeinschaft mit Schlagwörtern, sogenannten Tags, annotiert werden. Mit Tags lassen sich Fotos leichter organisieren und auffinden. Manuelles Taggen ist allerdings sehr zeitintensiv. Animierte von der schier unermesslichen Menge an im Web zugänglichen, von Usern getaggen Fotos, erforschen Wissenschaftler derzeit neue Techniken der automatischen Bildannotation. Dahinter steht die Idee, ein noch nicht beschriftetes Bild auf der Grundlage visuell ähnlicher, bereits beschrifteter Community-Fotos zu annotieren. Unlängst wurde eine immer größere Menge an Community-Fotos mit geographischen Koordinaten versehen (geotagged). Die Arbeit macht sich diesen geographischen Kontext zunutze und präsentiert einen Ansatz zur automatischen Annotation geogetaggtter Fotos. Ziel ist es, die wesentlichen Grenzen der bisher bekannten Ansätze in Hinblick auf die Qualität der produzierten Tags und die Geschwindigkeit des gesamten Annotationsprozesses aufzuzeigen.

Um dieses Ziel zu erreichen, wurden zunächst Bilder mit entsprechenden Metadaten aus den Online-Quellen gesammelt. Darauf basierend, wird eine Strategie zur Datensammlung eingeführt, die sich sowohl der geographischen Informationen als auch der sozialen Verbindungen zwischen denjenigen, die die Fotos zur Verfügung stellen, bedient. Um die Qualität der gesammelten User-Tags zu verbessern, wird eine Methode zur Auflösung ihrer Ambiguität vorgestellt, die auf der Information der Tag-Ähnlichkeiten basiert. In diesem Zusammenhang wird ein Ansatz zur Darstellung von Tags als Wahrscheinlichkeitsverteilungen vorgeschlagen, der auf den Algorithmus der sogenannten Laplacian Score (LS) aufbaut. Des Weiteren wird eine Erweiterung der Jensen-Shannon-Divergence (JSD) vorgestellt, die statistische Fluktuationen berücksichtigt. Zur effizienten Identifikation der visuellen Nachbarn werden in der Arbeit zwei Erweiterungen des Speeded Up Robust Features (SURF)-Algorithmus vorgestellt. Zur Beschleunigung des Abgleichs wird eine Lösung auf der Basis von Klassifikationstechniken präsentiert, die die Anzahl der miteinander verglichenen SURF Deskriptoren minimiert, während die SURF-Genauigkeit durch eine effiziente Methode des schrittweisen Bildabgleichs verbessert wird. Des Weiteren wird ein statistisches Modell basierend auf der Bayes'schen Regel vorgeschlagen, um die erlangten Annotationen entsprechend ihrer Relevanz in Bezug auf das Zielbild zu ranken. Schließlich wird die Effizienz jedes einzelnen, erwähnten Beitrags experimentell evaluiert. Darüber hinaus wird die Performanz des vorgeschlagenen automatischen Annotationsansatzes durch umfassende experimentelle Studien als Ganzes demonstriert.

Schlagwörter: Bildannotation, SURF, Tagging, Geokodierung, Flickr, Folksonomie.