

Automatische Geodaten-Annotation von Fotos

Christoph Pohl

Institut für Software- und Multimediatechnologie, TU Dresden
s2804966@mail.zih.tu-dresden.de

Zusammenfassung Metadaten, insbesondere *GeoTags* (geographische Koordinaten), gewinnen vor allem bei der Archivierung und Sortierung großer Fotoarchive immer mehr an Bedeutung. In diesem Zusammenhang stellt der vorliegende Artikel neuartige Technologien zur automatischen Erkennung von geographischen Koordinaten überblicksweise vor. Dabei wird zunächst die Verwendung von User-Tags beschrieben und es werden Herausforderungen, welche bei der Bestimmung der Geokoordinaten mit Hilfe dieser Tags auftreten, analysiert. Im Anschluss daran werden weitere Methoden vorgestellt, welche auch *nachträglich* den Aufnahmeort bestimmen können. Jede der vorgestellten Technologien wird dabei zunächst beschrieben sowie anschließend auf mögliche Defizite beziehungsweise auftretende Herausforderungen untersucht. Abschließend werden die gewonnenen Erkenntnisse zusammengefasst und zukünftige Fragen dieses Forschungsgebietes diskutiert.

Keywords: Geotagging, Georeferencing, Placing, User-Tags, Bildannotationen

1 Einleitung

Eine häufig genutzte Funktion von Tablet-PCs und vor allem Smartphones ist das Aufnehmen von Fotos und Videos. Aufgrund der Tatsache, dass sich mehrere 1000 Bilder mühelos auf solch einem ultra-mobilen Endgerät abspeichern lassen, werden jeden Tag unzählige Aufnahmen von Millionen von Menschen auf der ganzen Welt gemacht. Des Weiteren sind Smartphones oder normale Mobiltelefone zum alltäglichen Begleiter vieler Menschen geworden und somit in jeder Situation griffbereit. Auch soziale Netzwerke wie Facebook oder Twitter, welche den Nutzern helfen ihre Aufnahmen ohne großen Aufwand weit zu verbreiten, sind maßgeblich an der stetig wachsenden Menge von digitalen Aufnahmen beteiligt.

Seit der Veröffentlichung der ersten Digitalkamera mit integriertem GPS¹-Empfänger im Jahr 2008 [4] können Bild- und Videoaufnahmen mit exakten geographischen Koordinaten des Aufnahmeortes annotiert werden. Das dazu verwendete Metadaten- bzw. Dateihheaderformat ist *Exif*² und kann neben dem Längen- und Breitengrad auch weitere Metadaten wie bspw. die Brennweite, den

¹ Global Position System

² Exchangeable Image File Format

Belichtungszustand oder den ISO-Wert einer Aufnahme im Header der JPEG- oder TIFF-Bilddatei abspeichern. Heutzutage ist neben den Digitalkameras jedes Smartphone in der Lage Geodaten zu sammeln, um diese nach Wunsch an das jeweilige Bild anzuhängen. Diese Positionsbestimmung funktioniert mit Hilfe des bereits erwähnten GPS-Systems, Mobilfunktürmen oder anhand von WLAN-Netzwerken, welche sich in der Nähe des Anwenders befinden.

Die steigende Zahl der Anwender von Foto-Archivierungsprogrammen (z. B. *iPhoto* von Apple oder *Picasa* von Google) und Webseiten wie *Flickr*, welche es ermöglichen Bilder mit Geotags zu versehen und ortsbezogene Suchanfragen zu stellen, lässt darauf schließen, dass sich die Nutzer über den Mehrwert dieser Geodaten bewusst werden [9]. Die Lokalisierung des Aufnahmeortes von Bild- und Videodaten ermöglicht neue Such- bzw. Archivierungsstrategien, welche zum Beispiel in der Tourismusbranche eingesetzt werden könnten, um den Kunden aktuelle, ortsbezogene Aufnahmen präsentieren zu können.

Trotz dieser Erkenntnis ist der Anteil „getaggtter“ Fotos oder Videos im Verhältnis zu allen hochgeladenen Aufnahmen sehr gering. Der Ausgleich dieses Defizits im Fotobereich steht im Fokus der vorliegenden Arbeit. Dazu werden im weiteren Verlauf verschiedene Technologien und Systeme zur (semi-)automatischen Geodaten-Annotation vorgestellt, welche ohne die Zuhilfenahme von GPS-Koordinaten den Aufnahmeort eines Bildes ermitteln können. Der nachfolgende Abschnitt 2 untersucht zunächst konventionelle Informationsquellen wie *User-Tags* und geht dabei auf deren Herausforderungen für Computersysteme ein. Im Anschluss daran beschreibt Abschnitt 3 *neuartige* Möglichkeiten zur Bestimmung geographischer Metadaten. Danach werden Algorithmen präsentiert, welche anhand visueller Ähnlichkeiten zweier Bilder (Abschnitt 3.1) oder anhand räumlich-zeitlicher Zusammenhänge von Bildsequenzen (Abschnitt 3.2) den Aufnahmeort bestimmen können. Der darauffolgende Abschnitt 3.3 erläutert, wie mit Hilfe von sozialen Netzwerken die gewünschten Geo-Koordinaten ermittelt werden können. Abschließend werden die gewonnenen Erkenntnisse in Abschnitt 4 noch einmal zusammengefasst und ein Überblick über zukünftige Forschungsschwerpunkte gegeben.

2 User-Tags als Grundlage für die Geodaten-Annotation

Durch die Ablösung des klassischen Fotoapparates durch die Digitalkamera wurde es den Anwendern ermöglicht, zu sehr geringen Preisen eine Vielzahl von Bildern aufzunehmen und abzuspeichern. Denn aufgrund der Tatsache, dass man nur eine Auswahl der geschossenen Bilder „entwickeln“ lassen kann, werden automatisch guten Gewissens mehrere Aufnahmen einer Szene gemacht. Oftmals werden dann die favorisierten Bilder ausgedruckt, in ein Fotoalbum geklebt und *beschriftet*. Dies ist im Grunde die analoge Form des „*Taggings*“. Mit der Einführung des Web 2.0 verlagerten sich die analogen Fotoalben ins Internet und wurden so digital. Mit der immer größer werdenden Datenmenge effizient umzugehen wurde zunehmend schwerer, da oft nur einfache Ordnerstrukturen zum Archivieren verwendet wurden. Mit der Einführung der in Abschnitt 1 erwähnten

Fotoprogramme *iPhoto* und *Picasa* sowie Online-Plattformen wie *Flickr*, wurde es den Nutzern ermöglicht, weitere, für sie relevante Informationen in einem Bild zu annotieren.

Diese *Tags* bestehen meist aus einer Liste von Stichworten, welche die gezeigte Aufnahme beschreiben [10]. Im Vergleich zu kategorie- und ontologiebasierten Systemen, resultieren diese Tags in unstrukturiertem Wissen ohne einer vordefinierten Semantik. Trotzdem, oder gerade deswegen, liegt darin der größte Nutzen von User-Tags, denn sie bieten zwar einen großen Interpretationsspielraum, beschreiben in den meisten Fällen aber dennoch das Wesentliche. Aufgrund der Tatsache, dass es gegenwärtig unmöglich für Computersysteme ist ein komplettes, semantisches Verständnis der Tags zu entwickeln, fokussiert sich die Forschung, im Bereich Geodaten-Annotation, auf die Extraktion strukturierter Wissens aus den Annotationen der Anwender [10]. Um dies zu ermöglichen werden unter anderem die folgenden Methoden angewandt, welche im nächsten Abschnitt näher beschrieben werden:

1. Qualitätsprüfung der User-Tags
2. Unterscheidung von Event- und Place-Tags
3. Toponym Resolution

Die **Qualität der vorhandenen User-Tags** kann anhand vieler Kriterien unterschiedlichster Art gemessen werden. Aufgrund der Tatsache, dass Tags in einfache Textboxen, welche in den meisten Fällen keine Restriktionen vorschreiben, eingetragen werden, können die Anwender vollkommen willkürliche Wörter bzw. „Wortkreationen“ verwenden. Demzufolge können keine qualitativ sehr hochwertigen Beiträge erwartet werden, da Schreibfehler oder selbst erdachte Wortkreationen und Abkürzungen in den Tags enthalten sein könnten. Des Weiteren muss ermittelt werden, wie hoch der Detailgrad der Beschreibung ist und es muss festgestellt werden, ob die Tags überhaupt *inhaltlich korrekt* sind. In diesem Zusammenhang werden in [14] Lösungsansätze dieser eben beschriebenen Herausforderungen detailliert vorgestellt.

Ein weiterer Ansatz strukturierte (insbesondere geographische) Information aus unstrukturierten User-Tags zu extrahieren ist die explizite Unterscheidung von **ereignis- und ortsbezogenen Tags** [10]. Nach der Definition von Rattenbury et al. weisen Event- bzw. ereignisbezogene User-Tags signifikante *zeitliche* Muster auf, wohingegen Place- bzw. ortsbezogene Tags signifikante *räumliche* Muster offenbaren [10]. Wenn ein Anwender eine Aufnahme beispielsweise mit „*New York Marathon*“ annotiert, dann ist davon auszugehen, dass diese Aufnahme in New York City, in der Zeit um November gemacht wurde. Im Gegensatz dazu ist bei dem Tag „*Karneval*“ nicht ersichtlich an welchem Ort die Aufnahme erstellt wurde. Die *Zeitspanne* des Aufnahmezeitraums lässt sich bei diesem Beispiel hingegen schon eingrenzen.

Ein letzter Ansatz, mit dessen Hilfe insbesondere geographische Information gewonnen werden kann, ist die sogenannte **Toponym Resolution** [11]. Das Ziel dieses Ansatzes ist es, Ortsnamen, welche in einem User-Tag gefunden wurden, Geo-Koordinaten zuzuordnen. Dabei muss vor allem die *Mehrdeutigkeit* der

gefundenen Toponyme in Bezug auf deren Referenten aufgelöst werden. Ein typisches Beispiel ist „Aesch“ in der Schweiz. Dieses Toponym hat sowohl im Kanton Zürich, Baselland als auch Luzern einen Referenten. Dieses Problem wird als *Geo-Geo-Ambiguität* bezeichnet und lässt sich mit Hilfe verschiedener Methoden, welche Tobias Brunner in seiner Masterarbeit [2] ausführlich beschreibt, auflösen. An dieser Stelle sollen lediglich zwei Vorgehensweisen beschrieben werden, um einen Einblick in die Problematik zu geben:

Auflösung durch beschriebenen Kontext: In vielen Fällen sind Toponyme von beschreibenden Wörtern umgeben, welche sie eindeutig auflösbar machen. Beispielsweise kann der Referent des Toponyms „Zürich“ sowohl der Kanton als auch die Stadt sein. Sollte der Nutzer allerdings „Kanton Zürich“ geschrieben haben, so ist das Toponym „Zürich“ eindeutig dem Referenten „Kanton“ zuzuweisen.

Verwendung von „Default Referents“: Bei dieser Vorgehensweise wird aus einer Vielzahl möglicher Referenten derjenige bestimmt, welcher mit höchster Wahrscheinlichkeit der richtige (Ort) ist. Dieser Wahrscheinlichkeitswert wird oftmals anhand der Population bestimmt. So gibt es z. B. 1867 Referenten³ für „San Francisco“ im GNS-Gazatteer⁴ und nur äußerst selten ist ein anderer Referent gemeint, als die Millionenstadt San Francisco im US-Bundesstaat Kalifornien.

Zusammenfassend kann festgestellt werden, dass **User-Tags** qualitativ sehr abhängig von den jeweiligen Bildautoren sind, da diese sowohl für die Korrektheit der Tags verantwortlich sind, als auch für den Detailgrad der Bildannotationen.

Aufgrund der Tatsache, dass es für viele Anwender anstrengend ist, ihre (alten) Bildaufnahmen selbst zu taggen, werden neuartige Technologien und Ansätze entwickelt, um Bilder (semi-)automatisch mit Geodaten zu versehen. In diesem Zusammenhang werden in den nachfolgenden Abschnitten Ansätze beschrieben, welche anhand visueller Ähnlichkeiten zweier Bilder oder aufgrund räumlich-zeitlicher Zusammenhänge von Aufnahmesequenzen die geographischen Koordinaten des Aufnahmeortes auch lange nach dem Aufnahmezeitpunkt bestimmen können.

3 Möglichkeiten zur automatischen Erkennung geographischer Metadaten

Am Ende des letzten Abschnittes wurde festgestellt, dass User-Tags zur Erfassung von geographischen Koordinaten oft unzureichend sind und für den Fall, dass gar keine Tags hinterlegt sind, wäre die Lokalisierung des Aufnahmeortes - auf diese eine Technologie beschränkt - sogar unmöglich. Aus diesem Grund wird an einer Vielzahl weiterer Ansätze zur automatischen Erfassung geographischer Metadaten geforscht. Die vorliegende Arbeit befasst sich in diesem Zusammenhang mit den folgenden aktuellen Methoden:

³ Stand: 04.12.2012

⁴ GeoNames-Ortsverzeichnis <http://geonames.nga.mil>

1. Auswertung der visuellen Ähnlichkeiten zweier Bilder
2. Analyse der räumlich-zeitlichen Zusammenhänge von Aufnahmesequenzen
3. Beobachtung der Aktivitäten des Bildurhebers in sozialen Netzwerken

In den folgenden Abschnitten 3.1, 3.2 und 3.3 werden diese Vorgehensweisen nacheinander überblicksweise beschrieben.

3.1 Visuelle Ähnlichkeiten zweier Bilder

In [12] erläutern Sergieh et al. ihren entwickelten Ansatz zur automatischen Bildannotation, welcher auf einer Kombination von High-Level-Informationen (User-Tags) und Low-Level-Informationen (visuelle Ähnlichkeiten zweier Bilder, wobei eines der Bilder bereits geographische Metadaten beinhaltet) basiert. Dieser Abschnitt fokussiert sich auf den Vergleich zweier Bildaufnahmen. Bei einer Übereinstimmung zweier Bilder werden die fehlenden Geodaten von dem gefundenen „Duplikat“ übernommen. Demzufolge werden verschiedene Low-Level-Eigenschaften, wie Farbhistogramme oder Texturen, identifiziert, kombiniert sowie miteinander verglichen. Voraussetzung für diesen Ansatz ist das Vorhandensein einer Datenbank mit Bildern, welche bereits geographische Metadaten enthalten.



Abbildung 1. Links: Referenzbild - rechts: Testbild [1]

Bilder, welche zwar am selben Ort, aber von zwei unterschiedlichen Fotografen aufgenommen wurden, unterscheiden sich in verschiedenen Eigenschaften wie Kameraperspektive, Zoomfaktor oder vorherrschende Lichtverhältnisse (siehe Abbildung 1). Deshalb ist es notwendig, diese Aufnahmen basierend auf Low-Level-Charakteristiken, welche robust gegen diese eben genannten Eigenschaften sind, miteinander zu vergleichen. Der gesamte Prozess kann dabei in drei Schritte unterteilt werden, welche nachfolgend überblicksweise beschrieben werden [12]:

1. Im ersten Schritt werden sogenannte „**Interest Points**“ - markante Ecken - innerhalb eines Bildes ermittelt (siehe Abbildung 2).

2. Anschließend wird jeder *Interest Point* durch einen „**Descriptor**“ repräsentiert, welcher als ein unterscheidbarer Eigenschaftsvektor angesehen werden kann.
3. Der letzte Schritt bestimmt schließlich die Übereinstimmungen zweier *Deskriptoren* auf Grundlage des euklidischen Abstandes. Demzufolge ist die visuelle Ähnlichkeit zweier Bilder abhängig von der Anzahl der *Interest Points*, die sie sich teilen.



Abbildung 2. ermittelte *Interest Points* einer Bildaufnahme [13]

Derzeitig existiert eine Vielzahl von Algorithmen zum Vergleich von Bildern, darunter *SIFT*⁵[7], *PCA*⁶-*SIFT*[5] und *SURF*⁷[1]. Sergej et al. haben sich in ihrer Arbeit, aufgrund seiner überragenden Geschwindigkeit verglichen mit den anderen, für den letztgenannten Algorithmus entschieden. Des Weiteren haben sie festgestellt, dass zwei Bilder als „ähnlich“ identifiziert werden können, sobald eine Übereinstimmung von vier *Interest Points* besteht [12].

⁵ Scale-Invariant Feature Transform

⁶ Principal Component Analysis

⁷ Speeded Up Robust Features

3.2 Räumlich-zeitliche Zusammenhänge von Aufnahmesequenzen

Der in diesem Absatz vorgestellte Ansatz erweitert die Technologie des vorherigen Abschnittes um den Zeitpunkt der Aufnahme eines jeden Bildes und bestimmt auf diesem Weg die geographische Lage des jeweiligen Aufnahmeortes. Aufgrund der Nutzung *zeitlicher* Bedingungen, sind markante Orientierungspunkte innerhalb eines Bildes für die Lokalisierung des Aufnahmeortes nicht zwingend notwendig [6]. Für ein besseres Verständnis soll an dieser Stelle ein Beispiel erläutert werden:

Der Aufnahmeort eines Bild, auf welchem ausschließlich das Meer zu sehen ist, kann unmöglich bestimmt werden. Es kann lediglich festgestellt werden, dass das Bild auf oder an einem Ozean der Erde gemacht wurde. Sollten allerdings weitere Bilder zur Verfügung stehen, welche zwei Stunden später aufgenommen wurden und auf denen die Akropolis zu sehen ist, dann kann man daraus schlussfolgern, dass sich der Aufnahmeort des ersten Bildes innerhalb eines Umkreises von maximal zwei Stunden Fahrzeit von Athen befinden muss. Werden an die Bildsequenz weitere Aufnahmen angehängt, so können, mit Hilfe jedes weiteren Bildes, die möglichen Aufnahmeorte aller anderen Bilder immer weiter eingegrenzt bzw. konkretisiert werden.



Abbildung 3. In dieser Abbildung aus [6] sind häufige Ausflüge, basierend auf aufeinanderfolgenden Bildaufnahmen eines Nutzers, welche mindestens fünf Tage auseinander liegen, dargestellt. Die Informationen wurden dabei aus der Flickr-Datenbank entnommen.

Kalogerakis et al. lassen in das Modell zur Erfassung von räumlich-zeitlichen Zusammenhängen, welches in [6] ausführlich vorgestellt wird, zwei Kernfaktoren einfließen:

1. Ähnlichkeit zweier Bildaufnahmen (vgl. Abschnitt 3.1)
2. Oft besuchte Reiseziele der letzten Jahre (siehe Abbildung 3)

Aufgrund der Tatsache, dass die erste Kernkomponente mit dem im vorangegangenen Abschnitt beschriebenen Ansatz vergleichbar ist, liegt der Fokus dieses Abschnittes auf dem zweiten Punkt.

Der vorgestellte Algorithmus aus [6] benötigt eine Sequenz von Bildern sowie die verstrichene Zeit zwischen den Aufnahmen als Eingabeparameter, um den Aufnahmeort jedes Bildes zu lokalisieren. Dabei ist es nicht wichtig, ob die Uhr der Kamera korrekt eingestellt ist oder nicht. Des Weiteren wird die Weltkugel in 3186 Quadrate mit jeweils 400km x 400km aufgeteilt. Das erarbeitete Modell ist dabei eine Variante des *Hidden Markov Models* (siehe Abbildung 4) und beschreibt die mögliche Position L_k eines Bildes I_k , wobei der Aufnahmeort der vorherigen Aufnahme L_{k-1} sowie das Zeitintervall ΔT_k zwischen diesen beiden Bildern bekannt sein muss.

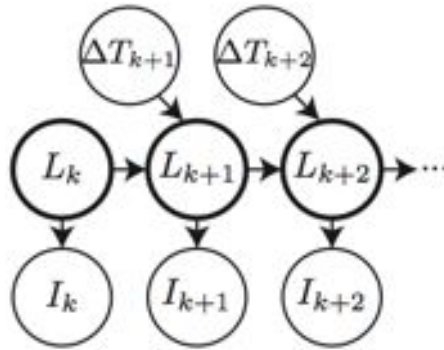


Abbildung 4. Darstellung des erarbeiteten Modells auf Basis des *Hidden Markov Models* aus [6]; I_k ...aufgenommenes Foto; L_k ...mögliche Position des Aufnahmeortes; ΔT_k ...das jeweilige Zeitintervall von zwei aufeinanderfolgenden Bildern

Die Basis dieses Ansatzes bildet eine Flickr-Bilddatenbank, welche 6 Millionen geographisch-referenzierte Fotoaufnahmen beinhaltet, sowie ein Algorithmus, welcher wie folgt arbeitet:

1. Bestimmung des ersten Aufnahmeortes
 - (a) Bestimmung einer Aufnahmesequenz x (mehrere Fotos, welche an einem Ort erstellt wurden) unter Zuhilfenahme des vorhandenen Zeitintervalls
 - (b) Visueller Vergleich der aufgenommenen Fotos mit den Bildaufnahmen der gesamten Flickr-Datenbank
2. Bestimmung der nächsten Aufnahmesequenz
3. Errechnung möglicher Orte mit Hilfe des Zeitintervalls von Sequenz x und $x + 1$
4. Vergleich der aufgenommenen Fotos mit den Bildern der eingeschränkten Flickr-Datenbank (ausschließlich Orte, welche in Punkt 3 ermittelt wurden)
5. Gehe zu Punkt 2

Die Wahrscheinlichkeit, den richtigen Aufnahmeort der Fotos zu bestimmen, steigt mit der Anzahl der Bilder innerhalb einer Aufnahmesequenz. Kalogerakis

et al. konnten bei einem Vergleich mit dem SIG⁸-Algorithmus sehr gute Ergebnisse erzielen. So bestimmte deren Algorithmus *SEQ*⁹ den Aufnahmeort der Fotos von drei Anwendern mit einer Genauigkeit von mehr als 95%, wohingegen *SIG* lediglich eine Genauigkeit von 29% erreichte [6].

Nachdem dieser Abschnitt ein Verfahren vorgestellt hat, bei welchem sowohl die Ähnlichkeit zweier Bilder, als auch die Reiseroute der Fotografen genutzt wird, um die geographischen Koordinaten der aufgenommenen Fotos zu bestimmen, erläutert der nachfolgende Abschnitt, wie klimatische Bedingungen, Einwohnerzahlen und Aktivitäten in sozialen Netzwerken die Bestimmung geographischer Koordinaten beeinflussen.

3.3 Klimatische Bedingungen, Einwohnerzahlen und Aktivitäten in sozialen Netzwerken zur Bestimmung geographischer Koordinaten

Webseiten, Blogs, Wikis und soziale Netzwerke enthalten Informationen, welche sehr vielfältig und facettenreich in Bezug auf deren Inhalt sowie deren Art und Weise sind. Dies resultiert aus der Tatsache, dass die Medien von sehr vielen Anwendern genutzt werden und die Informationen dabei in vielen Fällen von Laien bereitgestellt werden. Dienste wie Twitter ermöglichen eine automatisierte Geo-Tagging-Funktionalität, welche einer Nachricht den aktuellen Standpunkt des Nutzers beifügt. Dadurch wird es der Community ermöglicht, bezogen auf Raum und Zeit, aktuelle Geschehnisse zu identifizieren, zu beurteilen und darüber zu sprechen [8]. Die Informationen stammen aus öffentlichen Quellen, welche von unzähligen (anonymen) Menschen konstruiert und moderiert werden. Demzufolge ist es oft schwierig, die Wahrheit von der Unwahrheit zu unterscheiden. Des Weiteren ist die Objektivität der Informationen fraglich, da diese durch politische, regionale oder persönliche Ansichtsweisen verzerrt sein können.

In [3] versuchen Hauff et al. mit Hilfe der Microblogging-Plattform Twitter die Genauigkeit des bestimmten Aufnahmeortes zu erhöhen, indem sie die Kommentare des jeweiligen Nutzers, welche im selben Zeitraum „getwittert“ wurden, untersuchen. Aufgrund der Tatsache, dass Twitter-Nachrichten immer mit einem *Zeitstempel* versehen sind, können Kommentare wie „*Mein erster Urlaubstag in Miami!*“ dabei helfen, den exakten Aufnahmeort zu bestimmen. Darüber hinaus sind auch *Straßennamen*, *besondere Geschäfte*, *Sehenswürdigkeiten* oder erwähnte *Baudenkmäler* bei der Bestimmung geographischer Koordinaten hilfreich. Zusätzlich zu den Twitter-Nachrichten fließen zwei weitere Faktoren in den Prozess zur Bestimmung geographischer Koordinaten ein:

1. Klimatische Gegebenheiten
2. Einwohnerzahlen

Die untersuchte Technologie arbeitet unter der Annahme, dass der Großteil aller Menschen in Gebiete reist, in denen ein *mildes Klima* vorherrscht und keine

⁸ Single Image Geolocation

⁹ Image Sequence Geolocation

extremen Wetterbedingungen vorzufinden sind [3]. In diesem Zusammenhang wurden eine Reihe von geographisch referenzierten Fotos gesammelt und um die Durchschnittstemperatur des Aufnahmemonats erweitert.

Ein weiterer Aspekt, der eine entscheidende Rolle in dem Algorithmus von Hauff et al. spielt, ist die Einwohnerzahl der Orte. Die Autoren vermuten, ähnlich wie beim Faktor Klima, dass der Großteil der Bildaufnahmen an Orten erstellt wird, welche eine *hohe Einwohnerzahl* vorweisen. Demzufolge wird erwartet, dass mehr Bilder in New York City geschossen werden, als zum Beispiel in der Großen Victoria-Wüste Australiens. Nach diesen einführenden Fakten, wird im Folgenden der Ansatz aus [3] von Hauff et al. schrittweise beschrieben:

1. Zunächst werden alle URLs sowie sämtliche Benutzernamen (*@user*) aus der maximal 140-Zeichen-langen Twitter-Nachricht entfernt.
2. Anschließend wird anhand verschiedener, vordefinierter Sprachmodelle die Region (Asien, Europa, Amerika etc.) bestimmt, aus der die Nachricht gesendet wurde.
3. Im nächsten Schritt werden die Terme der Nachricht mit einem geographischen Wörterbuch wie z. B. *GeoNames* abgeglichen. Dabei werden geographisch irrelevante Wörter und Terme, die auf der ganzen Welt verwendet werden können (z. B. „Bowling“), herausgefiltert.
4. Abschließend werden die möglichen Kandidaten mit Hilfe der eingangs spezifizierten Faktoren ein weiteres Mal gefiltert.

Die Evaluation dieses Ansatzes in [3] hat gezeigt, dass durch die Kombination einer Vielzahl verschiedenartiger Faktoren die Trefferwahrscheinlichkeit verbessert werden kann.

Neben all diesen positiven Aspekten des Systems gibt es aber trotzdem auch Defizite. Beispielsweise wird bei den Twitter-Nachrichten nicht zwischen *lokalen* und *globalen* Nachrichten unterschieden. Dies führt bei globalen News zu einer Verfälschung der Ergebnisse und vergibt die Chance, nützliche (lokale) Indikatoren, welche Hinweise auf den Heimatort enthalten könnten, zu erfassen. Des Weiteren wäre ein Ausbau der unterstützten sozialen Netzwerke wünschenswert (z. B. Facebook und LinkedIn).

4 Fazit und Ausblick

In diesem Artikel wurden Ansätze zur automatischen Geodaten-Annotation von Fotos vorgestellt. Der Fokus wurde in diesem Zusammenhang auf Technologien und Systeme gelegt, welche einerseits Bildaufnahmen analysieren und andererseits mit Hilfe zusätzlicher Faktoren den Aufnahmeort dieser Fotos bestimmen können. Zu diesen Faktoren zählen *User-Tags*, *Referenzbilder* aus Datenbanken *für einen visuellen Vergleich*, *räumlich-zeitliche Zusammenhänge von Bildsequenzen* sowie *klimatische Bedingungen*, *Einwohnerzahlen* und *Aktivitäten in sozialen Netzwerken*. In jedem Unterabschnitt wurde zunächst überblicksweise beschrieben, wie der jeweilige Ansatz arbeitet, um anschließend auf eventuell

auftretende Herausforderungen einzugehen und mögliche Lösungsvorschläge zu erläutern.

Auch in Zukunft werden geographische Koordinaten eine wichtige Rolle in Bezug auf Such- und Archivierungsalgorithmen spielen. In diesem Zusammenhang werden demzufolge weitere Methoden zur automatischen Erkennung solcher Geodaten erforscht werden. Dabei wäre beispielsweise denkbar, dass die vorgestellten Ansätze miteinander kombiniert werden oder Systeme entwickelt werden, welche zum Zeitpunkt der Eingabe eines User-Tags bereits positionsrelevante Vorschläge anbieten können. Alle Technologien verfolgen allerdings ein Hauptziel, welches auch in zukünftigen Forschungsprojekten relevant bleiben wird: Ein Computer soll in der Lage sein Bilder zu analysieren, um diese möglichst so gut zu interpretieren, dass er aus dem Kontext der Aufnahme den Aufnahmeort extrahieren kann.

Literatur

- [1] H. Bay, T. Tuytelaars und L. Gool. „SURF: Speeded Up Robust Features“. In: *Computer Vision – ECCV 2006*. Hrsg. von Aleš Leonardis, Horst Bischof und Axel Pinz. Bd. 3951. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, S. 404–417. ISBN: 978-3-540-33832-1. DOI: 10.1007/11744023_32. URL: http://dx.doi.org/10.1007/11744023_32.
- [2] T. Brunner. „Geographic Information Retrieval: Identifikation der geographischen Lage von Zeitungartikeln“. Master’s thesis. Geographisches Institut, Winterthurerstrasse 190, 8057 Zürich: Universität Zürich - Abteilung GIS, Mai 2008.
- [3] C. Hauff und G.-J. Houben. „Placing images on the world map: a microblog-based enrichment approach“. In: *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. SIGIR ’12*. Portland, Oregon, USA: ACM, 2012, S. 691–700. ISBN: 978-1-4503-1472-5. DOI: 10.1145/2348283.2348376. URL: <http://doi.acm.org/10.1145/2348283.2348376>.
- [4] Jan J. *Wegweisend: GPS-Kamera Nikon Coolpix P6000 im Test*. Okt. 2008. URL: <http://www.netzwelt.de/news/78589-wegweisend-gps-kamera-nikon-coolpix-p6000-test.html>.
- [5] Yan K. und R. Sukthankar. „PCA-SIFT: a more distinctive representation for local image descriptors“. In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Bd. 2. Juni 2004, pages. DOI: 10.1109/CVPR.2004.1315206.
- [6] E. Kalogerakis, O. Vesselova, J. Hays, A.A. Efros und A. Hertzmann. „Image sequence geolocation with human travel priors“. In: *Computer Vision, 2009 IEEE 12th International Conference on*. Okt. 2009, S. 253–260. DOI: 10.1109/ICCV.2009.5459259.

- [7] D. Lowe. „Distinctive Image Features from Scale-Invariant Keypoints“. In: *Int. J. Comput. Vision* 60.2 (Nov. 2004), S. 91–110. ISSN: 0920-5691. DOI: 10.1023/B:VISI.0000029664.99615.94. URL: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [8] J. Luo, D. Joshi, J. Yu und A. Gallagher. „Geotagging in multimedia and computer vision—a survey“. English. In: *Multimedia Tools and Applications* 51 (1 2011), S. 187–211. ISSN: 1380-7501. DOI: 10.1007/s11042-010-0623-y. URL: <http://dx.doi.org/10.1007/s11042-010-0623-y>.
- [9] A. Rae, V. Murdock, P. Serdyukov und P. Kelm. „Working notes for the placing task at MediaEval 2011“. In: *MediaEval 2011 Workshop* (2011).
- [10] T. Rattenbury, N. Good und M. Naaman. „Towards automatic extraction of event and place semantics from flickr tags“. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. SIGIR '07*. Amsterdam, The Netherlands: ACM, 2007, S. 103–110. ISBN: 978-1-59593-597-7. DOI: 10.1145/1277741.1277762. URL: <http://doi.acm.org/10.1145/1277741.1277762>.
- [11] P. Serdyukov, V. Murdock und R. van Zwol. „Placing flickr photos on a map“. In: *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. SIGIR '09*. Boston, MA, USA: ACM, 2009, S. 484–491. ISBN: 978-1-60558-483-6. DOI: 10.1145/1571941.1572025. URL: <http://doi.acm.org/10.1145/1571941.1572025>.
- [12] H. Sergieh, G. Gianini, M. Döllner, H. Kosch, Elöd Egyed-Zsigmond und J.-M. Pinon. „Geo-based automatic image annotation“. In: *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval. ICMR '12*. Hong Kong, China: ACM, 2012, 46:1–46:8. ISBN: 978-1-4503-1329-2. DOI: 10.1145/2324796.2324850. URL: <http://doi.acm.org/10.1145/2324796.2324850>.
- [13] J. Stottinger, A. Hanbury, T. Gevers und N. Sebe. „Lonely but attractive: Sparse color salient points for object retrieval and categorization“. In: *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. Juni 2009, S. 1–8. DOI: 10.1109/CVPRW.2009.5204286.
- [14] H. Zhu und H. Wu. „Sloppy Tags and Metacrap? Quality of User Contributed Tags in Collaborative Social Tagging Systems“. In: *Quality of User Contributed Tags in Collaborative Social Tagging Systems (May 21, 2009)* (2010).