TECHNISCHE UNIVERSITÄT DRESDEN

ZENTRUM FÜR INFORMATIONSDIENSTE
UND HOCHLEISTUNGSRECHNEN
PROF. DR. WOLFGANG E. NAGEL

Master-Arbeit

zur Erlangung des akademischen Grades
Master of Science

Image Retrieval für Historische Bilder

Philipp Langen
(Geboren am 26. Dezember 1994 in Münsterlingen)

Hochschullehrer: Prof. Dr. Wolfgang E. Nagel
Betreuer: Dr. Christoph Lehmann & Dr. Taras Lazariv

Dresden, 3. Juni 2020

Hier Aufgabenstellung einfügen!

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die von mir am heutigen Tag dem Prüfungsausschuss der Fakultät Informatik eingereichte Master-Arbeit zum Thema:

Image Retrieval für Historische Bilder

vollkommen selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Dresden, den 3. Juni 2020

Philipp Langen

Kurzfassung

Abstract

Inhaltsverzeichnis

0.1 zitate 4

Literaturverzeichnis **5**

Motivation

Mit der zunehmenden Digitalisierung unserer Datenwelt stehen jedem von uns heute mit wenigen Klicks mehr Informationen zur Verfügung, als wir je analog aufnehmen könnten. Auch wenn diese Flut an Informationen ein enormes Potential für uns darstellt, so bürgt sie auch neue Herausforderungen. Um große Datenmengen sinnvoll nutzbar zu machen sind effiziente Suchwerkzeuge von zentraler Bedeutung. Dies gilt insbesondere auch für die Suche in großen Bilderdatenbanken. Bei der klassischen Bildersuche wird vom Nutzer eine Anfrage als String formuliert, auf welche das System eine Liste der Bilder liefert, die die größte Relevanz im Bezug auf die Anfrage haben. Dabei werden häufig nützliche Zusatzinformationen, sogenannte Metadaten, der Bilder wie Titel, Beschreibung, Aufnahmeort und Datum genutzt, um bessere Ergebnisse liefern zu können. Eine alternativer Ansatz der Suche ist die inhaltsbasierte Bildersuche (engl. Content-Based Image Retrieval kurz CBIR). Hierbei werden statt formulierten Anfragen Bilder als Anfragen gestellt. Ziel ist es Bilder mit gleichem oder ähnlichem Bildinhalt wie in der Anfrage als Ergebnis zurückzugeben. Dabei arbeitet das System direkt mit den Pixelinformationen der Bilder. Zusätzliche Metadaten sind daher nicht erforderlich. Im Folgenden wird die inhaltsbasierte Suche auch als Image Retrieval bezeichnet.

Ein interessanter Anwendungsbereich für Image Retrieval Systeme ist die Suche auf historischen Bildern. Diese weit gefasste Domäne enthält sehr heterogene Daten. Dabei finden sich nicht nur sehr unterschiedliche Bildmotive wie Gebäude, Naturaufnahmen oder Portraits, sondern auch unterschiedliche Aufnahmetechniken bedingt durch den technologischen Fortschritt von Zeichnungen, Malerei und Druck bis hin zur Photographie. Da Metadaten zu historischen Bildern erst bei der Digitalisierung hinzugefügt werden können sind diese oft gar nicht oder nur lückenhaft vorhanden, was die inhaltsbasierte Suche hier zu einem besonders geeigneten Ansatz macht. Mit der Umsetzung einer unterstützenden Suche für das UrbanHistory4D Projekt [1] ergibt sich ein konkreter Anwendungsfall für Image Retrieval Systeme. Hierbei handelt es sich um einen aktiven Forschungsbereich, in dem momentan unterschiedliche Suchsysteme analysiert werden.

Bei dem Image Retrieval Verfahren DELF (attentive DEep Local Features) [2], welches in dieser Arbeit untersucht wird handelt es sich um einen Deep Learning Ansatz. Durch den raschen Fortschritt im Bereich tiefer neuronaler Netzwerkarchitekturen der letzten Jahre erfreuen sich gelernte Ansätze immer größerer Beliebtheit. DELF erzielt auf bekannten Benchmarkdatensätzen wie Oxford5k [3] und Paris6k [4] sehr gute Ergebnisse. Besonders gut schneidet DELF im Vergleich auf dem eigens erstellten Google Landmarks Datensatz [5] ab. Dieser enthält mit über 1 mio. Bilder und 13k unterschiedlichen Motiven eine deutlich heterogenere Mischung an Objekten. Die gute Performanz auf diesem Datensatz lässt also hoffen, dass sich das Verfahren auch für die historische Domäne eignet.

Verwandte Arbeiten

Bei Information Retrieval handelt es sich um ein Problem aus dem Bereich der Computer Vision, welches bereits seit langem intensiv erforscht wird. In frühen Ansätzen versuchte man vor allem globale Beschreibungen von Bildern zu erstellen, um diese untereinander vergleichen zu können. Diese basierten zum Beispiel auf Farbhistogrammen oder Texturbeschreibungen [6]. Allerdings waren diese Ansätze oft sehr anfällig für Unterschiede in Beleuchtung, Skalierung und anderen Transformationen, wie sie bei unterschiedlichen Aufnahmen des selben Motivs auftreten können.
Ein wesentlicher Durchbruch gelang David G. Lowe 2004 mit der Entwicklung des SIFT-Verfahrens (Scale Invariant Feature Transform) [7]. Hierbei werden mehrere Konzepte vereint um Bildbeschreibungen zu erzeugen, die robuster gegenüber unterschiedlichen Transformationen sind. So arbeitet der SIFT Algorithmus beispielsweise nicht direkt auf den Bildern sondern im sogenannten Scale Space basierend auf unterschiedlich skalierten Versionen des Ursprungsbildes, um Resistenz gegen Skalierung zu schaffen. Das Verfahren nutzt Gauß-Filteroperationen um Merkmale zur Beschreibung hervorzuheben. Im ersten Schritt des Verfahren werden über die Suche nach lokalen Extrema bedeutsame Bildpunkte gewählt, für die anschließend einzelne Deskriptoren gefertigt werden. Das Bild wird also nicht global beschrieben, sondern über viele lokale Deskriptoren dargestellt. Die lokalen Deskriptoren ergeben sich aus Histogrammen der Gradientrichtungen der umliegenden Bildpunkte. Diese werden relativ zu der dominanten Gradientenrichtung in der Umgebung berechnet, was die Deskriptoren invariant gegenüber Rotationen macht. Lowes Entwicklung bildet den Ursprung für viele abgeleitet Verfahren, wie SURF[8], PCA-SIFT[9] und RIFT[10]. Auch in aktueller Forschung werden noch Image Retrieval Verfahren erforscht, die mit SIFT-Deskriptoren arbeiten [11].
Da Image Retrieval Systeme meist auf großen Bilddatenbanken eingesetzt werden und somit für eine Suchanfrage viele Vergleiche zwischen Bildern durchgeführt werden müssen ist es sinnvoll Bildrepräsentationen so kompakt wie möglich zu gestalten, um die Laufzeit in Grenzen zu halten. Insbesondere bei Verfahren, die lokale Deskriptoren erstellen und häufig hunderte oder tausende Merkmale pro Bild extrahieren, kann mit einer guten Kodierung viel Rechenzeit gespart werden. Ein beliebter Ansatz zur Erstellung kompakter Darstellungen aus lokalen Deskriptoren ist das BOVW-Modell (Bag-of-Visual-Words) [12]. Hierbei werden zunächst alle aus einem Datensatz extrahierte Deskriptoren mittels Clusteranalyse (bspw. K-Means-Clustering [13]) in Gruppen eingeteilt. Deskriptoren in der gleichen Gruppe werden dabei auf das selbe "visuelle Wort" abgebildet. Als Beschreibung des Gesamtbilds dient ein Histogramm über die im Bild enthaltenen visuellen Wörter. Bei diesem Verfahren geht durch den Quantisierungsfehler ein Teil der Information verloren. Das ebenfalls auf Clustering basierte VLAD-Verfahren [14] versucht diese Information nutzbar zu machen, indem es statt der Vorkommen die Quantisierungsfehler akkumuliert, die beim abbilden auf die nächsten visuellen Worte entstehen.
Der finale Teil einer Image Retrieval Pipeline ist das Vergleichen des Anfragebildes mit den Bildern in der Suchdatenbank auf Basis der erzeugten Deskriptoren, mittels einer geeigneten Distanzmetrik (z.B. euklidische Distanz). Dabei ist ein erschöpfender Suchansatz auf Grund der Menge an Daten oft nicht sinnvoll. Stattdessen

0.1 zitate

[6][7][8][9][11][12][14][15][16][17][18][19][20][21][22][23][24][25][26][27][28][2][29]

Literaturverzeichnis

[1] Ferdinand Maiwald, Jonas Bruschke, Christoph Lehmann, and Florian Niebling. A 4D Information System for the Exploration of Multitemporal Images and Maps using Photogrammetry, Web Technologies and VR/AR. *Virtual Archaeology Review*, 10:1, 07 2019.

[2] Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, and Bohyung Han. Large-Scale Image Retrieval with Attentive Deep Local Features. pages 3476–3485, 10 2017.

[3] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[4] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[5] Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig Adam, Ulrich Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua, Hartmut Neven, and Jay Yagnik. Tour the World: A Technical Demonstration of a Web-Scale Landmark Recognition Engine. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, page 961–962, New York, NY, USA, 2009. Association for Computing Machinery.

[6] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, December 2000.

[7] David Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60:91–, 11 2004.

[8] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up Robust Features. volume 3951, pages 404–417, 07 2006.

[9] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more Distinctive Representation for Local Image Descriptors. volume 2, pages II–506, 05 2004.

[10] Cordelia Schmid and J. Ponce. Semi-Local Affine Parts for Object Recognition. *BMVC04*, 08 2004.

[11] Miaojing Shi, Yannis Avrithis, and Hervé Jégou. Early Burst Detection for Memory-Efficient Image Retrieval. 06 2015.

[12] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. Understanding Bag-of-Words Model: A Statistical Framework. *International Journal of Machine Learning and Cybernetics*, 1:43–52, 12 2010.

[13] James B. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. 1967.

[14] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Perez. Aggregating Local Descriptors into a Compact Image Representation. pages 3304 – 3311, 07 2010.

[15] Jerome Friedman, Jon Bentley, and Raphael Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Softw.*, 3:209–226, 09 1977.

[16] Hervé Jégou, Matthijs Douze, and Cordelia Schmid. Product Quantization for Nearest Neighbor Search. *IEEE transactions on pattern analysis and machine intelligence*, 33:117–28, 01 2011.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. 7, 12 2015.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25, 01 2012.

[19] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv 1409.1556*, 09 2014.

[20] Matthew Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Neural Networks. volume 8689, 11 2013.

[21] Ali Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN Features Off-the-Shelf: an Astounding Baseline for Recognition. *Arxiv*, 03 2014.

[22] Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky. Neural Codes for Image Retrieval. volume 8689, 04 2014.

[23] Joe Ng, Fan Yang, and Larry Davis. Exploiting Local Features from Deep Networks for Image Retrieval. pages 53–61, 06 2015.

[24] Eva Mohedano, Kevin McGuinness, Noel O'Connor, Amaia Salvador, Ferran Marques, and Xavier Giró-i Nieto. Bags of Local Convolutional Features for Scalable Instance Search. pages 327–331, 04 2016.

[25] Albert Gordo, Jon Almazan, Jerome Revaud, and Diane Larlus. Deep Image Retrieval: Learning Global Representations for Image Search. volume 9910, pages 241–257, 10 2016.

[26] Filip Radenović, Giorgos Tolias, and Ondřej Chum. CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples. volume 9905, pages 3–20, 10 2016.

[27] Ali S. Razavian, Josephine Sullivan, Stefan Carlsson, and Atsuto Maki. [paper] visual instance retrieval with deep convolutional networks. *ITE Transactions on Media Technology and Applications*, 4(3):251–258, 2016.

[28] Lingxi Xie, Richang Hong, Bo Zhang, and Qi Tian. Image classification and retrieval are one. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ICMR '15, page 3–10, New York, NY, USA, 2015. Association for Computing Machinery.

[29] Amaia Salvador, Xavier Giró-i Nieto, Ferran Marques, and Shin'ichi Satoh. Faster R-CNN Features for Instance Search. *Arxiv*, 04 2016.