Über die Ungleichheit im Gleichen. Erkennung unterschiedlicher Reproduktionen desselben Objekts in kunsthistorischen Bildbeständen

Schneider, Stefanie
stefanie.schneider@itg.uni-muenchen.de
Ludwig-Maximilians-Universität München, Deutschland

Es finden sich mittlerweile mannigfaltige Studien, die Objekte historischen Ursprungs in den Fokus rücken und zwischen ihnen bestehende Relationen und Ähnlichkeitsverhältnisse mathematisch abzubilden versuchen (darunter Bergel et al., 2013; Monroy et al., 2014; Hentschel et al., 2016). Eine fundamentale Herausforderung *per se* heterogener historischer Inventare bleibt dabei unberührt: digitale Reproduktionen ein und desselben Objekts, sogenannte *Near-duplicates* oder *Near-replicas*, die als separate Einträge in Aggregatordatenbanken vorgehalten werden und sich bspw. hinsichtlich ihres Farbstichs oder ihrer Helligkeit unterscheiden. Aufgrund nicht-standardisierter, teils unstrukturierter oder selten kontrollierten Vokabularen zugeordneter Metadaten ist es zumeist nicht oder nur mit größtem Aufwand möglich, derartige „Kopien" auf Basis textueller Information sowohl nachhaltig als auch zuverlässig ohne händische Nacharbeit zu verknüpfen.

Unser Ansatz zielt auf dreierlei: erstens die automatische Zusammenführung unterschiedlicher Reproduktionen desselben Objekts; die zweitens das *Retrieval* variierender Reproduktionen erlaubt, um weiterführende Analysen, z. B. quellenkritischer Art, über eben jenes Objekt anstoßen zu können; und drittens die Extraktion textueller Information von Objekten, die ausschließlich visuell, d. h. als digitales Bild, vorliegen und, im Sinne einer *Reverse Image Search*, mit einem bereits annotierten Inventar von Objekten abzugleichen sind. Auf diese Weise wird nicht nur eine effiziente, bildbasierte Suche in Datenbanken insbesondere (aber nicht ausschließlich) kunsthistorischer Objekte ermöglicht, sondern nahezu *Bias*-freie statistische Untersuchungen, wie sie durch den Einfluss häufig reproduzierter Werke bislang nicht gegeben waren.[1]

Methode

Wir stützen uns hauptsächlich auf *Scale Invariant Feature Transform* (*SIFT*; Lowe, 1999; Lowe, 2004), um aus Bildern historischer Objekte lokale Schlüsselpunkte (*Keypoints*) zu extrahieren und mittels Deskriptoren (*Descriptors*) weiterverarbeiten zu können. Schlüsselpunkte bilden einzelne Interessenregionen eines Bildes ab, die statistisch, aber nicht notwendigerweise semantisch relevante Merkmale tragen, und stellen sie in einem 128-dimensionalen Histogramm dar. Im mathematischen Sinne sind sie Extremwerte, die über einen Raum mehrfach skalierter und mit einem Gaußfilter geglätteter Bilder ermittelt werden. Mit üblichen Ähnlichkeits- und Distanzmaßen, z. B. der euklidischen Distanz, ist es so möglich, die Nähe zwischen zwei Schlüsselpunkten zu quantifizieren, und demnach auch, über die Summe der *matchenden* Schlüsselpunkte, die Nähe zwischen zwei Digitalisaten; wobei anzunehmen ist, dass die Anzahl der übereinstimmenden Schlüsselpunkte für variierende Abbildungen desselben Objekts höher ist als für Abbildungen unterschiedlicher Objekte.

Ein Bild wird je nach Größe und Detailgrad mit Hunderten bis Tausenden derartiger Schlüsselpunkte assoziiert. Daraus resultierende computationale Kosten fangen wir durch drei Erweiterungen des Verfahrens ab. Erstens reduzieren wir die Dimensionalität der Deskriptoren mittels *Principal Component Analysis* (*PCA*). Im Gegensatz zu Ke und Sukthankar (2004) greifen wir nicht in den Deskriptionsprozess selbst ein, sondern ermitteln den Eigenraum auf Basis der standardmäßig durch *SIFT* eruierten Deskriptoren. Zweitens verringern wir die Anzahl der Schlüsselpunkte, indem wir zunächst die mit dem höchsten Kontrast auswählen und auf Basis dessen jene filtern, welche die größten Interessenregionen charakterisieren. Damit werden auf flächenmäßig kleinen Arealen zu findende, kontrastreiche Schlüsselpunkte getilgt, die bspw. in textuellen Ergänzungen von Kupferstichen auftreten und für das *Matching* irrelevant bis schädlich sind. Drittens setzen wir mit *Hierarchical Navigable Small World* (*HNSW*; Malkov und Yashunin, 2016) einen *Approximate Nearest Neighbor*-Ansatz mit polylogarithmischer Komplexität ein, der in aktuellen repräsentativen Benchmarks andere Graph-basierte Ansätze in Präzision und Schnelligkeit übertrifft (Aumueller et al., 2018). Eine adaptive Intensitätskorrektur jedes Bildes durch *Contrast Limited Adaptive Histogram Equalization* (*CLAHE*; Zuiderweld, 1994) wird vor der Extraktion der Schlüsselpunkte durchgeführt, um stark über- oder unterbelichtete Reproduktionen anzupassen.

Daten

Ein geeigneter *Gold Standard* wird in drei Schritten etabliert. Zunächst ziehen wir eine Zufallsstichprobe von 3.581 kunsthistorischen Objekten, die in der Datenbank *ArteMIS* des Instituts für Kunstgeschichte der Ludwig-Maximilians-Universität München verzeichnet sind[2] und einen angemessenen Querschnitt verschiedener kunsthistorischer Stile und Epochen erlauben; Holzschnitte sind ebenso inkludiert wie realistische Landschaftsmalerei

und Werke des französischen Impressionismus.[3] In einem zweiten Schritt speisen wir Titel und Künstler jener Objekte in die 94 Datenbanken kumulierende *Application Programming Interface* von Prometheus.[4] Da in den jeweiligen Suchergebnissen nicht nur Digitalisate ein und desselben Objekts zu finden sind – Vorzeichnungen und aufgrund ihrer Metadaten ähnliche Reproduktionen sind auch darunter –, schließen wir einen dritten Schritt an, in dem auf unterschiedliche Objekte referenzierende Abbildungen manuell entfernt werden. Es verbleiben 9.934 Reproduktionen. Ein derart selegiertes Digitalisat trägt einen eineindeutigen Identifikator, der sowohl auf das es abbildende Objekt zeigt als auch auf die an das Digitalisat gekoppelten Metadaten weist.

Um weitere in der bildarchivarischen Praxis existente, aber durch zuvor extrahierte *reale* Digitalisate unzureichend abgedeckte Modifikationen, bspw. größere Änderungen des Kontrasts oder der Sättigung eines Bildes, untersuchen zu können, generieren wir 278.152 zusätzliche, sogenannte *synthetische* Kopien. Jede Ursprungsreproduktion wird dementsprechend dupliziert und 28 mathematischen Transformationen unterzogen; ähnlich zu jenen in Ke et al. (2004), Qamra et al. (2005) und Foo et al. (2007). Unter anderem modelliert werden sich in ihrer Stärke unterscheidende nicht-lineare Verzerrungen, die Wölbungen nahe des Buchrückens von Fotografien kunsthistorischer Publikationen suggerieren.

Evaluation

Drei im *Information Retrieval* gewöhnliche Gütekriterien dienen der Evaluation der angewandten Methoden: Precision, Recall und F_1-Maß. Wir gehen wie folgt vor. Der Satz an Objekten, die mit realen und synthetischen Reproduktionen assoziiert sind, wird unterteilt in 25 zufällig separierte Trainings- und Teststichproben, wobei jeweils 80 Prozent der Objekte der Trainings- und 20 Prozent der Teststichprobe zuzuordnen sind. Auf Basis der 25-fachen Kreuzvalidierung erhalten wir durchschnittliche Werte für Precision, Recall und F_1-Maß, die von einem jeweils gegebenen Schwellenwert abhängen, der die minimale Anzahl der Schlüsselpunkte bezeichnet, die zwischen zwei Digitalisaten übereinstimmen müssen, damit diese als unterschiedliche Reproduktionen desselben Objekts gelten und entsprechend zusammengeführt werden können. Der jeweils für eine Parameterkonstellation optimale Schwellenwert bildet sich aus dem Modus der 25 Einzelschwellenwerte, die mit dem höchsten F_1-Maß verknüpft sind.

Ergebnisse

Wir stellen fest, dass aufgrund des hohen Anteils in den Reproduktionen enthaltener digitaler Artefakte – unter anderem Bildrauschen und Unschärfe –, Konfigurationen mit im Vergleich zu Standardwerten höherem Skalenparameter #, der die Stärke des in *SIFT* angelegten gaußschen Weichzeichners reguliert, und niedrigeren Schwellenwerten, welche die Aufnahme von kontrastarmen oder auf Kanten situierten Schlüsselpunkten steuern, zu bevorzugen sind. Eine Reduktion der Dimensionalität der Deskriptoren und der Anzahl der Schlüsselpunkte auf jeweils 50 führt zu marginalen Einbußen in Precision und Recall, steigert jedoch maßgeblich die Performanz und mindert den notwendigen Speicherbedarf. Eine so klassifizierte, aus 500 Objekten bestehende Zufallsstichprobe resultiert in Precision = 0,9857, Recall = 0,9820 und F_1-Maß = 0,9839, wenn für *HNSW* moderate Kompromisse zwischen der Geschwindigkeit, die der Aufbau des Index und die eine Suche im Index benötigt, formuliert werden; mindestens 7 näherungsweise übereinstimmende Schlüsselpunkte sind erforderlich, damit Digitalisate als demselben Objekt zugehörig erkannt werden. Größere Einbrüche in Recall, d. h. bis zu 5 Prozentpunkte, sind für stärkere Farbänderungen und nicht-lineare Verzerrungen zu beobachten. Insbesondere drei Gruppen von Objekten erfordern weitere Anpassungen: Digitalisate von Druckgrafiken mit hohen Kontrastunterschieden; Reproduktionen, die Rahmen oder rahmenähnliche Strukturen abbilden; Werke des Impressionismus und nicht-gegenständliche oder diffuse Werke, die unterdurchschnittlich viele Schlüsselpunkte, teilweise nur bis zu 10, produzieren.

Auch ohne zusätzliche Modifikationen zeigt sich, dass die hier präsentierte Methode hinreichend exakte Ergebnisse erwarten lässt und kaum, oder lediglich im Falle hoch spezialisierter Korpora, manuelle Adjustierungen erfordert; selbst wenn stärkere Abweichungen in Kontrast oder Sättigung auftreten. Durch die Integration eines *Approximate Nearest Neighbor*-Ansatzes ist weiterhin gewährleistet, dass das Verfahren auch auf größere historische Bildbestände skaliert.

Fußnoten

1. Dies schließt anderweitige Verzerrungen, bspw. einen *Selection Bias*, natürlich nicht aus.
2. Eine Online-Schnittstelle ist zu erreichen unter http://artemis.lmu.de/ (25.09.2018).
3. Ausgenommen werden Reproduktionen eindeutig dreidimensionaler Objekte, z. B. Plastiken und Skulpturen, da sich diese zusätzlich durch den bei der Aufnahme eingenommenen Blickwinkel unterscheiden können und folglich gesondert zu evaluieren wären.
4. http://www.prometheus-bildarchiv.de/ (25.09.2018).

Bibliographie

Aumueller, Martin / Bernhardsson, Erik / Faitfull, Alec (2018): *ANN Benchmarks*, http://sss.projects.itu.dk/ann-benchmarks/index.html (26.09.2018).

Bergel, Giles / Franklin, Alexandra / Heaney, Michael / Arand-Jelovic, Relja / Zisserman, Andrew / Funke, Donata (2013): „*Content-based Image Recognition on Printed Broadside Ballads. The Bodleian Libraries' Imagematch Tool*“, in: Proceedings of the IFLA World Library and Information Congress.

Foo, Jun Jie / Zobel, Justin / Sinha, Ranjan (2007): „*Clustering Near-duplicate Images in Large Collections*“, in: Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval 21–30.

Hentschel, Christian / Wiradarma, Timur P. / Sack, Harald (2016): „*An Approach to Large Scale Interactive Retrieval of Cultural Heritage*“, in: Proceedings of the 23th IEEE International Conference on Image Processing 3693–3697.

Ke, Yan / Sukthankar, Rahul (2004): „*PCA-SIFT. A More Distinctive Representation for Local Image Descriptors*“, in: Proceedings of the IEEE International Conference on Computer and Pattern Recognition 506–513.

Ke, Yan / Sukthankar, Rahul / Huston, Larry (2004): „*An Efficient Parts-based Near-duplicate and Sub-image Retrieval System*“, in: Proceedings of the >12th ACM International Conference on Multimedia 869–876.

Lowe, David G. (1999): „*Object Recognition from Local Scale-invariant Features*“, in: Proceedings of the 7th IEEE International Conference on Computer Vision 1150–1157.

Lowe, David G. (2004): „*Distinctive Image Features from Scale-invariant Keypoints*“, in: International Journal of Computer Vision 60 (2): 91–110.

Malkov, Yury A. / Yashunin, Dmitry A. (2016): *Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs*.

Monroy, Antonio / Bell, Peter / Ommer, Björn (2014): „*Morphological Analysis for Investigating Artistic Images*“, in: Image and Vision Computing 32 (6): 414–423.

Qamra, Arun / Meng, Yan / Chang, Edward Y. (2005): „*Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition*“, in: IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3): 379–391.

Zuiderveld, Karel (1994): „*Contrast Limited Adaptive Histogram Equalization*“, in: Academic Press 474–485.