

Harald Eberle

# Inhaltsbasierte Bilderschließung durch Crowd und Cloud

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 489–502  
<https://doi.org/10.25364/978390337423226>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz, ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Harald Eberle, Vorarlberger Landesbibliothek, [harald.eberle@vorarlberg.at](mailto:harald.eberle@vorarlberg.at) | ORCID iD: 0000-0001-9251-2924

## Zusammenfassung

Die Vorarlberger Landesbibliothek ist bestrebt, ihre Bildsammlungen durch Retrodigitalisierung zu sichern und über ihr Repositorium volare benutzungsfreundlich bereitzustellen. Allerdings können die durch Massendigitalisierung entstandenen Datenmengen mit den vorhandenen personellen Ressourcen nicht in der gewünschten Qualität abgearbeitet werden. Deshalb wird mit neuartigen Methoden versucht, eine möglichst effiziente Bilderschließung zu bewerkstelligen. Über die Crowdsourcing-Plattform sMapshot lassen sich Landschaftsbilder nicht nur verorten, sondern exakt in einem digitalen 3D-Globus platzieren. Die auf diese Weise produzierten Geoinformationen ermöglichen es, Namen der abgebildeten Orte, beispielsweise von Städten, Ortsteilen oder Bergen, zu berechnen. Zudem wird mit maschinellem Sehen<sup>1</sup> experimentiert, um historische Aufnahmen automatisiert zu beschreiben.

**Schlagwörter:** Crowdsourcing; Georeferenzierung; Bildbeschreibung; Maschinelles Sehen; Künstliche Intelligenz; Metadatenanreicherung

## Abstract

### Content-Based Image Description Through Crowd and Cloud

The Vorarlberg State Library uses retro digitisation to facilitate long-term and user-friendly access to its image collections through the institutional repository volare. As traditional resources and means are limited when it comes to processing large data volumes, new and innovative methods are being explored to generate high-quality descriptions and metadata in a more efficient way. The crowdsourcing platform sMapshot not only allows the georeferencing of digital image files to a geographical location but also the positioning of landscape images on a virtual 3D globe. This geoinformation enables the calculation of place names such as cities, districts or mountains. In the context of machine-based descriptions for historic photographs, the Vorarlberg State Library is also experimenting with machine vision.

**Keywords:** Crowdsourcing; georeferencing; image description; machine vision; artificial intelligence; metadata enrichment

---

1 Das maschinelle Sehen als Teilgebiet der Informatik beschäftigt sich mit der Erfassung visueller Informationen eines Computers in seiner Umwelt. Dabei geht es in erster Linie um das einfache Erfassen von Gegenständen und Personen, und nicht um die Interpretation und das Verständnis eines tieferen Sinnes.

## 1. Einleitung

Zu den Aufgaben von Gedächtnisinstitutionen gehört es, das kulturelle Erbe in all seinen medialen Ausprägungen zu sammeln, zu dokumentieren und zu vermitteln. Seit der Jahrtausendwende werden in vielen Institutionen immer größere Retrodigitalisierungsprojekte in Angriff genommen, um analoge Vorlagen digital zu sichern. Dadurch können die Originalvorlagen geschont und eine bessere Zugänglichkeit ermöglicht werden. Waren diese Projekte in der Vergangenheit oft auf Textdokumente beschränkt, erweitert sich der Fokus nun auch auf Lichtbilderwerke.<sup>2</sup> Aufgrund des Umfangs der Projekte stellt die manuelle Erschließung der Bilddaten die einzelnen Institutionen vor große finanzielle und personelle Herausforderungen. „Der klassische Ansatz – die manuelle Annotation des Bildinhaltes mittels alphanumerischer Texte – hat sich in der Vergangenheit als zu fehleranfällig und zu kostenintensiv erwiesen“.<sup>3</sup> Die ständig wachsende Menge an digitalisierten und digital entstandenen Bilddaten erfordert neuartige Methoden, um eine regelwerkskonforme Erfassung und eine verbesserte Auffindbarkeit in den Repositorien zu ermöglichen. Die Vorarlberger Landesbibliothek verfolgt zwei Ansätze, um die anfallenden Bilddaten effizient und qualitativ hochwertig zu erschließen und so einen erheblichen Mehrwert für die Forschung zu generieren. Der erste Ansatz bezieht sich auf die dreidimensionale Georeferenzierung historischer Landschaftsbilder durch Crowdsourcing über die Plattform sMapshot. Die Webanwendung ermöglicht es, einen virtuellen Globus der Vergangenheit aufzubauen und die genauen Ortsnamen zu berechnen, die im Bild sichtbar sind.<sup>4</sup> Der zweite Ansatz nutzt maschinelles Lernen, ein Teilgebiet der künstlichen Intelligenz, um Bildklassifizierung durch Objekterkennung und Personenidentifizierung durch Gesichtserkennung zu ermöglichen. Dabei wird versucht, die abgebildeten Objekte und Personen automatisiert und mit entsprechenden Normdaten zu beschreiben. Um die Möglichkeiten der oben beschriebenen Ansätze sinnvoll einschätzen zu können, wird in diesem Beitrag auf die Technik und Workflows sowie deren Vor- und Nachteile eingegangen. Ziel ist es, einen Überblick über die Möglichkeiten und Anwendungsszenarien des Repositorienmanagements zu geben. Einige Abschnitte dieses Aufsatzes wurden bereits veröffentlicht.<sup>5</sup>

---

2 Helm, W. (2019), S. 127–134.

3 Volmer, S. (2012), S. 3.

4 Produit, T.; Ingensand, J. (2019), S. 273f.

5 Eberle, H. (2020)

## 2. Georeferenzierung historischer Landschaftsbilder durch Crowdsourcing

Historische Landschaftsbilder liefern Informationen über die Veränderung der Landschaft und sind eine wertvolle Informationsquelle, beispielsweise im Bereich der Raum- und Stadtplanung, aber auch in der Regional- und der Kulturgeschichtsforschung. Um historische Aufnahmen der Öffentlichkeit zugänglich machen zu können, sind die in den Gedächtnisinstitutionen meist nur rudimentär vorhandenen Bildbeschreibungen oft nicht ausreichend. Das von der westschweizer Fachhochschule für Management und Ingenieurwesen Waadt (HEIG-VD) entwickelte Projekt sMapshot hat sich zum Ziel gesetzt, eine dreidimensionale Georeferenzierung historischer Bilder mit Hilfe des geografischen Wissens von Freiwilligen zu realisieren. Diese Geoinformationen können verwendet werden, um einen „virtuellen Globus der Vergangenheit aufzubauen und um die genauen Ortsnamen zu berechnen, die im Bild sichtbar sind“.<sup>6</sup> Damit entsteht die Möglichkeit, historische Aufnahmen exakt zu lokalisieren, mit Geoinformationen anzureichern und so die Bildrecherche zu erleichtern.

### 2.1. Methoden zur Georeferenzierung von Bildern

In der Literatur werden drei Methoden zur Georeferenzierung von Einzelbildern beschrieben.<sup>7</sup> Die erste und einfachste Methode ist die Verortung durch eine Punktkoordinate. Dabei kann der Punkt des Aufnahmestandorts oder die exakte Position des Objektes erfasst werden. Die Referenzierung des Aufnahmepunktes beherrschen bereits viele GPS-Kameras automatisiert. Zudem ermitteln einige Geräte die Blickrichtung mit Hilfe eines digitalen Kompasses. Die Punktverortung ermöglicht zwar eine Beschreibung des Standortes, nicht aber die genaue Definition des sichtbaren Bildbereichs. Für die Nachnutzung hilfreicher ist die Erfassung des Objektmittelpunktes auf der Erdoberfläche. Für die Verortung von senkrecht aufgenommenen Bildern wird meist die zweite Methode, die Georeferenzierung einer Fläche durch mindestens vier Randkoordinaten, angewendet. Die Flächenverortung eignet sich hervorragend für senkrecht aufgenommene Luftbilder, sogenannte Orthofotos. Bei Bildern, die schräg aus der Luft, beispielsweise aus Flugzeugen, Hubschraubern oder Heißluftballons aufgenommen werden, eignet sich diese Methode nicht. Bei solchen Aufnahmen werden nämlich nicht rechteckige Flächen, sondern komplexe Polygone im Raum abgebildet. Für diese Art der Bilder ist die dritte Me-

---

<sup>6</sup> Produit, T.; Ingensand, J. (2019), Anm. 3, S. 273.

<sup>7</sup> Vgl. Produit, T.; Ingensand, J. (2016)

thode, das Monoplotting, geeignet. Es basiert auf dem Konzept, mehrere Referenzpunkte, sogenannte Ground Control Points, im Bild und im virtuellen Globus zu identifizieren und deren genaue Position zu berechnen. Neben dem Bildaufnahme-punkt lassen sich so zusätzlich die drei Bildorientierungswinkel Omega, Phi und Kappa ermitteln.<sup>8</sup> Diese Informationen dienen einer genauen Positionierung des Rasterbildes über dem dreidimensionalen virtuellen Globus und ermöglichen so die Berechnung des im Bild sichtbaren Bereiches.

### 2.1.2. sMapshot

Das Konzept zur Verortung einzelner Aufnahmen in dreidimensionalen Landschaftsmodellen wurde bereits im Jahr 2015 vorgestellt.<sup>9</sup> Seit 2016 entwickelt die HEIG-VD die Plattform sMapshot, die es freiwilligen Teilnehmenden ermöglicht, historische Bilder mittels Monoplotting zu georeferenzieren. Als erste Institution stellte die Bibliothek der ETH Zürich im Jänner 2018 mit ihrer Kampagne Schweizer Luftbilder von Walter Mittelholzer auf sMapshot zur Verfügung.<sup>10</sup> Die Nutzung von sMapshot außerhalb der Schweiz war vorerst nicht vorgesehen, da lediglich das 3D-Modell und die Luftbilder der swisstopo (Bundesamt für Schweizer Landestopografie) als hochauflösende Basisdaten zur Verfügung standen, nicht aber die Daten anderer Länder. Aus diesem Grund beauftragten die Vorarlberger Landesbibliothek und das Landesamt für Vermessung und Geoinformation die Erweiterung von sMapshot zur Georeferenzierung von Bildern auf dem gesamten Globus. Zudem sollten die bereits vorhandenen und qualitativ hochwertigen Geobasisdaten aus Vorarlberg bzw. Österreich eingebunden werden. Mit der Sammlung *Historische Schrägluftbilder der Alpine Luftbild GmbH* lancierte die Vorarlberger Landesbibliothek im Sommer 2020 ihre erste Kampagne auf Basis der österreichischen Verwaltungsgrundkarten (basemap.at) auf sMapshot.<sup>11</sup>

Die Bearbeitung historischer Landschaftsbilder auf sMapshot nutzt spielerische Elemente und verläuft in mehreren Schritten. Bilder können sowohl anonym als auch unter einem Namen bzw. einem Pseudonym referenziert werden. Die Registrierung bietet den Vorteil, dass die Arbeit der jeweiligen Person sichtbar wird. Die teilnehmende Person wählt zunächst auf einer Landkarte ein Bild aus, welches georeferenziert werden soll. Im ersten Schritt wird der ungefähre Aufnahmestandort bestimmt. Im zweiten Schritt muss die ungefähre Blickrichtung festgelegt wer-

---

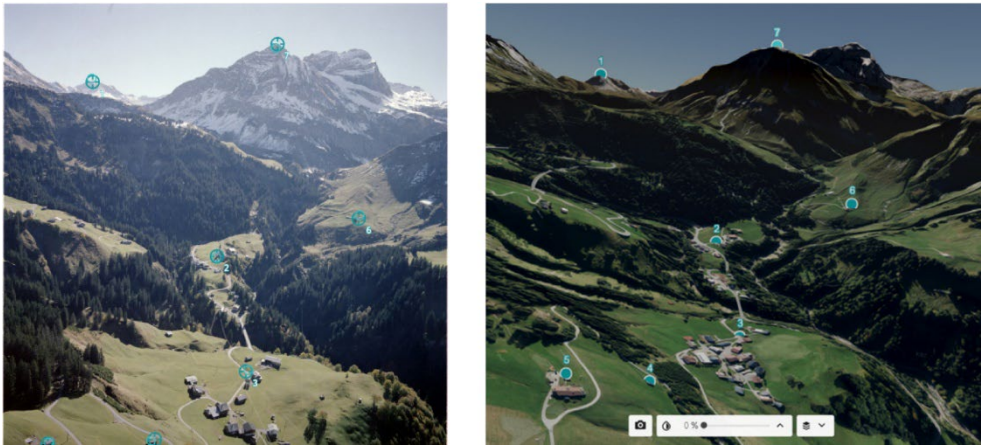
8 Produit, T.; Ingensand, J. (2016), Anm. 3, S. 273.

9 Produit, T.; Ingensand, J. (2018), S. 113–126.

10 Graf, N. (2018)

11 Huonder, F. (2020)

den. Anschließend platziert sMapshot das Bild neben dem bereits provisorisch ausgerichteten Kartenausschnitt. Nun beginnt die eigentliche Georeferenzierung, bei der mindestens sechs übereinstimmende Punkte zwischen Bild und virtuellem 3D-Globus gefunden werden müssen.



**Abb. 1: Zur Georeferenzierung müssen mindestens sechs übereinstimmende Punkte sowohl im Bild als auch im virtuellen Globus angegeben werden.**

Als Punkte eignen sich besonders geografisch markante Bergspitzen, Straßenkreuzungen oder Gebäudeecken. Die Verwendung von Gebäudedachflächen eignet sich hingegen nicht, da die Gebäudehöhen aus dem Geländemodell herausgerechnet wurden. Idealerweise sollten die Punkte gleichmäßig über das Bild verteilt sein, damit eine möglichst große Abdeckung gewährleistet ist. Wird der Vorgang abgeschlossen, so muss die Referenzierung noch systemseitig validiert werden. Mitarbeitende der Vorarlberger Landesbibliothek prüfen dann die Qualität der Verortung und haben die Möglichkeit, entsprechende Verbesserungen vorzunehmen. Bei einer mangelhaften Verortung besteht zudem die Option, diese abzulehnen und den Freiwilligen erneut zur Referenzierung bereitzustellen. Insgesamt zeigt sich, dass die bereitgestellten Kampagnen von der Crowd sehr gut angenommen und rasch abgearbeitet werden. Zudem ist die Qualität der erstellen Daten in der Regel so gut, dass lediglich in Ausnahmefällen Verbesserungen vorgenommen werden müssen.

## 2.2. Metadatenanreicherung

Nach Abschluss einer Kampagne können für alle Bilder der Aufnahmestandort, die Aufnahmehöhe und der Footprint aus sMapshot exportiert werden. Beim Footprint handelt es sich um die Koordinaten des Polygons der auf der Aufnahme abgebildeten Fläche. Diese Polygone lassen sich mittels Geoverarbeitungswerkzeugen mit beliebigen Geodatenätzen verschneiden, um so weitere Stichwörter zum abgebildeten Bildausschnitt zu schaffen. Besonders gut eignen sich dazu Geodaten von Katastralgemeinden, Bergspitzen, Fließgewässer, Parzellen oder Flurnamen. Weitreichende Bildausschnitte generieren eine Fülle an neuen Metadaten, umso mehr sollte bei den jeweiligen Stichwörtern auf Relevanz geachtet werden. Zudem können auch der exakt erfasste Ausschnitt, der Aufnahmepunkt, die Flughöhe sowie die Blickrichtung einen Mehrwert für künftige Nutzungsszenarien und Forschungen generieren.

## 2.3. Aktueller Stand

Mit Sommer 2021 hat die Vorarlberger Landesbibliothek drei Kampagnen mit rund 13.000 Einzelbildern zur Georeferenzierung bereitgestellt.<sup>12</sup> Pro Tag werden von der Crowd durchschnittlich 100 Bilder abgearbeitet und von den Mitarbeitenden der Vorarlberger Landesbibliothek validiert. Um Aufmerksamkeit für das Projekt im eigenen Bundesland zu erregen, strahlte das ORF-Landesstudio Vorarlberg am 2. August 2021 einen Beitrag im Rahmen der Nachrichtensendung Vorarlberg Heute aus.<sup>13</sup> Nach Ausstrahlung dieses Beitrags wurden rund 15 neue Personen, die sich an der Verortung beteiligten, verzeichnet. Insgesamt ist allerdings festzustellen, dass sich besonders die bereits bestehenden Mitglieder der Schweizer sMapshot-Community am intensivsten an dem Projekt beteiligen. Die beiden aktivsten Personen verorteten bereits jeweils über 20.000 Bilder in allen von sMapshot-Partnern zur Verfügung gestellten Kampagnen.<sup>14</sup> Die Logfiles zeigen eine durchschnittliche Bearbeitungsdauer von 5,3 Minuten pro Bild. Bei diesen beiden gerade erwähnten Personen ergibt das eine erbrachte Leistung von jeweils über 1.700 Stunden, also mehr als das jährliche Arbeitszeitausmaß einer vollzeitbeschäftigten Arbeitskraft. Insgesamt wurden über die Plattform sMapshot bereits über 187.000 Bilder aus den diversen Institutionen verortet, das entspricht einer erbrachten Leistung der Crowd von rund 16.500 Stunden.

---

12 <https://smapshot.heig-vd.ch/owner/vlb>

13 sMapshot. Die partizipative Zeitmaschine. vorarlberg.ORF.at, 02.08.2021: <https://vorarlberg.orf.at/stories/3115329/>

14 <https://smapshot.heig-vd.ch/#Teilnehmer>

### **2.3.1. Inhaltsbasierte Bilderschließung mittels visueller Merkmale durch Maschinelles Sehen**

Das Thema Objekterkennung in digitalen Bildern durch künstliche Intelligenz ist derzeit in aller Munde. Ob in industriellen Herstellungsprozessen, beim autonomen Fahren oder bei Videoüberwachungssystemen – die Auswertung des visuellen Materials spielt eine immer größere Rolle. Viele wissenschaftliche Disziplinen setzen Objekterkennung auf Basis von maschinellem Lernen bereits erfolgreich ein. In Gedächtnisinstitutionen hingegen haben Technologien für die maschinelle inhaltsbasierte Bilderschließung noch kaum Einzug gehalten. Im Rahmen einer Masterthesis<sup>15</sup> wurde für die Vorarlberger Landesbibliothek ein Prototyp entwickelt, mit dem historische Lichtbildwerke unter Einsatz von künstlicher Intelligenz maschinengestützt erschlossen und in weiterer Folge an Metadatenverwaltungssysteme zur Archivierung und Recherche übergeben werden können. Ein besonderes Augenmerk wurde dabei auf die Bildklassifizierung durch Objekterkennung, die Personenidentifizierung durch Gesichtserkennung sowie die optische Zeichenerkennung gelegt. Zudem wurde geprüft, inwieweit das Modell auch trainiert werden kann, um markante landeskundliche Elemente wie beispielsweise Gebäude, Berge oder Landschaften erkennen zu können.

### **2.3.2. Identifizierung von landeskundlich relevanten Personen durch Gesichtserkennung**

Die Identifizierung von landeskundlich relevanten Personen auf historischen Bildern stellt für die Nutzbarkeit und Auffindbarkeit einen enormen Mehrwert dar. Allerdings müssen bei Personenbildern die rechtlichen Rahmenbedingungen genau geprüft werden. So ist jedenfalls der Bildnisschutz gemäß § 78 UrhG der abgebildeten Person zu beachten. Sind Personen abgebildet, so sind die berechtigten Interessen der abgebildeten Person (bei bereits verstorbenen abgebildeten Personen die naher Angehöriger) sowie die berechtigten Interessen der beteiligten Institution gegeneinander abzuwägen. Grundsätzlich bedarf es bei der Veröffentlichung eines Personenbildnisses der Zustimmung der abgebildeten Person bzw. nach deren Tod der Zustimmung der Angehörigen. Unabhängig davon stehen einer Veröffentlichung von Bildnissen von Personen der Öffentlichkeit, etwa Politikern, Schauspielern und Menschen von zeitgeschichtlicher Bedeutung, grundsätzlich keine berechtigten Interessen der Abgebildeten beziehungsweise deren Angehörigen entgegen, sofern die Abbildung nicht im Einzelfall geeignet ist, das Privatleben der Person preiszugeben, sie zu entwürdigen, herabzusetzen oder bloßzustellen.<sup>16</sup>

---

<sup>15</sup> Eberle, H. (2020)

<sup>16</sup> Vgl. § 23 KunstUrhG Abs. 1f.



Ebenfalls stellen Personenbildnisse auch personenbezogene Daten gemäß der Datenschutzgrundverordnung dar. In Bezug auf Abbildungen lebender Personen bedarf es einer Rechtsgrundlage zur Verarbeitung dieser Daten. Als solche kommt grundsätzlich die Einwilligung des Betroffenen in Betracht.<sup>17</sup> Erwägungsgrund Nr. 27 der DSGVO stellt allerdings klar, dass die DSGVO keine Anwendung auf personenbezogene Daten Verstorbener findet. Aufgrund dieser rechtlichen Rahmenbedingungen wurden sowohl die Trainingsdaten als auch die Testdaten sehr sorgsam gewählt. Für die Entwicklung des Prototyps wurde auf die Face-Recognition-Plattform der Microsoft Cognitive Services zurückgegriffen, da sich diese im Besonderen bei der Qualität der Gesichtserkennung von anderen Dienstleistern abhebt.<sup>18</sup> Jede im Erkennungsmodell individualisierte Person kann bis zu 248 Trainingsbilder besitzen. Microsoft empfiehlt zudem, die Gesichter aus den Bildern freizustellen, damit pro hochgeladenem Bild nur ein Gesicht abgebildet ist.<sup>19</sup> Um ein möglichst optimales Erkennungsergebnis zu erreichen, verfügen Trainingsdaten nach Möglichkeit über unterschiedliche Blickwinkel und verschiedene Belichtungen. Zudem sollten die Bilder zumindest eine Auflösung von 200 x 200 Pixel haben. Der Abstand zwischen den Augen sollte mindestens 100 Pixel betragen.<sup>20</sup> Probleme in der Gesichtsdetektion bereiten beispielsweise Aufnahmen mit extremer Über- oder Unterbelichtung, Verdeckungen von zumindest einem Auge, extreme Gesichtsausdrücke oder markante Kopf- oder Gesichtsbehaarung.<sup>21</sup> Um in weiterer Folge eine regelwerkskonforme Personenbeschreibung zu ermöglichen, wird im Modell nicht nur der Name der Person, sondern auch der eindeutige Identifier aus der Gemeinsamen Normdatei (GND) hinterlegt. Ist das Modell trainiert, so kann es validiert und verwendet werden. Im Rahmen erster Tests funktionierte die Erkennung einer landeskundlich relevanten Person immer korrekt und die Erkennungssicherheit schwankte zwischen 60 und 79 Prozent.<sup>22</sup>

### 2.3.3. Bilddatierung mittels Age Prediction

Der internationale Katalogisierungsstandard Ressource Description and Access (RDA) sieht für jede Ressource die Erfassung einer Veröffentlichungsangabe vor. Diese beinhaltet unter anderem das Erscheinungsdatum. Ist für eine Ressource kein Erscheinungsdatum angegeben, wird versucht, das wahrscheinliche Jahr zu

---

17 Vgl. DSGVO Art. 6, Abs. 1, lit. a.

18 Singh, J. (2019)

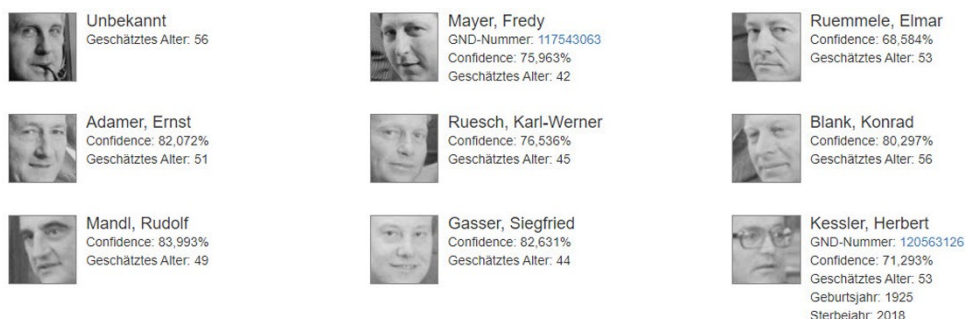
19 <https://docs.microsoft.com/en-us/azure/cognitive-services/face/concepts/face-recognition>

20 <https://docs.microsoft.com/en-us/rest/api/cognitiveservices/>

21 <https://docs.microsoft.com/en-us/azure/cognitive-services/face/concepts/face-recognition>

22 Eberle, H. (2020), Anm. 14, S. 52.

ermitteln.<sup>23</sup> Bei historischen Lichtbildwerken erweist sich eine genaue Datierung oft als schwierig, da sie in den meisten Fällen nicht überliefert ist. Im Rahmen des Prototyps wurde ein Algorithmus entwickelt, um eine automatisierte Schätzung der Datierung zu ermöglichen: ist eine abgebildete Person über einen Normdatensatz wie beispielsweise die GND individualisiert, so ist in vielen Fällen das Geburts- und Sterbedatum bekannt. Diese Daten lassen sich über eine Schnittstelle maschinengestützt abgreifen und verarbeiten. Die Microsoft Cognitive Services bieten neben der Gesichtsdetektion und der Gesichtserkennung auch Methoden, um Attribute wie beispielsweise das Geschlecht, die Emotion oder das Alter zu bestimmen. Durch die Berechnung der Differenz des durch die künstliche Intelligenz berechneten Alters und des Geburtsjahrganges kann das ungefähre Aufnahmejahr des jeweiligen Lichtbildwerkes bestimmt werden. Werden mehrere via GND individualisierte Personen mit vorhandenen Lebensdaten erkannt, so wird für jede Person ein potentielles Aufnahmejahr ermittelt, ein Mittelwert berechnet und als geschätztes Aufnahmejahr ausgegeben.



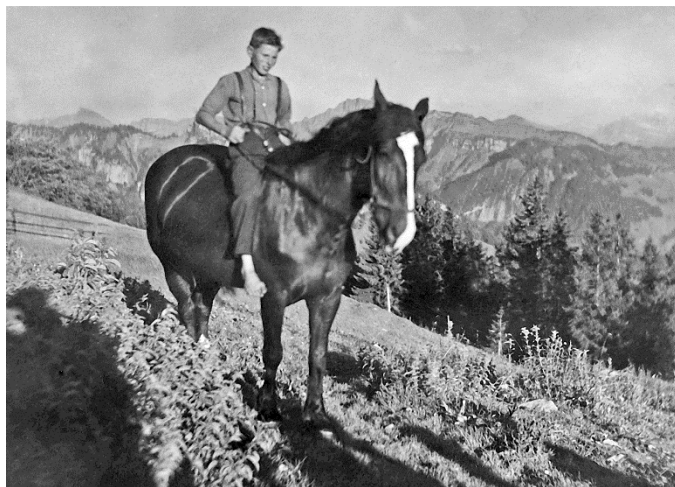
Datierungsversuch über das geschätzte Alter der Personen mit den verfügbaren Lebensdaten aus der GND:  
Das Bild könnte um das Jahr 1978 entstanden sein.

**Abb. 2: Bilddatierung mittels Age Prediction anhand des Bildes**  
<https://pid.volare.vorarlberg.at/o:87126>

<sup>23</sup> Wiesenmüller, H.; Horny, S. (2015), S. 48f.

### 2.3.4. Erkennung von visuellen Merkmalen anhand von maschinelltem Sehen

Das Erkennungsmodell von Microsoft für maschinelles Sehen, welches mit umfangreichen Datensätzen trainiert und laufend weiterentwickelt wird, basiert auf tausenden Objekten, Landschaften und Lebewesen.<sup>24</sup> Auf Basis dieses Modells werden erweiterte Algorithmen angeboten, welche Daten verarbeiten und daraus generierte Informationen zurückgeben können. Für die verbale Beschreibung von Bildern sind die sogenannten freien Schlagwörter von Interesse. Dabei sind diese weder innerhalb von Normdaten organisiert, noch in hierarchische Strukturen eingeordnet. Dennoch kann eine solche Auflistung von Schlagwörtern die Grundlage für eine verbale Bildbeschreibung bilden und ist deshalb wertvoll für die Erschließung von Bildmaterial.<sup>25</sup>



#### Tags

Outdoor (Confidence: 99,9%)  
 Gras (Confidence: 99,45%)  
 Berg (Confidence: 99,28%)  
 Pferd (Confidence: 97,78%)  
 Feld (Confidence: 80,71%)  
 Tier (Confidence: 80,61%)  
 Weiß (Confidence: 74,28%)  
 Säugetier (Confidence: 74,22%)  
 Hügel (Confidence: 61,44%)

**Abb. 3: Vergabe freier Schlagwörter am Beispiel des Objektes**

**<https://pid.volare.vorarlberg.at/o:113467>.**

**Foto: Vorarlberger Landesbibliothek, Oliver Benvenuti**

<sup>24</sup> Salvaris, M. (2019), S. 109f.

<sup>25</sup> <https://docs.microsoft.com/de-de/azure/cognitive-services/computer-vision/concept-tagging-images>

Einschränkend muss erwähnt werden, dass das zugrundeliegende Modell von Microsoft anhand von aktuellen Bildern trainiert wird. Aus diesem Grund liefert diese Methode besonders bei der Erkennung von historischen Objekten wie beispielsweise historischen Fahrzeugen oder historischen Gebrauchsgegenständen nur bedingt korrekte oder gar keine Ergebnisse.

### **2.3.5. Erkennung und Beschreibung markanter landeskundlicher Objekte**

Im Rahmen der Erstellung des Prototyps für die Vorarlberger Landesbibliothek wurde ebenfalls mit der Erstellung eines domänenspezifischen Modells experimentiert, um markante landeskundliche Elemente wie Gebäude, Berge oder sonstige relevante Orte zu erkennen. Microsoft empfiehlt, solche Modelle mit mindestens 250 Bildern pro Objekt zu trainieren.<sup>26</sup> Im historischen Kontext stellt dies ein großes Problem dar, da nur in den wenigsten Fällen so viele Aufnahmen in ausreichender Qualität überliefert wurden. Zudem werden neben den Trainingsdaten auch Daten zur Validierung benötigt. So wurde das Modell im Experiment lediglich mit sechs landeskundlichen Objekten trainiert. Die im Rahmen der Validierung durchgeführten Tests lieferten eine Erkennungsgenauigkeit zwischen 53 und 99 Prozent.<sup>27</sup> Aufgrund des unausgewogenen Datenbestandes kam es in vielen Fällen zu einer sogenannten Überanpassung. Das führte dazu, dass für Objekte, die im Modell nicht trainiert worden sind, das ähnlichste Objekt als wahrscheinlichste Klassifizierung ermittelt wurde. Aus diesen Gründen und weil der Aufwand in keinem Verhältnis zum erwarteten Nutzen steht, wurde die Idee eines domänenspezifischen Erkennungsmodells zur Erschließung landeskundlicher Elemente wieder verworfen.

---

<sup>26</sup> <https://docs.microsoft.com/de-de/azure/cognitive-services/Custom-Vision-Service/overview>

<sup>27</sup> Eberle, H. (2020), S. 65.

### 3. Fazit

Seit vielen Jahren werden in Gedächtnisinstitutionen immer mehr Retrodigitalisierungsprojekte in Angriff genommen, um analoge Fotos digital zu sichern und nutzerfreundlich vermitteln zu können. Mit den wachsenden Datenmengen steigen aber auch die Anforderungen durch die Standardisierung und Normierung in der deskriptiven Bildbeschreibung. Gleichzeitig fehlt es oft an personellen und finanziellen Ressourcen, um diesen Anforderungen in Qualität und Quantität gerecht werden zu können. So erfolgt die Generierung von Metadaten in vielen Institutionen nach wie vor manuell durch eine verbale Beschreibung des Bildinhaltes. Für die Optimierung dieser Arbeitsabläufe bieten sich neue Methoden an, um das vorhandene Fachpersonal effizient einsetzen zu können. Sowohl die Georeferenzierung über die Plattform sMapshot als auch die Bildbeschreibung durch eine Künstliche Intelligenz könnten Institutionen bei der Erfüllung ihres Sammelauftrages unterstützen und einen Mehrwert bieten, da viele dieser maschinell generierten Informationen bei einer manuellen verbalen Beschreibung gar nicht erfasst werden würden. Dennoch können all diese Methoden nicht die Expertise und das Engagement der einzelnen Mitarbeitenden ersetzen, die für die Umsetzung eines erfolgreichen Erschließungsprojektes sowie das laufende Motivieren der Crowd essentiell sind. Da es zum aktuellen Zeitpunkt noch keine anderen Möglichkeiten gibt, müssen zudem alle Beiträge manuell kontrolliert und validiert werden. Die Nutzung künstlicher Intelligenz zur Bildbeschreibung kann das Personal entlasten, aber nicht vollständig ersetzen. Auch Erkennungsmodelle müssen trainiert, validiert und kontrolliert werden. Die Einordnung eines jeden Bildes in den historischen Kontext benötigt nach wie vor das Spezialwissen im jeweiligen Fachgebiet. Letztlich sind bei allen automatisierten Bildbeschreibungsmethoden nur begrenzte Erfolge zu erwarten, da die Interpretation eines Bildes häufig einen großen Spielraum zulässt.

### Bibliografie

- Eberle, Harald (2020): Inhaltsbasierte Bilderschließung mittels visueller Merkmale durch Maschinelles Sehen. Am Beispiel der Microsoft Azure Cognitive Services und den Daten des Vorarlberger Landesrepositoriums. Masterarbeit, FH Wien der WKW.
- Graf, Nicole (2018): sMapshot. Die Crowd lokalisiert Bilder im virtuellen Globus. In: Emenlauer-Blömers, Eva et. al. (Hg.): Konferenzband EVA Berlin 2018. Elektronische Medien & Kunst, Kultur und Historie: 25. Berliner Veranstaltung der internationalen EVA-Serie Electronic Media and Visual Arts. Berlin: Staatliche Museen.
- Graf, Nicole (2018): sMapshot ist lanciert. [blogs.ethz.ch](https://blogs.ethz.ch), 07.02.2018. <https://doi.org/10.35016/ethz-cs-4995-de>

- Helm, Wiebke; Mandl, Thomas; Putjenter, Sigrun; Schmideler, Sebastian; Zellhöfer, David (2019): Distant Viewing-Forschung mit digitalisierten Kinderbüchern. Voraussetzungen, Herausforderungen und Ansätze. In: b.i.t. online 22.
- Huonder, Flurina (2020): sMapshot Anpassungen für Vorarlberg. blogs.ethz.ch, 22.06.2020. <https://doi.org/10.35016/ethz-cs-12203-de>
- Produit, Timothée; Ingensand, Jens (2016): A 3D Georeferencer and Viewer to Relate Landscape Pictures with VGI. In: Ali Mansourian, Ali et al. (eds.): AGILE International Conference on Geographic Information, LINK-VGI Workshop.
- Produit, Timothée; Ingensand, Jens (2018): 3D Georeferencing of Historical Photos by Volunteers. In: Geospatial Technologies for All. Selected Papers of the 21st AGILE Conference on Geographic Information Science. Cham: Springer.
- Produit, Timothée; Ingensand, Jens (2019): smapshot. Georeferenzierung historischer Landschaftsbilder durch Crowdsourcing. In: Geomatik Schweiz 117 (9).
- Salvaris, Mathew (2019): Deep Learning mit Microsoft Azure. Bonn: Rheinwerk Verlag.
- Singh, Jake (2019): A Comparison of Public Cloud Computer Vision Service. Santa Clara University, 12.09.2019. <https://osf.io/9t5qf/> (abgerufen am 25.08.2021)
- sMapshot. Die partizipative Zeitmaschine. vorarlberg.orf.at. <https://vorarlberg.orf.at/stories/3115329/> (abgerufen am 17.08.2021)
- Volmer, Stephan (2012): Inhaltsbasierte Bildsuche mittels visueller Merkmale. Eine Alternative zur Erschließung digitaler Bildinformation. 2. Aufl. Saarbrücken: AV Akademikerverlag.
- Wiesenmüller Heidrun; Horny, Silke (2015): Basiswissen RDA. Eine Einführung für deutschsprachige Anwender. Berlin: De Gruyter Saur.

**Harald Eberle** ist stellvertretender Leiter der Abteilung „EDV und Katalog“ an der Vorarlberger Landesbibliothek. Seit 2014 leitet er das Projekt „volare – Vorarlberger Landesrepositorium“ und ist in der Vorarlberger Landesbibliothek als technischer Experte für die Digitalisierung und die Bereitstellung digitaler Sammlungen verantwortlich.