

3D Reconstruction

Furtwanger Medienstudierendentagung
2026

Prof. Dr.-Ing. Uwe Hahne

Vorstellung



- Uwe Hahne
 - Mail: uwe.hahne@hs-furtwangen.de
 - <https://uhahne.github.io/>
- 2000 – 2006: Studium der Mediensysteme an der Bauhaus Universität Weimar
- 2006 – 2012: Promotion an der TU Berlin im Bereich Computergrafik
- 2012 – 2021: SW-Entwicklung und Gruppenleitung 3D Kameras bei SICK in Waldkirch
- Seit 2021 hier an der HFU
- Interessen: 3D Computer Vision, Agilität, Medien, ...

Wer seid ihr?

- Kurze Vorstellungsrunde
 - Name, Studiengang und –ort
 - Erwartungen oder Vorerfahrung mit 3D Rekonstruktion

Plan für heute

- Vorstellungsrunde
- Kurzvortrag zur Geschichte der 3D-Rekonstruktion
- Praktischer Teil mit gemeinsamer 3D-Rekonstruktion
- Jederzeit: Fragen und Antworten sowie Diskussion

Was ist Computer Vision?

... und warum 3D CV?

Computer Vision

- Einfache Fragen:
 - Was ist hier zu sehen?



[i:colo07]

Computer Vision

- Einfache Fragen:
 - Was ist hier zu sehen?
 - Wie sieht das Gebäude von oben/hinten aus?

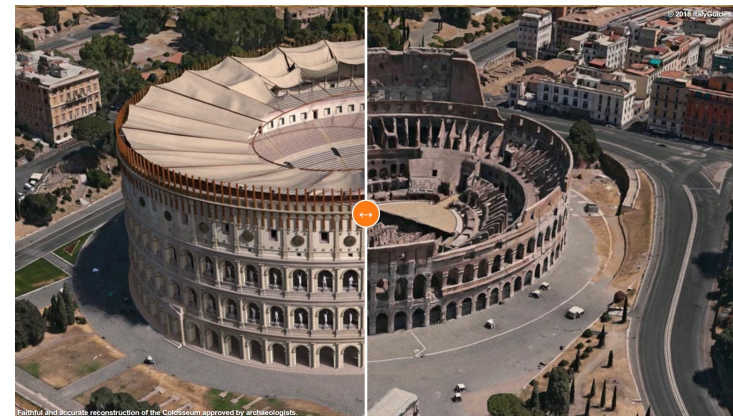
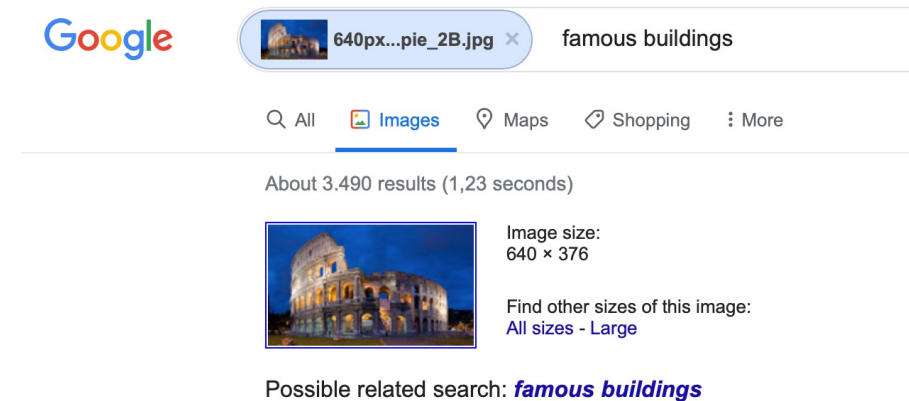


[i:colo07]

Computer Vision

- Einfache Fragen:
 - Was ist hier zu sehen?
 - Wie sieht das Gebäude von oben/hinten aus?

- → Schwierig für einen Computer zu beantworten.



<https://www.italyguides.it/en/lazio/rome/ancient-rome/colosseum>

Computer Vision

- Einfache Fragen:

- Was ist hier zu sehen?

Update from 2023:



einen Computer zu beantworten.

Computer Vision

- Einfache Fragen:


- Was ist hier zu sehen?

Update from 2025:

open/

- → Schwere Fragen:
• einen Code generieren
• beantworten

ChatGPT ▾ ↑ Gemeinsam nutzen UW



Was ist hier zu sehen?

Das Bild zeigt das Kolosseum in Rom, Italien. Es handelt sich um ein antikes Amphitheater, das während der römischen Kaiserzeit erbaut wurde. Es ist eines der bekanntesten Wahrzeichen Roms und ein bedeutendes architektonisches und historisches Symbol des Römischen Reiches. Das Kolosseum wurde für Gladiatorenkämpfe, Tierhetzen und öffentliche Spektakel genutzt und ist heute eine der meistbesuchten Touristenattraktionen der Welt.

📄 🌟 🗣️ 🔊 🎨 🔄



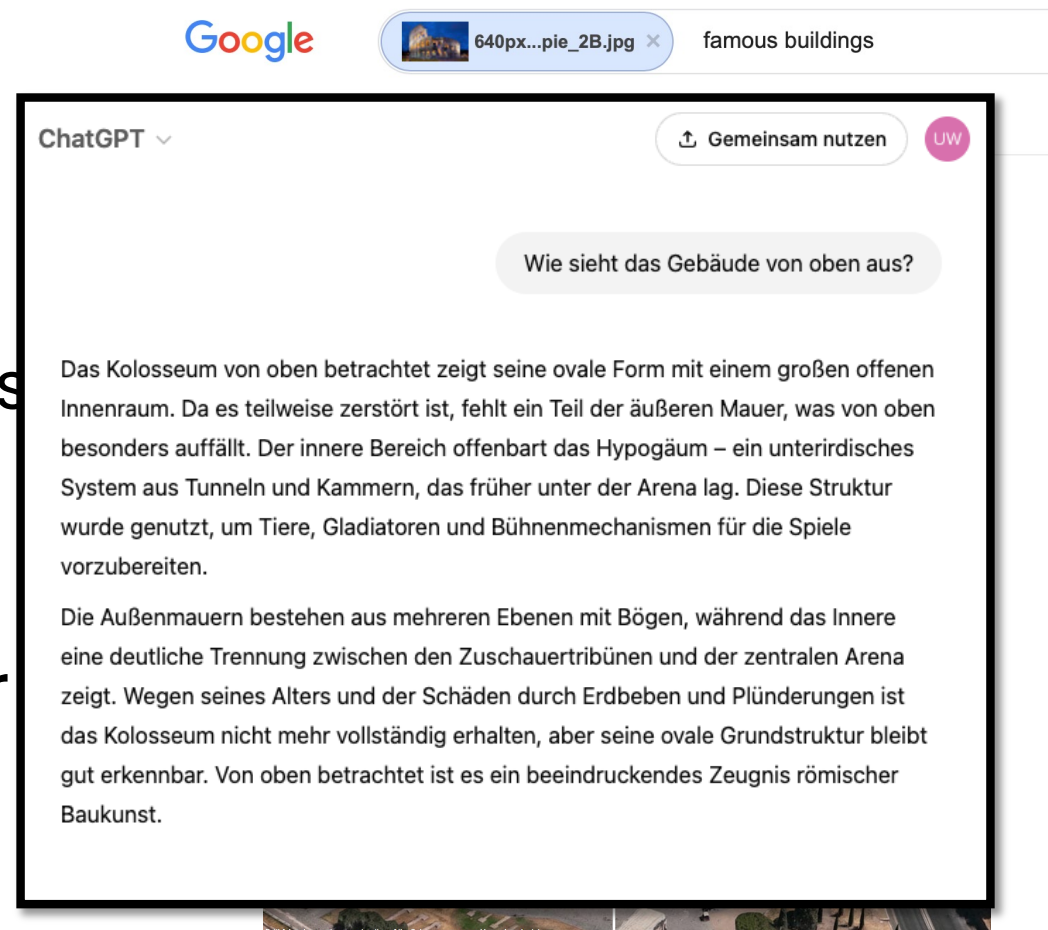
Computer Vision

- Einfache Fragen:
 - Was ist hier zu sehen?

Update from 2025:

Wie sieht das Gebäude von oben/ninten aus?

- → Schwierig für einen Computer beantworten.



Google 640px...pie_2B.jpg × famous buildings

ChatGPT ↑ Gemeinsam nutzen UW

Wie sieht das Gebäude von oben aus?

Das Kolosseum von oben betrachtet zeigt seine ovale Form mit einem großen offenen Innenraum. Da es teilweise zerstört ist, fehlt ein Teil der äußeren Mauer, was von oben besonders auffällt. Der innere Bereich offenbart das Hypogäum – ein unterirdisches System aus Tunneln und Kammern, das früher unter der Arena lag. Diese Struktur wurde genutzt, um Tiere, Gladiatoren und Bühnenmechanismen für die Spiele vorzubereiten.

Die Außenmauern bestehen aus mehreren Ebenen mit Bögen, während das Innere eine deutliche Trennung zwischen den Zuschauertribünen und der zentralen Arena zeigt. Wegen seines Alters und der Schäden durch Erdbeben und Plünderungen ist das Kolosseum nicht mehr vollständig erhalten, aber seine ovale Grundstruktur bleibt gut erkennbar. Von oben betrachtet ist es ein beeindruckendes Zeugnis römischer Baukunst.

Faithful and accurate reconstruction of the Colosseum approved by archaeologists.

Computer Vision

- Einfache Fragen:
 - Was ist hier zu sehen?


Update from 2025:

open/ninten aus

- → Schwierig für einen Computer beantworten.

ChatGPT ↑ Gemeinsam nutzen UW

Kannst Du mir eine Bild davon generieren?



Hier ist eine generierte Ansicht des Kolosseums aus der Vogelperspektive. Sie zeigt die ovale Struktur mit den teilweise zerstörten Außenwänden sowie das Hypogäum im Inneren. Die umliegenden Straßen und Grünflächen sind ebenfalls zu sehen. Lass mich wissen, falls du Änderungen möchtest!

Computer Vision

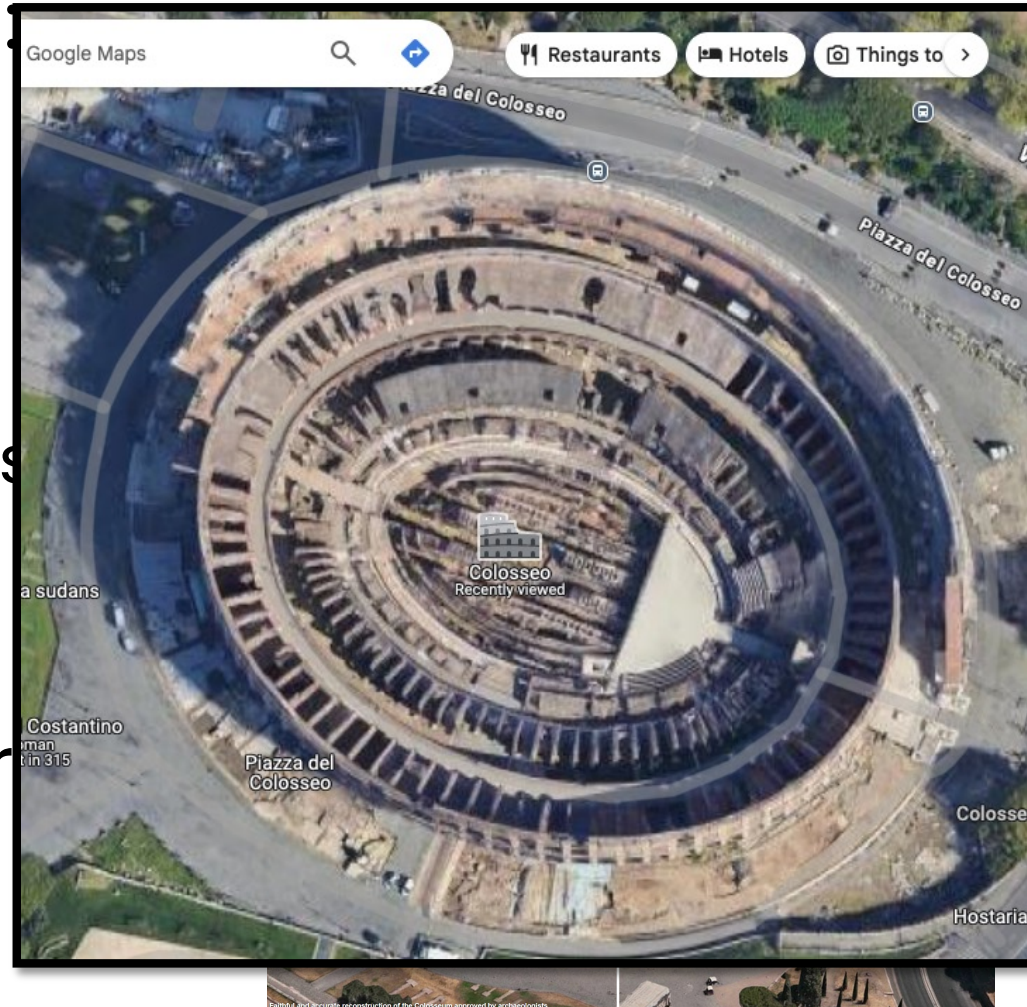
- Einfache Fragen

- Was ist hier zu sehen?

Update from 2025:

open/ninten aus

- → Schwierig für einen Computer beantworten.



Computer Vision vs Computer Graphics

Images



Vision



Model



Graphics



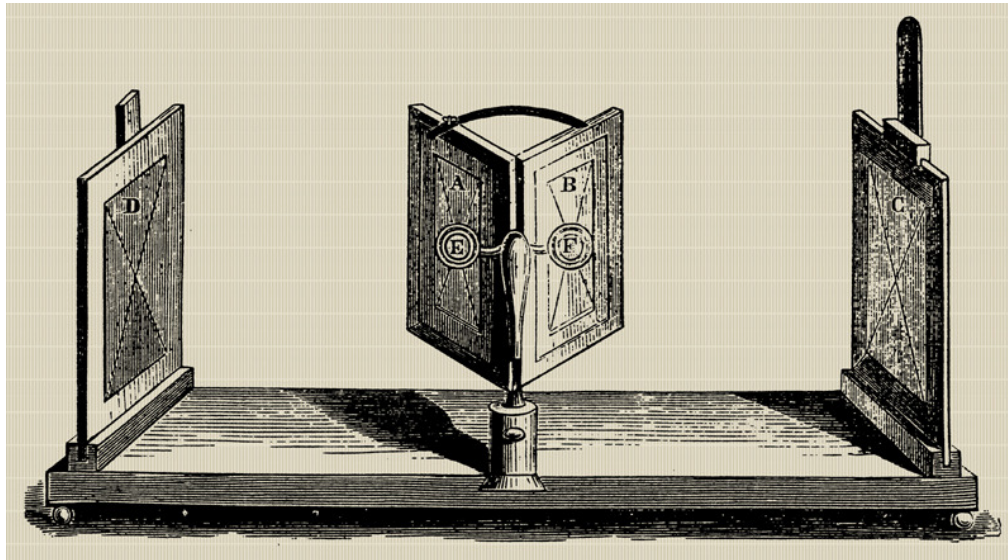
Slide is from <http://vision.cs.utexas.edu/376-spring2018/>

Inspiration: [Jared Heinly - The Evolution of Image Based 3D Reconstruction](#)

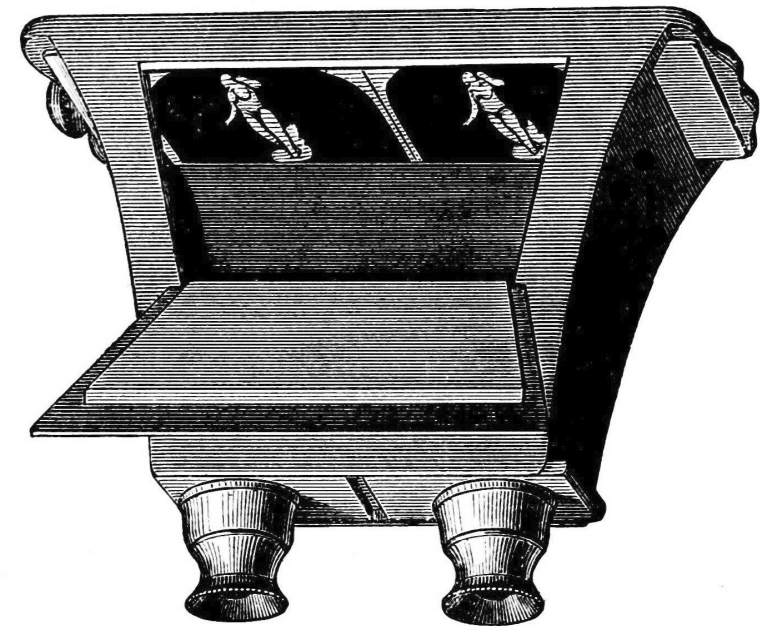
Historische Entwicklung

...wie machte man früher 3D Rekonstruktion und wie heute?

Stereoskopie



Wheatstone (1838)

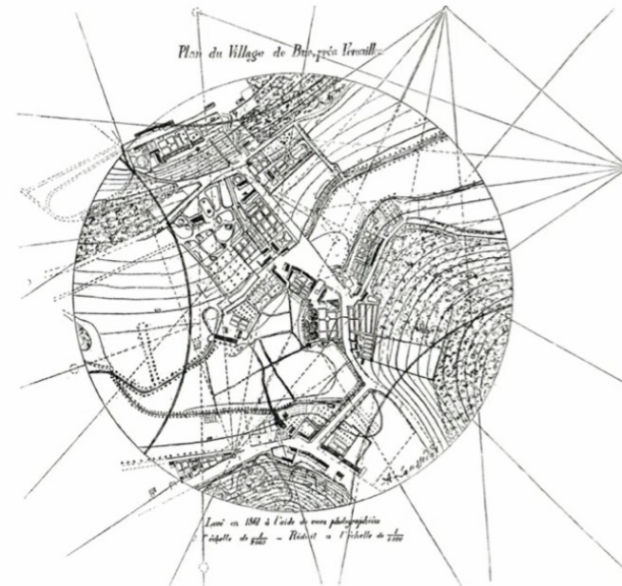


Brewster (1849)

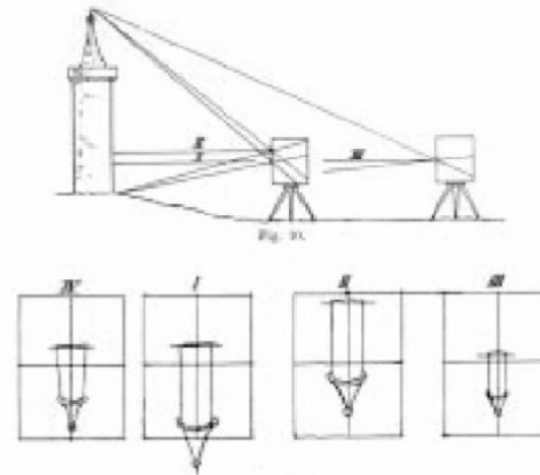
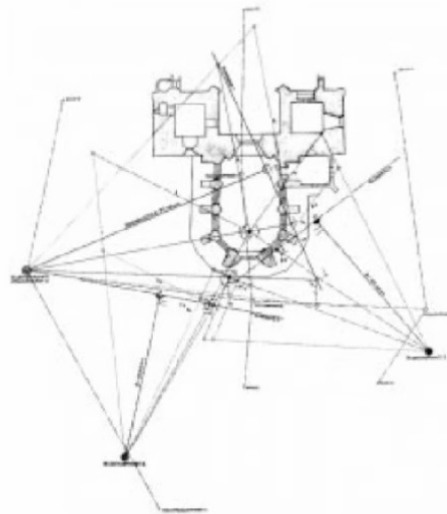
Métrophotographie



Aimé Laussedat, 1860s



Photogrammetrie



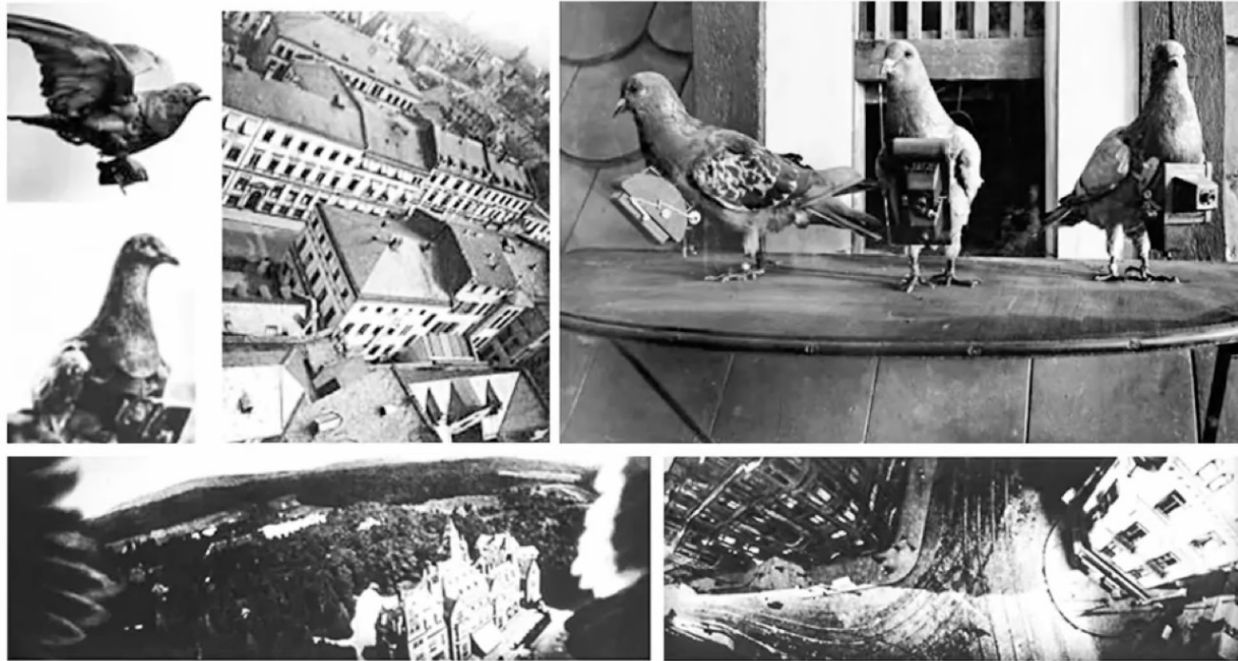
Albrecht Meydenbauer, 1860s

Aerial photography



1920s

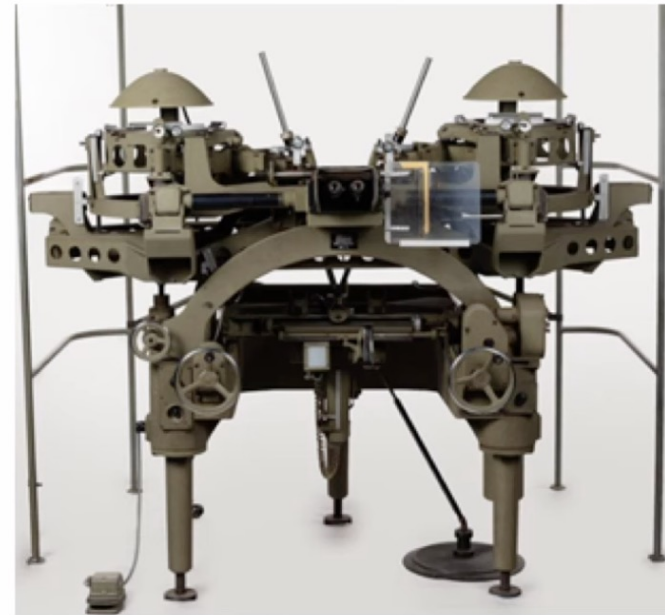
Aerial photography



1910s

Kein Witz: <https://de.wikipedia.org/wiki/Brieftaubenfotografie>

Stereoplotters (analog)



1930s

<https://en.wikipedia.org/wiki/Stereoplotter>

Stereoplotter (analytical)



1960s

<https://en.wikipedia.org/wiki/Stereoplotter>

Bundle adjustment

M. I. A. Lourakis and A. A. Argyros

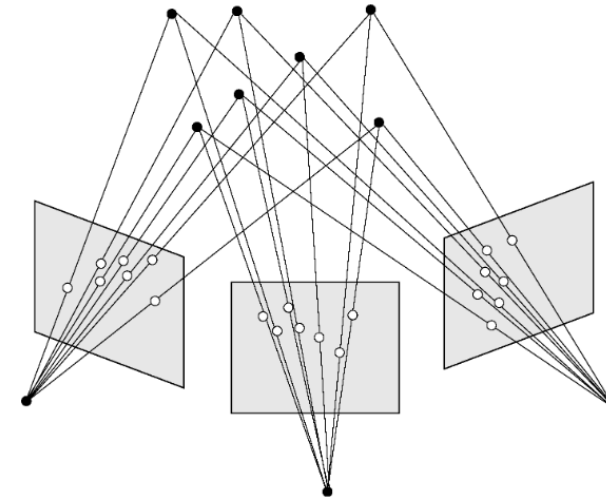
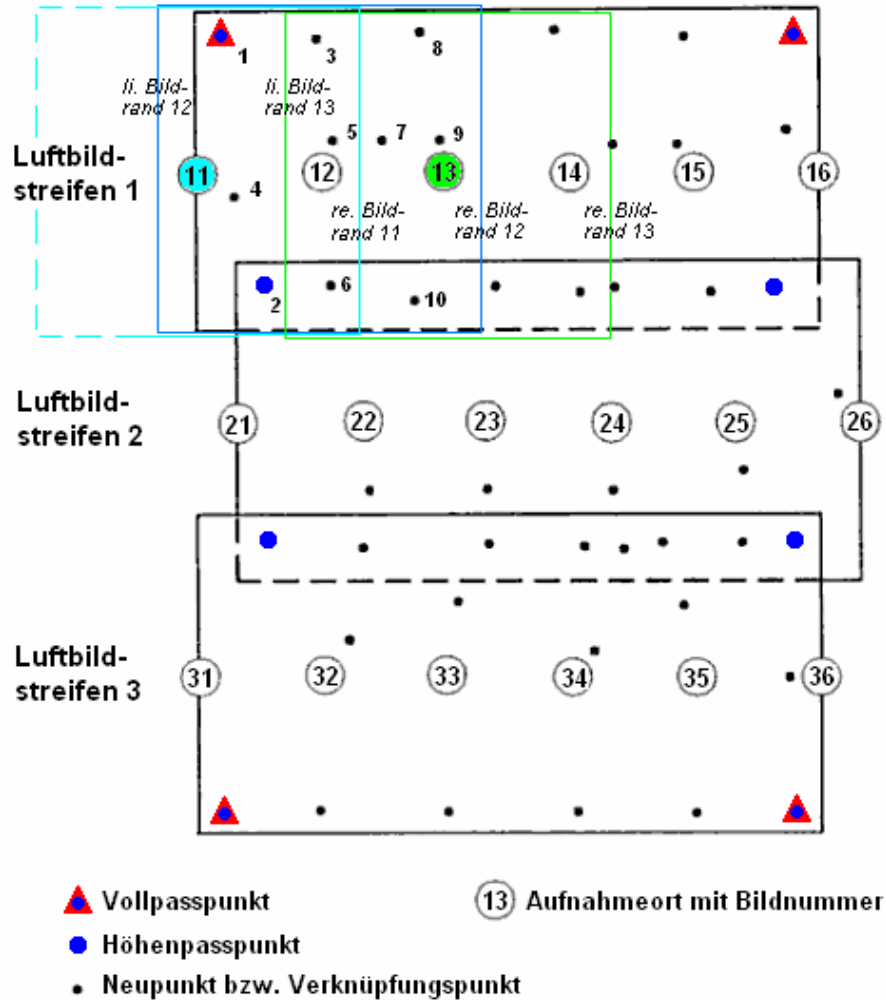
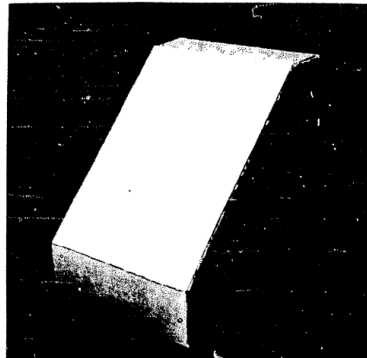


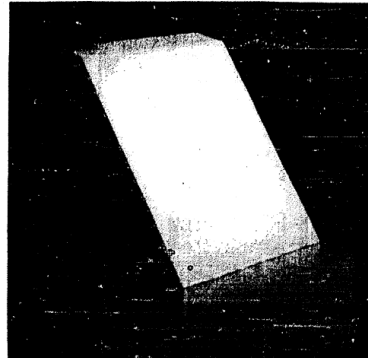
Fig. 3. Schematic illustration of $n = 7$ points projecting on $m = 3$ images.

$$\min_{\mathbf{a}_j, \mathbf{b}_i} \sum_{i=1}^n \sum_{j=1}^m v_{ij} d(\mathbf{Q}(\mathbf{a}_j, \mathbf{b}_i), \mathbf{x}_{ij})^2$$

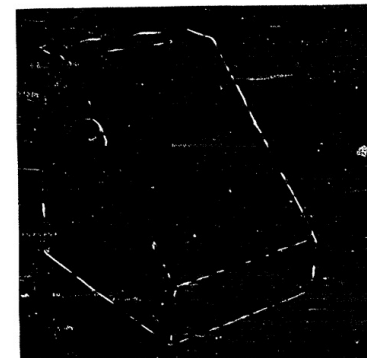
Frühe Computer Vision



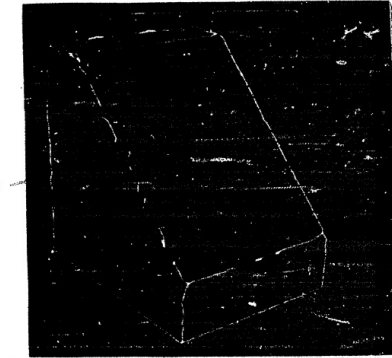
A. Original Picture



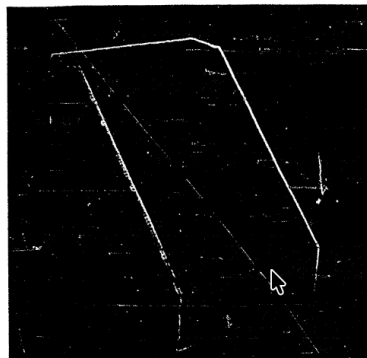
B. Computer Display of Picture
(Reflected by mistake)



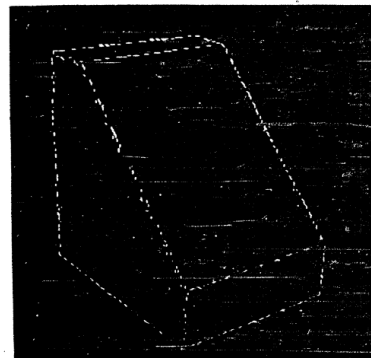
E. Connected Feature Points



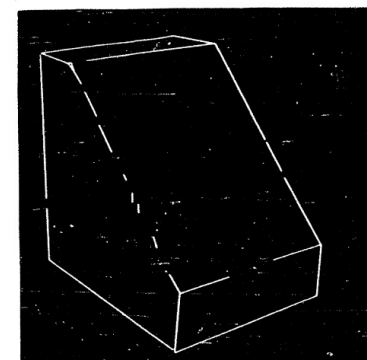
F. After Complexity Reduction



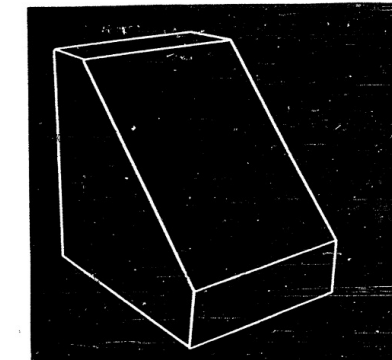
C. Differentiated Picture



D. Feature Points Selected



G. After Initial Line Fitting

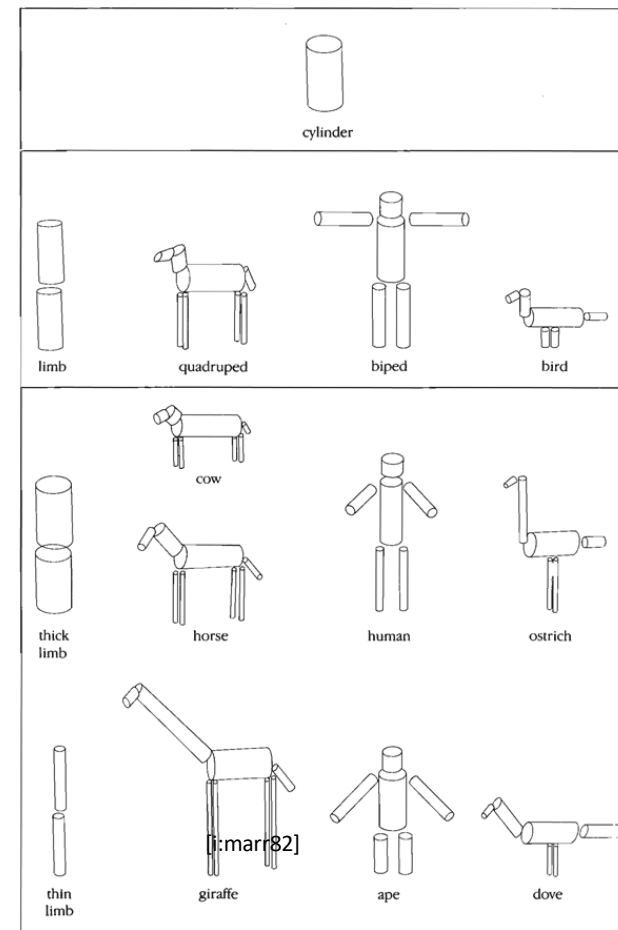


H. Final Line Drawing

aus Larry Roberts – *Machine Perception of 3D solids* – PhD thesis – 1963 - MIT

Computer Vision

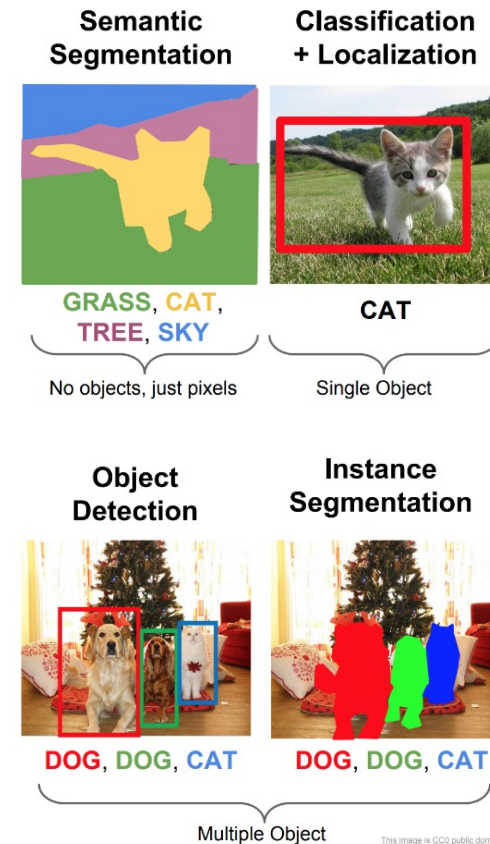
- Ein früherer Ansatz aus den 80er Jahren:
- **Wenn man die Objekte geometrisch rekonstruiert, kann man sie erkennen.**



[i:marr82]

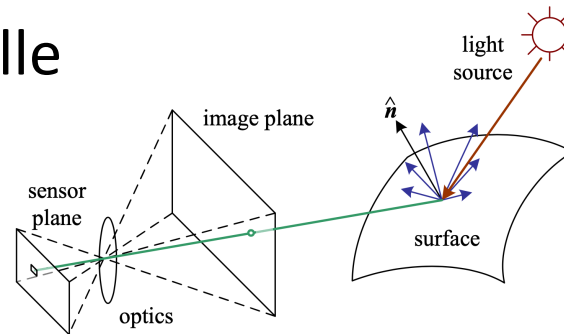
Computer Vision

- Begriffe
 - „Reconstruction“ → welche Geometrie hat das Objekt?
 - „Recognition“ → was ist das für ein Objekt?
 - „Detection“ → wo (im Bild) befindet sich das Objekt?
 - „Object“ vs „Scene“ → Geht es um einzelne Objekte oder um die gesamte Umgebung?

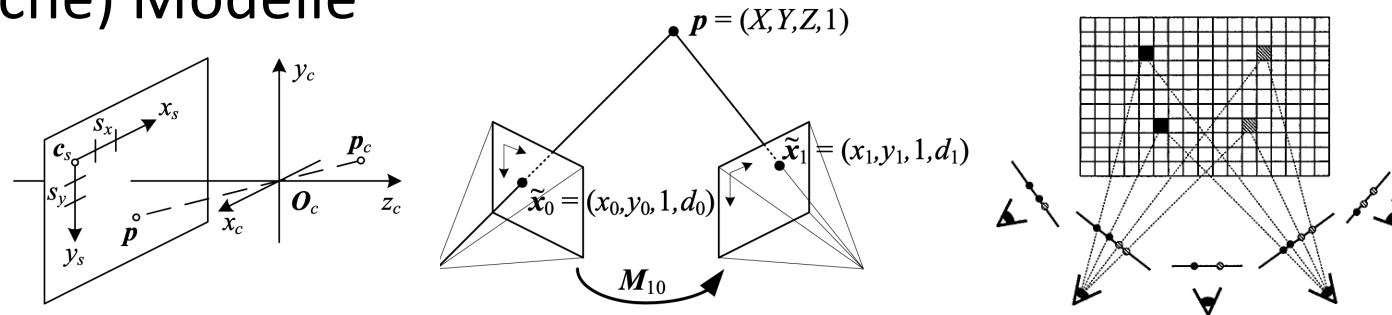
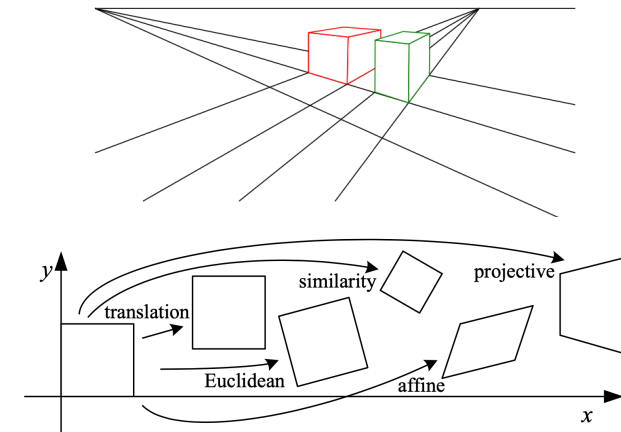


Klassischer Ansatz

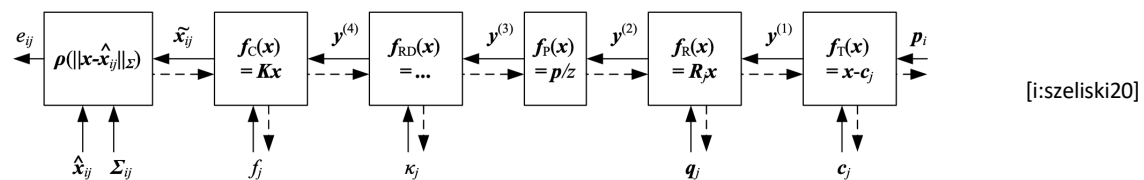
- Physikalische Modelle



- Mathematische (geometrische) Modelle

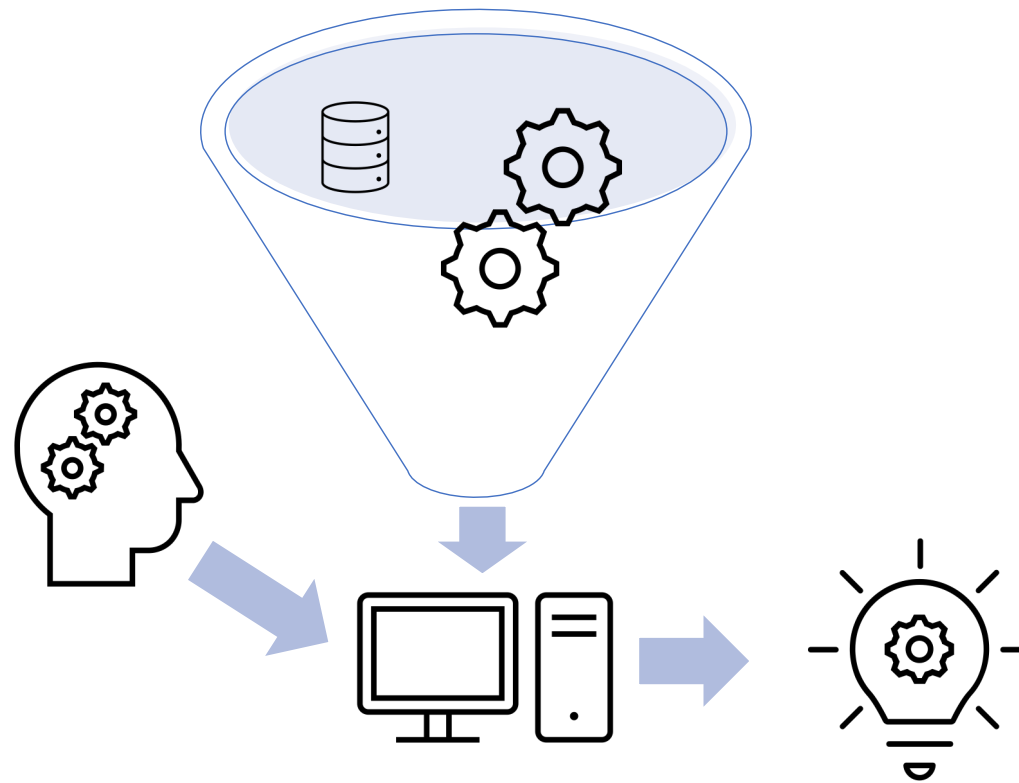


- Komplexe, aber nachvollziehbare Prozesse

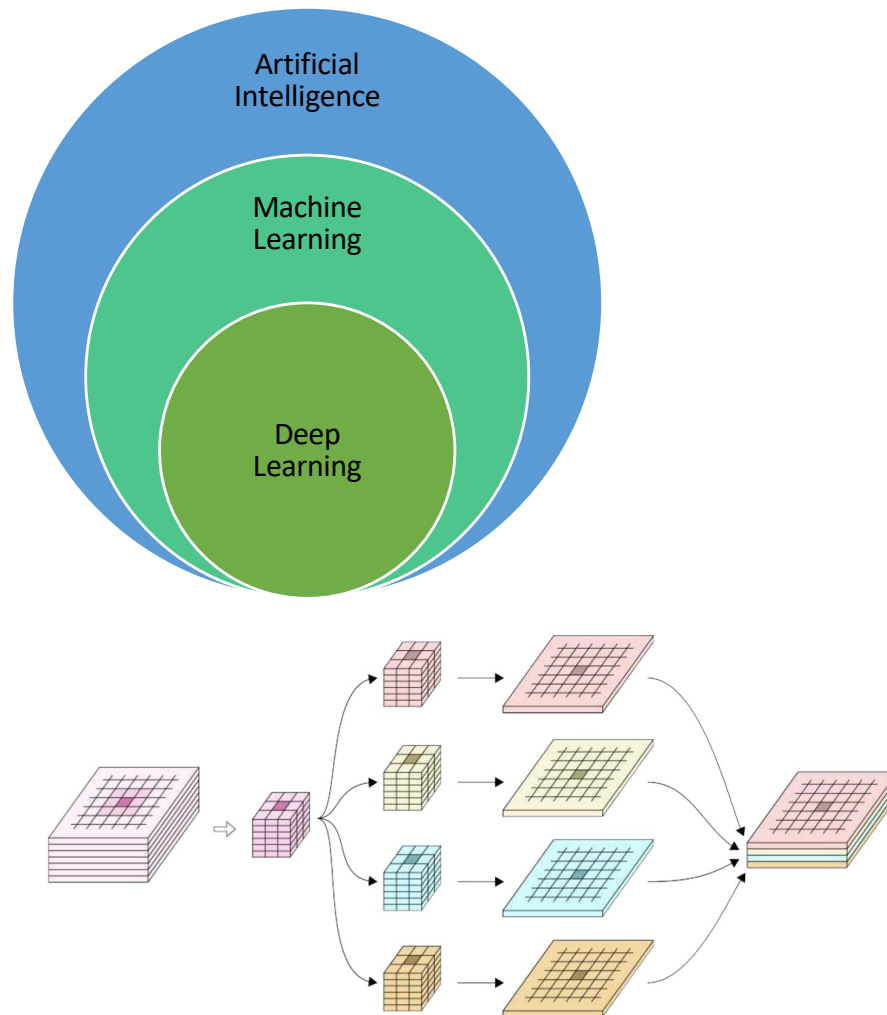


[i:szeliski20]

Klassische Computer Vision

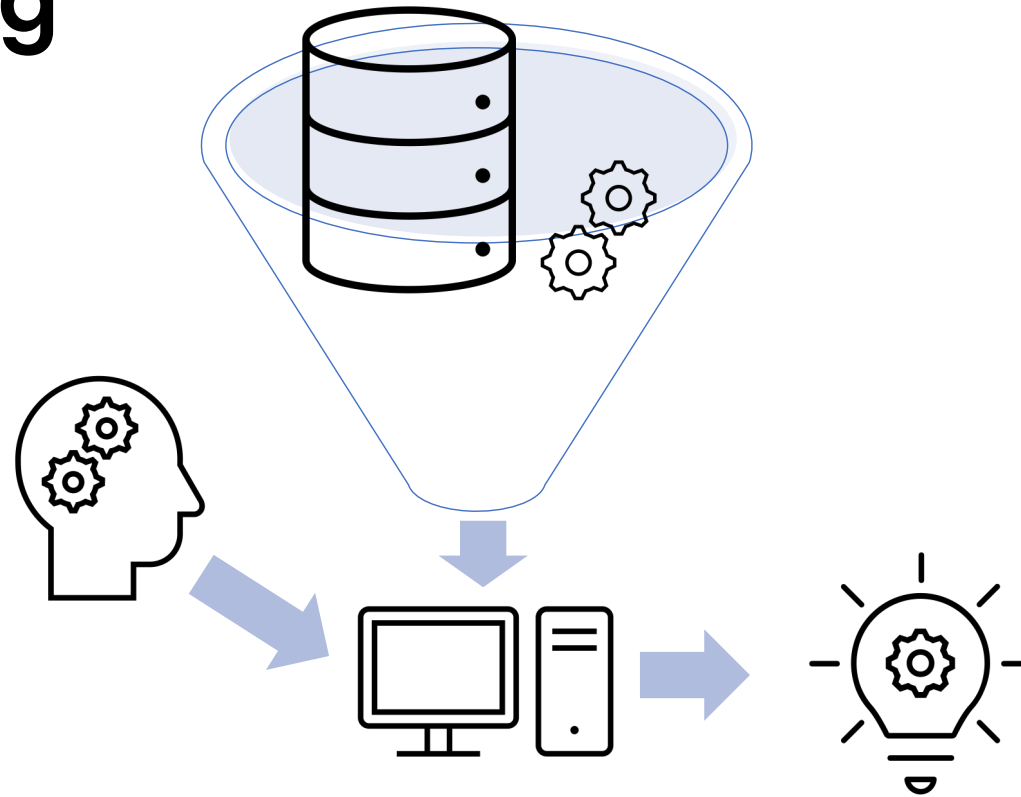


Deep Learning

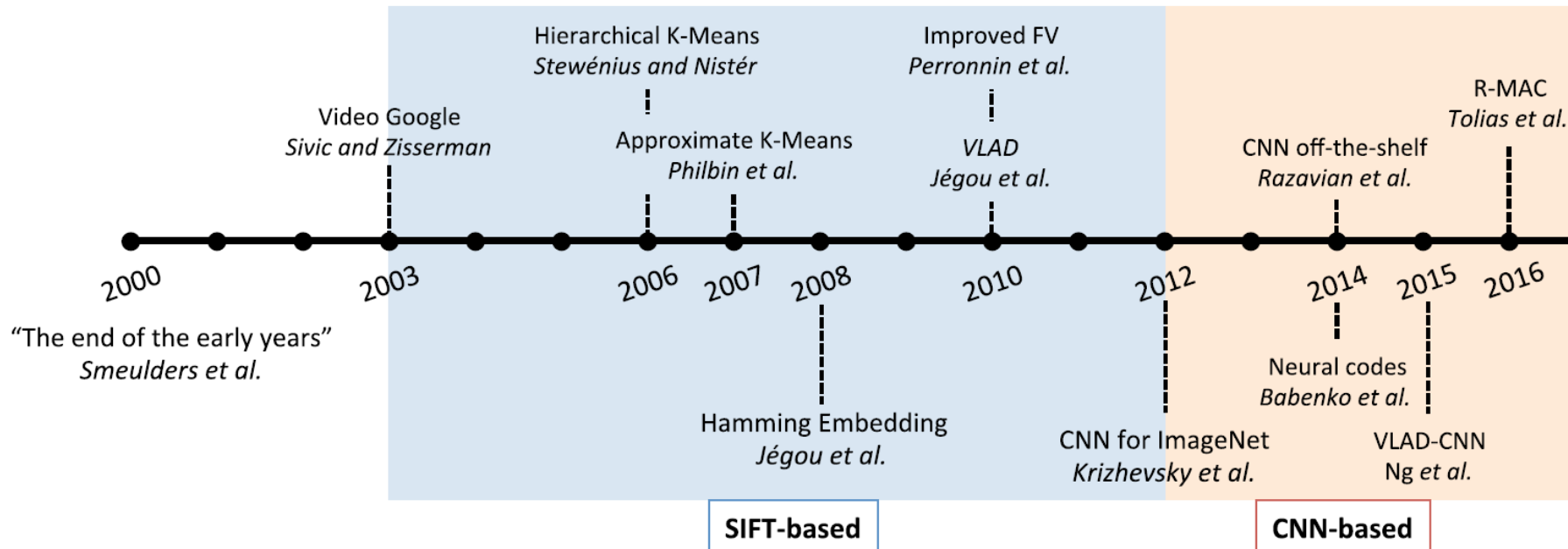


- Künstliche Intelligenz (Artificial Intelligence) als Überbegriff für wissensbasierte Entscheidungsfähigkeit
- Maschinelles Lernen (Machine Learning) als Gruppe von Algorithmen, die ein Modell aus Trainingsdaten erstellen.
 - Überwachtes Lernen (Supervised learning)
 - Unüberwachtes Lernen (Unsupervised learning)
 - Bestärkendes Lernen (Reinforcement learning)
- Deep Learning als Untergruppe, die künstliche Neuronale Netze (Artificial Neural Network) als Modell verwendet. [i:szeliski20]
 - Faltendes Neuronales Netz (convolutional neural network)

Deep Learning



Übergang von klassischer Computer Vision zu Deep Learning



[i:szeliski20]

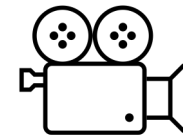
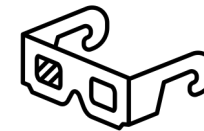
Methoden

Welche Ansätze es zur 3D Rekonstruktion gibt...

3D Rekonstruktion

- Aus Kamerabildern (passiv)

- Monokulare Tiefenkriterien
- Stereo
- Multi-View Stereo
- Structure from Motion
- ...



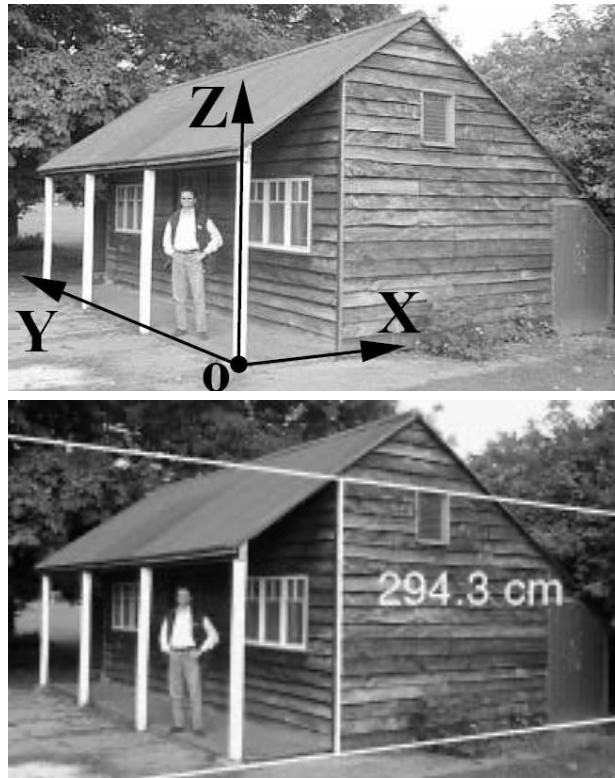
- Mit speziellen Sensoren (aktiv)

- Structured Light
- Time of Flight
- Laser-Triangulation
- ...



Aus einem einzelnen Bild

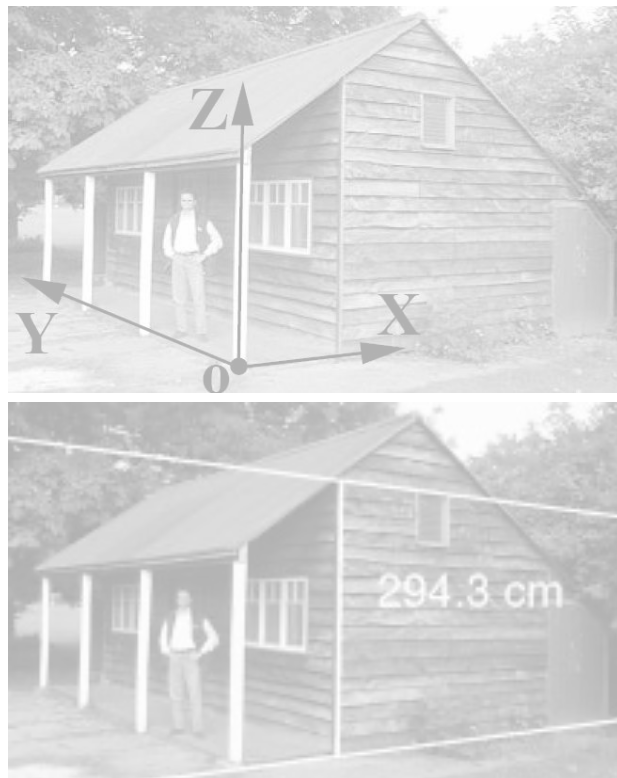
- Klassische Lösung



[i:criminsi99]

Aus einem einzelnen Bild

- Klassische Lösung
- Deep Learning (2020)

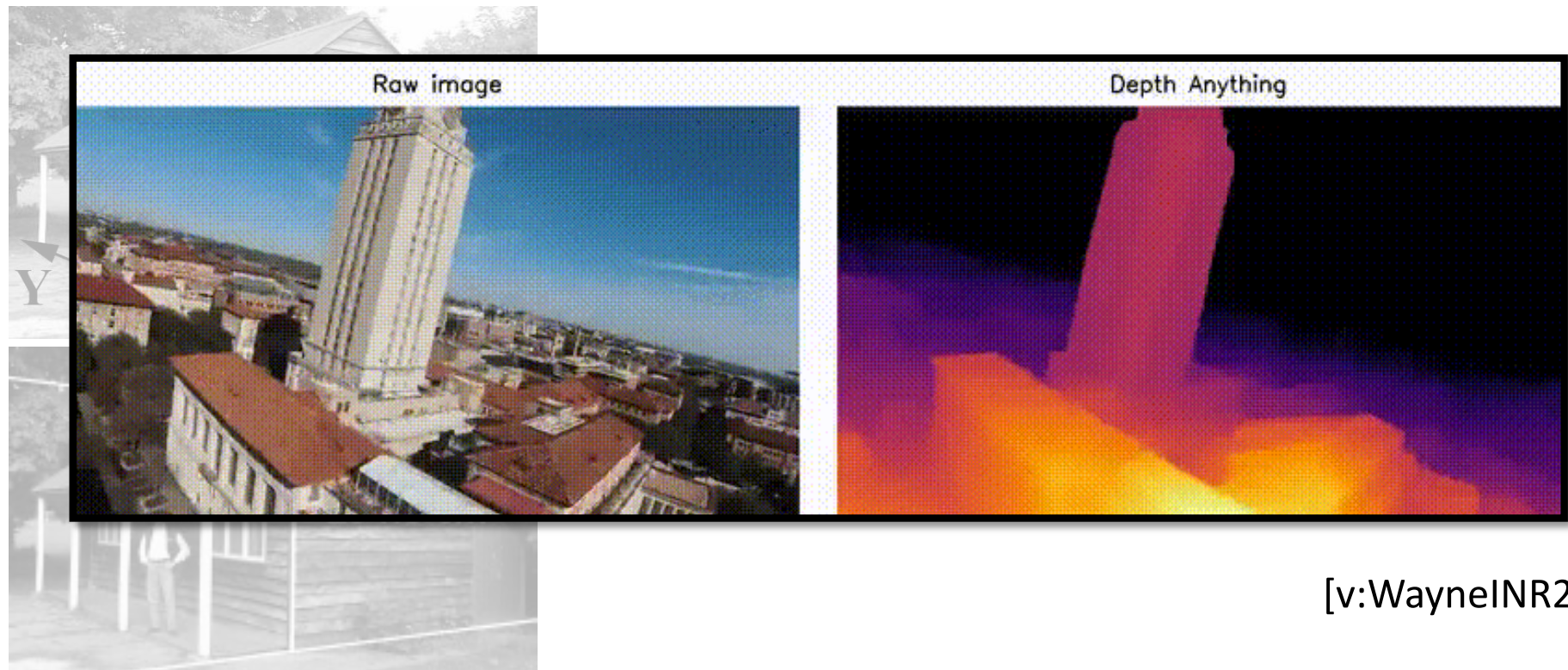


[v:facebook20]

[v:facebook20]

Aus einem einzelnen Bild

- Klassische Lösung
- Deep Learning (2024)

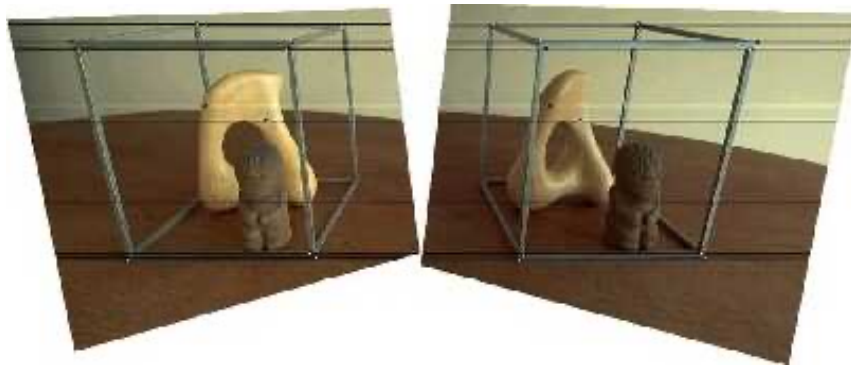


[v:WayneINR24]

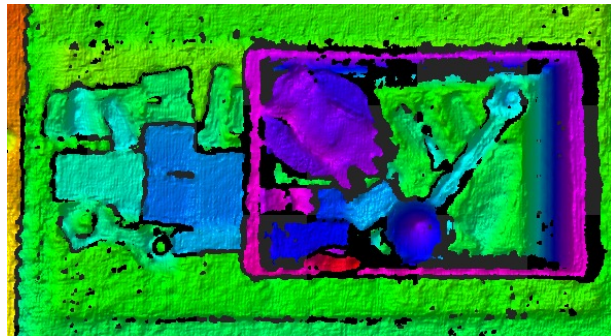
<https://depth-anything.github.io/> (2024)

Aus zwei Bildern

- Klassische Stereo Vision

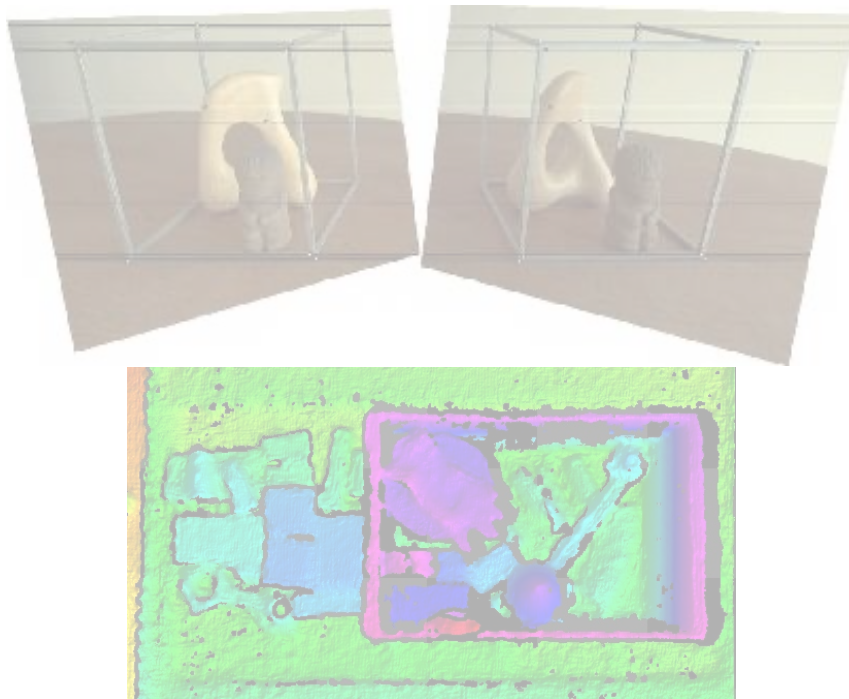


[i:szeliski20]



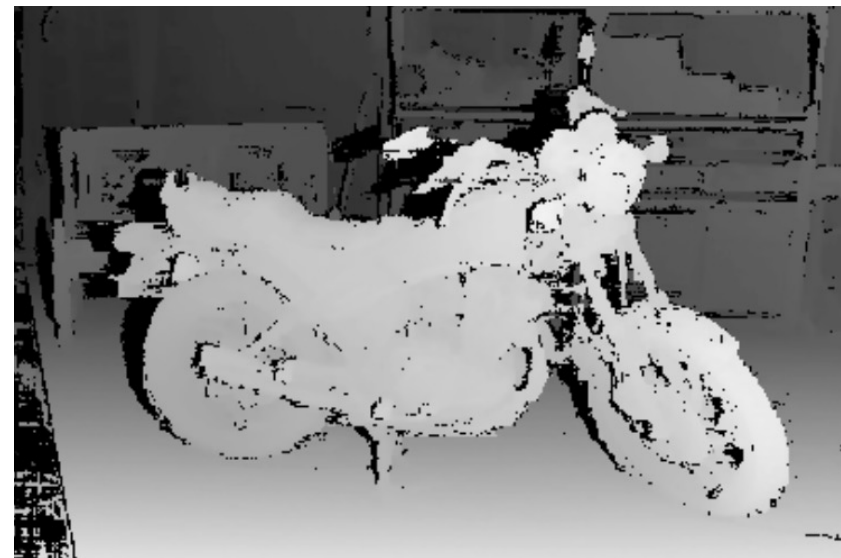
Aus zwei Bildern

- Klassische Stereo Vision



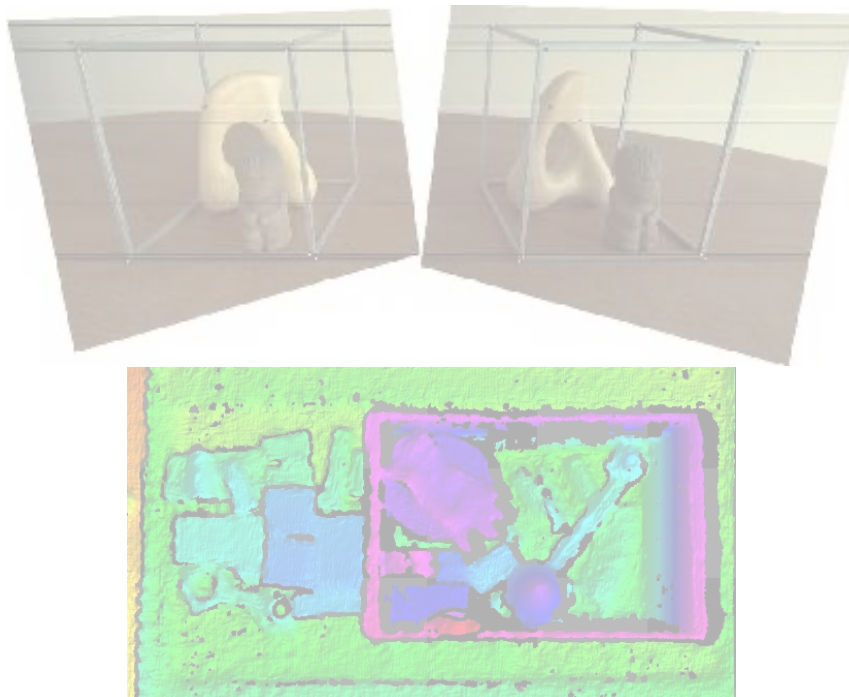
- Deep Learning (2020)

- [i:puglia20] – Intel SystemOnChip Lösung mit CNN Unterstützung und bis zu 250fps



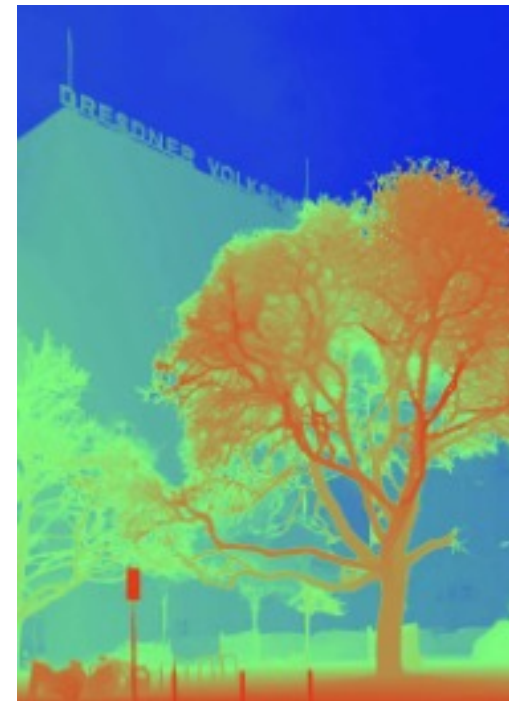
Aus zwei Bildern

- Klassische Stereo Vision



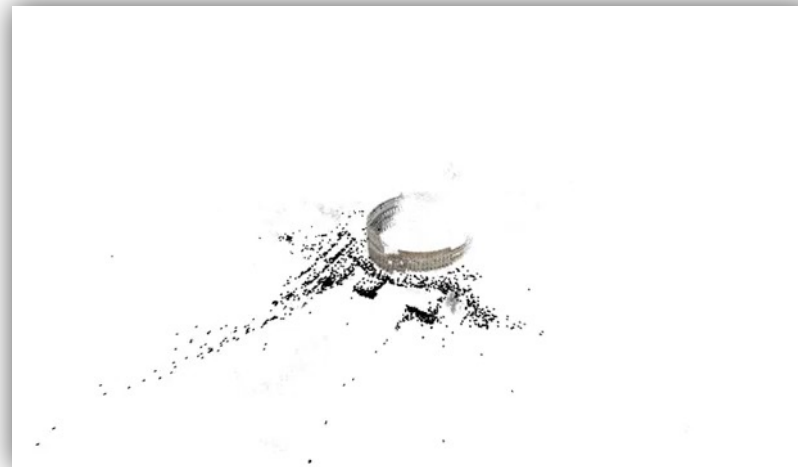
- Deep Learning (2025)

- DEFOM-Stereo: Depth Foundation Model Based Stereo Matching, [Jiang2025]



Aus vielen Bildern

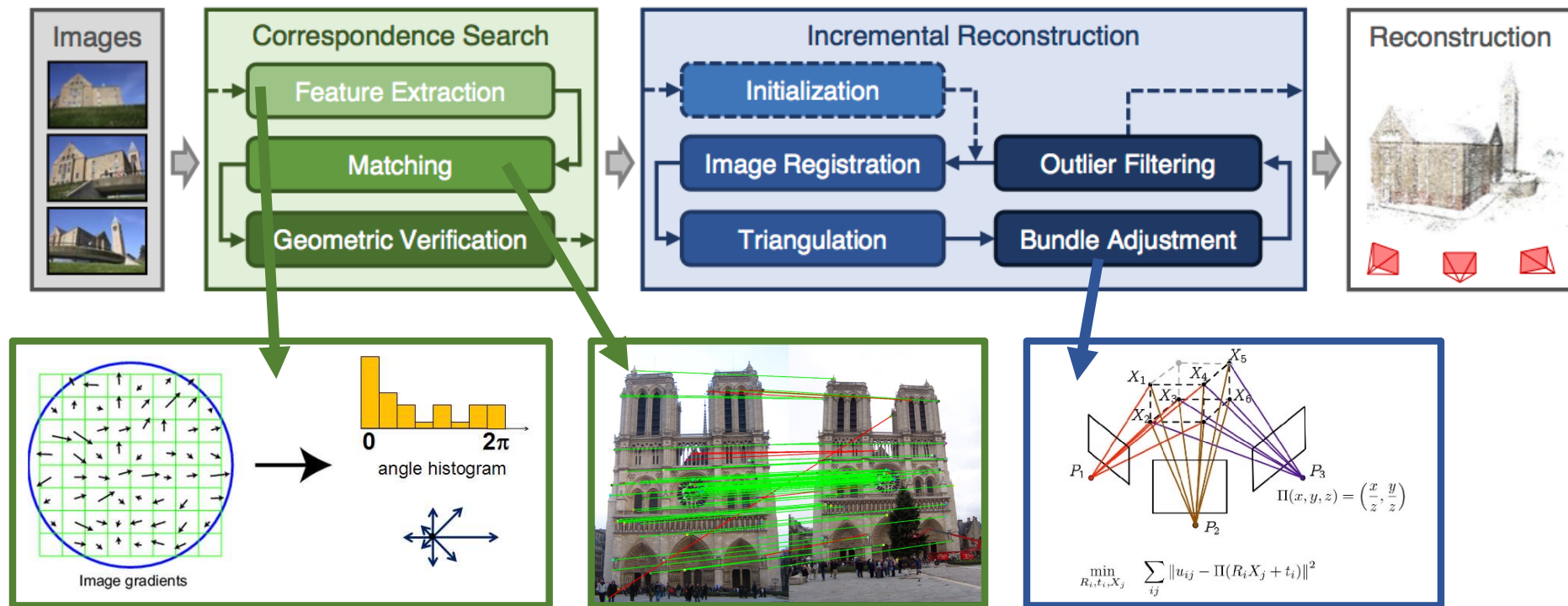
- klassisch: Structure from Motion (SfM)



[v:agarwal09]

Aus vielen Bildern

- klassisch: Structure from Motion (SfM)



Aus vielen Bildern

- Deep Learning: NeRF (2020)
 - Training: 1-2 Tage, Rendering: ca 2-3 Minuten pro Frame



[v:mildenhall20]

Aus vielen Bildern

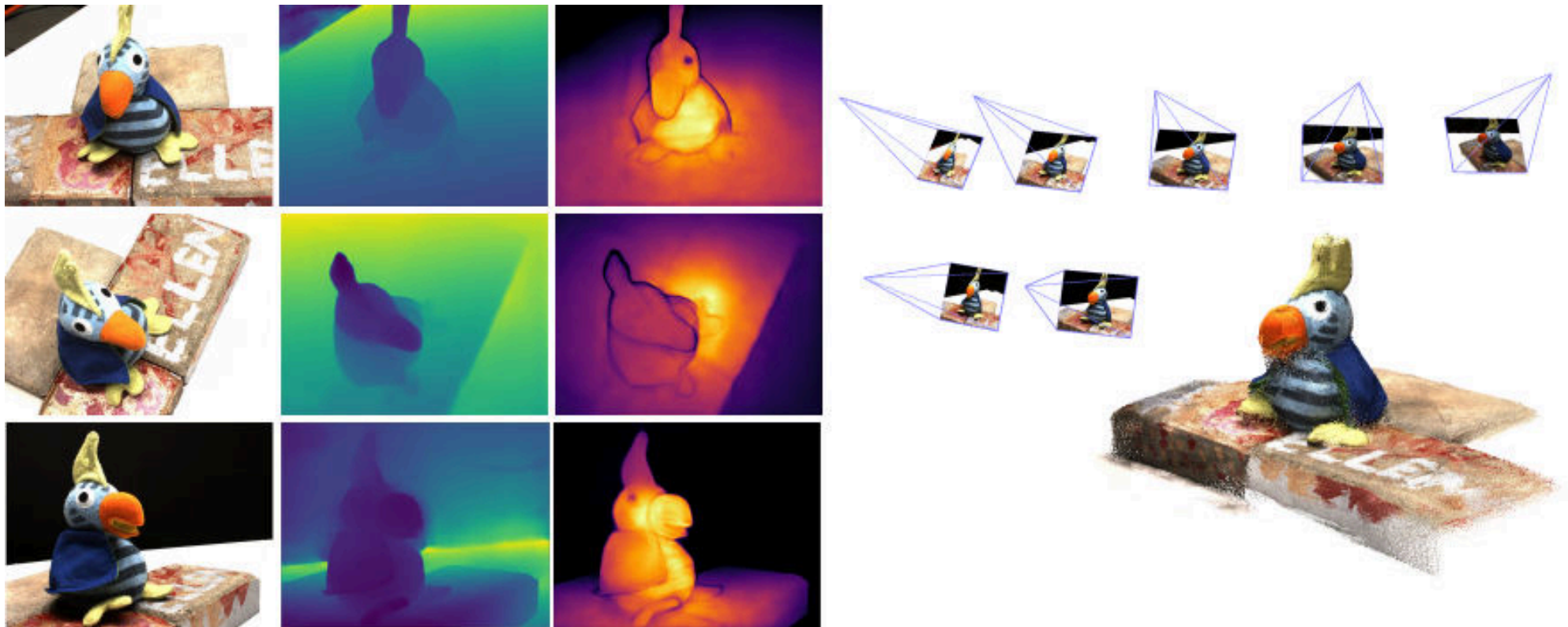
- Instant-NGP, 2022 (NVIDIA Implementierung)



[v:müller22]

Aus vielen Bildern

- DUS_t3R, Wang et al. from Naver Labs, CVPR 2024



[i:wang24]

Aus vielen Bildern



<https://github.com/ByteDance-Seed/depth-anything-3> (2025)

Aus vielen Bildern

32 Views



VGGT: <https://vgg-t.github.io/> (2025)

Deep Learning ist eine Revolution. Warum?

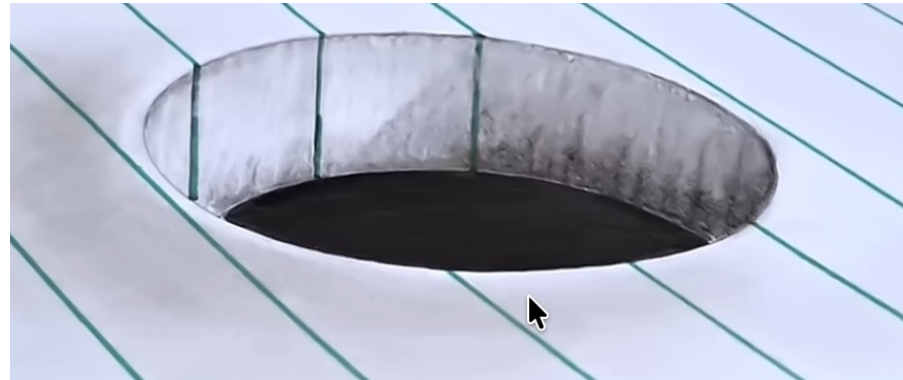
Zwei Gründe:

- Unvorstellbar gute Ergebnisse
 - Auf Basis von herkömmlichen Fotos
- Völlig andere Herangehensweise
 - Menschengemachte Modelle gegenüber aus Daten gelernte Modelle
 - Mehr Ausprobieren als Nachdenken
 - Mehr Datenaufbereitung als Algorithmenentwicklung

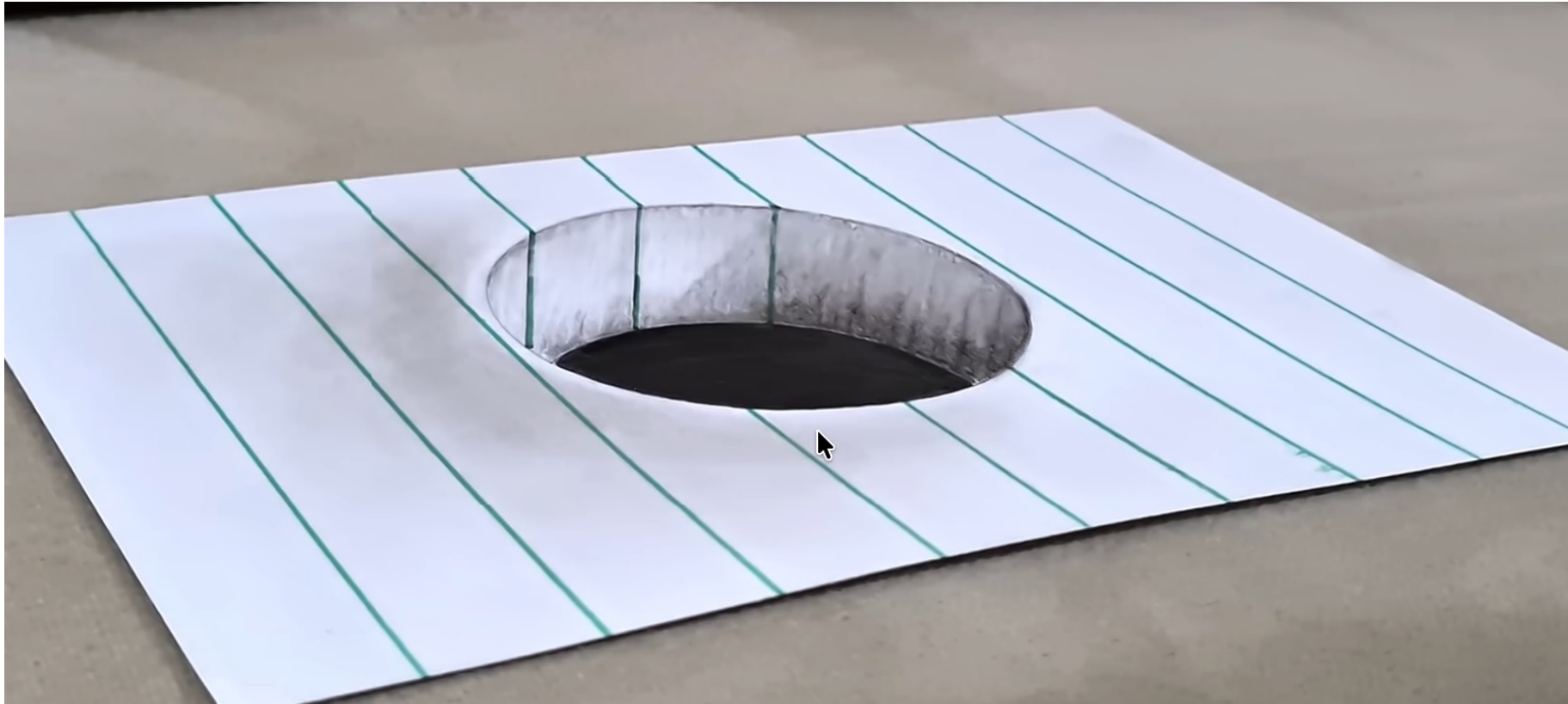
Erfolgsfaktoren

- **Moore's law** → exponentielle Entwicklung der verfügbaren Rechenleistung
- **Fotos** → Durch den Einbau von Kameras in Smartphones ist die Menge an verfügbaren Fotos explodiert
- **Open source** → Facebook, Google, IBM, Microsoft sind sehr offen
 - Viele Werkzeuge verfügbar (TensorFlow, PyTorch, Keras usw.)
- **Open education** → [MIT OpenCourseWare](#) hat über 5M Abonennten auf YouTube
 - [Gilbert Strang, Linear Algebra lectures](#) (fast) alle >100K Ansichten
 - A. Karpathy, [Stanford CS231n](#) alle Vorlesungen >30K Ansichten

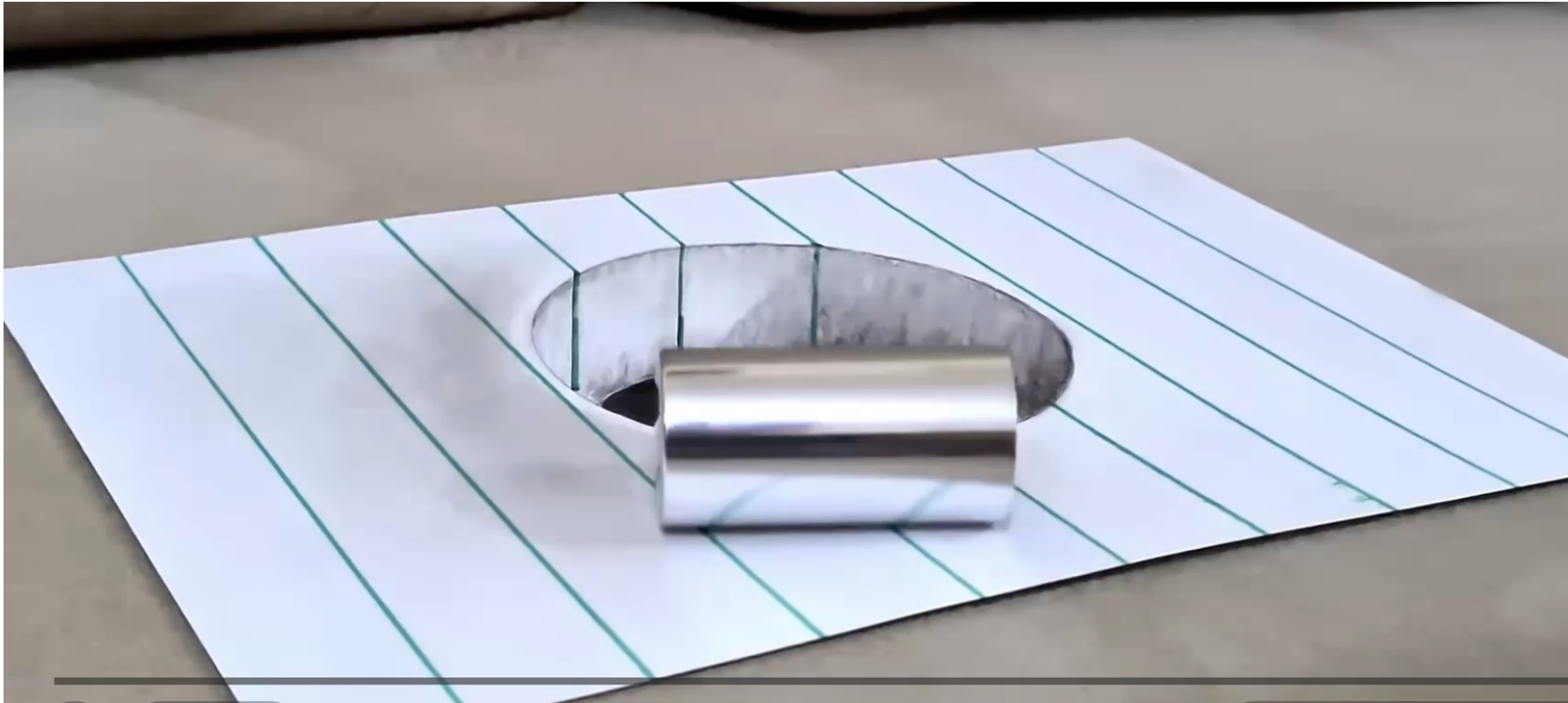
Grundidee



Grundidee



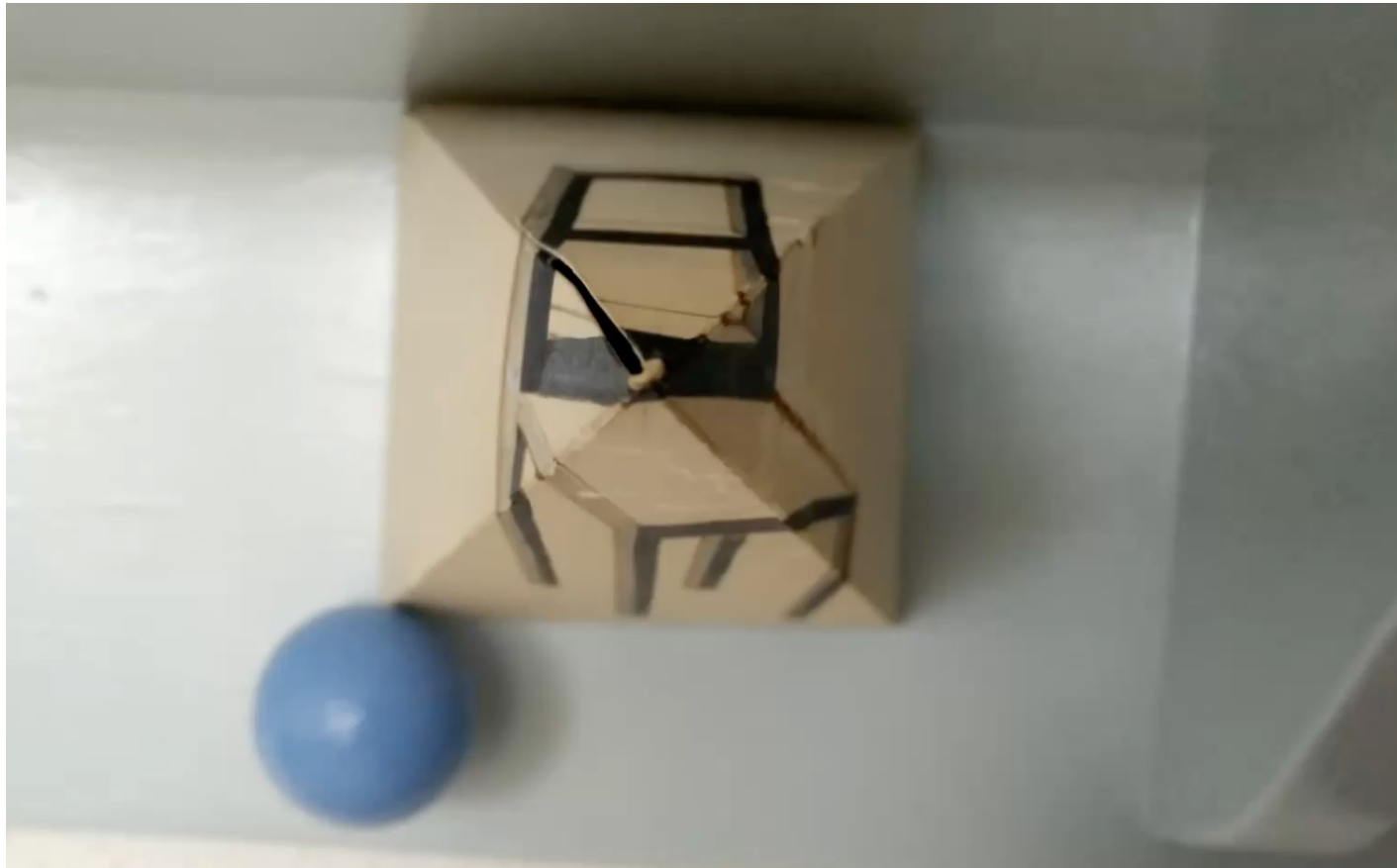
Grundidee



Grundidee



Grundidee



Grundidee

Wir nutzen unser Vorwissen und versuchen nicht immer die 3D Geometrie nur aus den Bilddaten zu gewinnen.

Praktischer Teil

Wir rekonstruieren gemeinsam diesen Raum.

Bilder aufnehmen

Jede*r bitte nur ein
Bild hochladen!



<https://3-dworkshop.vercel.app/>

COLMAP

- <https://colmap.github.io/index.html>



Sparse model of central Rome using 21K photos produced by COLMAP's SfM pipeline.



Dense models of several landmarks produced by COLMAP's MVS pipeline.



Hugging Face

HuggingFace Spaces (Collection):

- <https://huggingface.co/collections/uwehahne/3d-reconstruction>



DUST3R

- Project page:
 - <https://europe.naverlabs.com/research/publications/dust3r-geometric-3d-vision-made-easy/>

NAVER LABS
Europe

DepthAnything v3

Project page:

- <https://depth-anything-3.github.io/>

Code:

- <https://github.com/ByteDance-Seed/depth-anything-3>

VGGT- Ω

- Project page:
 - <https://vggt-omega.github.io/>
- Code:
 - <https://github.com/facebookresearch/vggt-omega>

Bildaufnahmen

- Sich wiederholende Strukturen können Computer-Vision-Algorithmen verwirren.
- Lichtverhältnisse beeinflussen die Bildqualität und die Rekonstruktionsgenauigkeit stark.
- Weitwinkelobjektive können dazu beitragen, einzigartigere Merkmale zu erfassen.
- Innenräume stellen einzigartige Herausforderungen dar, wie texturlose Wände.

Quellen

Bildquellen

- [i:colo07] – [source license](#)
- [i:colo10] – [source](#)
- [i:marr82] – Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco.
- [i:szeliski20] – Richard Szeliski – Computer Vision: Algorithms and Applications (September 20, 2020 draft)
- [i:hays18] – [source](#) – Project 2: Local Feature Matching, CS 6476: Computer Vision, James Hays, Samarth Brahmabhatt, and John Lambert,
- [i:crimini99] – [source](#) – Criminisi, A., Reid, I. & Zisserman, A. Single View Metrology. *International Journal of Computer Vision* 40, 123–148 (2000)
- [i:puglia20] – <https://arxiv.org/pdf/2001.04552v1.pdf> Deep Learning Stereo Vision at the edge, Luca Puglia and Cormac Brick, Intel R&D
- [i:jiang25]: Hualie Jiang, Zhiqiang Lou, Laiyan Ding, Rui Xu, Minglang Tan, Wenjie Jiang, and Rui Huang. [DEFOM-Stereo: Depth foundation model based stereo matching](#) (RVC version). arXiv:2501.09466, 2025.
- [i:wang24]: Shuzhe Wang and Vincent Leroy and Yohann Cabon and Boris Chidlovskii and Jerome Revaud. [DUSt3R: Geometric 3D Vision Made Easy](#), CVPR 2024.

Videoquellen

- [v:agarwal09] – [source](#) – Building Rome in a Day, Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski, International Conference on Computer Vision, 2009, Kyoto, Japan.
- [v:facebook20] [source](#) – Powered by AI: Turning any 2D photo into 3D using convolutional neural nets
- [v:WayneINR24] – <https://twitter.com/WayneINR/status/1750945037863551247>
Exploring the capabilities of DepthAnything – see [gallery](#)
- [v:cherabier18] [source](#) – Learning Priors for Semantic 3D Reconstruction, Ian Cherabier, Johannes L. Schönberger, Martin R. Oswald, Marc Pollefeys, Andreas Geiger
- [v:mildenhall20] [source](#) – NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, Ben Mildenhall and Pratul P. Srinivasan and Matthew Tancik and Jonathan T. Barron and Ravi Ramamoorthi and Ren Ng, 2020, ECCV
- [v:müller22] [source](#) – Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, Thomas Müller, Alex Evans, Christoph Schied, Alexander Keller (all from NVIDIA), ACM Transactions on Graphics (SIGGRAPH 2022, Best paper award)

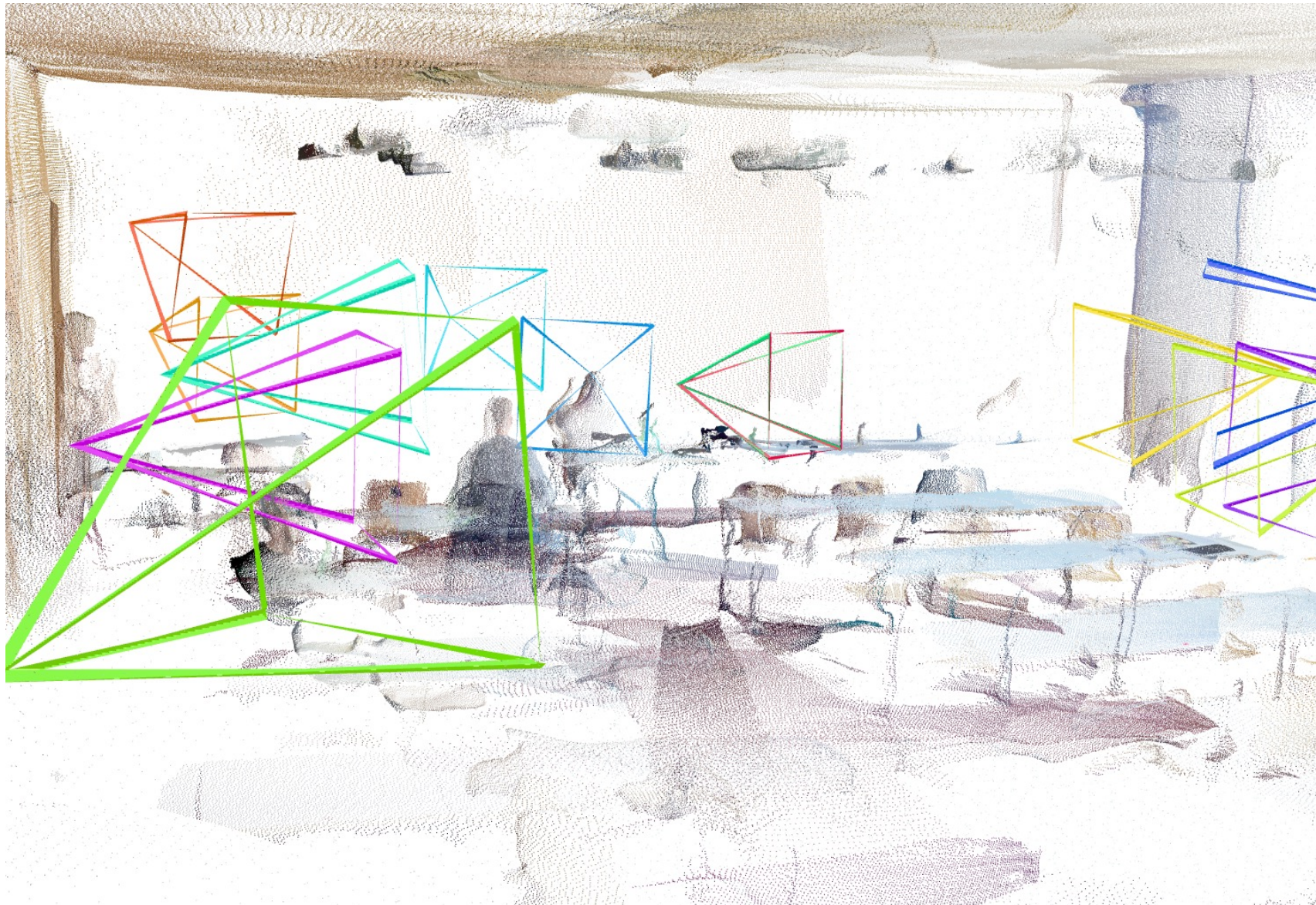
Textquellen

- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco.
- Richard Szeliski – Computer Vision: Algorithms and Applications (September 20, 2020 draft)
- <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
[Krizhevsky2012](#)

Ergebnisse DA3 1.Versuch



Ergebnis VGGT (2.Versuch)



Ergebnis VGGT (Personen)

