Die Software erkennt Gesichter besser als die Ermittler

Die Software erkennt Gesichter besser als die Ermittler

Was Gesichtserkennung bereits kann und womit sich die Forschung noch schwertut.

von Manuel Günther

Im Alltag verwenden viele die Gesichtserkennung, um ihre elektronischen Geräte automatisch zu entsperren. Es ist einfacher und sicherer, einmal das Gesicht in die Kamera zu halten, als sich eine PIN zu merken oder ein Muster auf dem Display zu zeichnen. Der automatischen Gesichtserkennung sind manche auch schon an Flughäfen begegnet, als sie durch ein sogenanntes E-Gate gegangen sind. Beide Anwendungen fallen in den Bereich der Verifikation, auch Authentifikation genannt, bei der es darum geht, ein aktuell aufgenommenes Foto mit einem vorher abgelegten Template zu vergleichen – sei es das Foto auf dem Reisepass oder ein Template, das während der Einrichtung des Smartphones aufgenommen wurde.

Der zweite Anwendungsbereich der Gesichtserkennung, nämlich die Identifikation einer Person, funktioniert zwar ähnlich, aber nicht ganz gleich: Es werden nicht nur zwei Bilder miteinander verglichen, sondern es wird versucht, der Person auf dem aktuellen Bild einen Namen zu geben. Dazu muss vorher eine Datenbank mit Gesichtern und zugehörigen Namen angelegt werden, die im Forschungsbereich der Biometrie häufig als Galerie bezeichnet wird. Eine solche Galerie kann zum Beispiel bei der Suche von vermissten Personen dienlich sein – sie kann aber auch sicherstellen, dass beispielsweise Hausverbote eingehalten werden. Im einfachsten Fall können aber auch nur die Personen auf einem persönlichen Gruppenfoto automatisch erkannt werden – als Galerie verwendet die Software dann Fotos, die bereits zu einem früheren Zeitpunkt aufgenommen wurden und bei denen der User selbständig die Namen der fotografierten Personen vermerkt hat.

In drei Schritten zur Erkennung

Die automatische Gesichtserkennung besteht typischerweise aus drei Schritten. Erstens benötigen alle bekannten Gesichtserkennungsalgorithmen ein Gesichtsbild in einem gewissen Format: Das Eingabebild muss also in einer bestimmten Auflösung vorliegen, und das Gesicht muss eine bestimmte Grösse und Position innerhalb des Bildes einnehmen. Um dies zu gewährleisten, wird das zu analysierende Gesicht im Originalbild detektiert und in das sogenannte Eingabebild extrahiert. In vielen Fällen werden dazu auch «Landmarken» im Gesicht detektiert, also auffällige Gesichtspartien wie der Mund, die Nase und die Augen. Anhand dieser wird das Eingabebild normiert.

Sobald das Eingabebild im gewünschten Format vorliegt, werden im zweiten Schritt Merkmale aus dem Bild extrahiert, die die wesentlichen Eigenschaften des Gesichts beinhalten, die zur Identifikation benötigt werden. Dazu gehören etwa die Grösse, Form und relative Position von Augen, Mund und Nase. Bestenfalls werden andere Einflussgrössen wie die Beleuchtung, der Gesichtsausdruck oder die Blickrichtung ausgeklammert.

Drittens werden die extrahierten Merkmale aus zwei Bildern miteinander verglichen, indem ein Ähnlichkeitswert berechnet wird. Ist dieser Wert grösser als ein vorher festgelegter Schwellenwert, legt der Algorithmus nahe, dass die beiden Bilder die gleiche Person zeigen. Wie genau der Schwellenwert festgelegt wird, ist von den Anwendungsfällen abhängig: Zum Beispiel wird zur Entsperrung des Smartphones meist ein niedrigerer Schwellenwert eingestellt als an sicherheitsrelevanten Punkten wie dem E-Gate am Flughafen.

Laufende Forschungsarbeiten

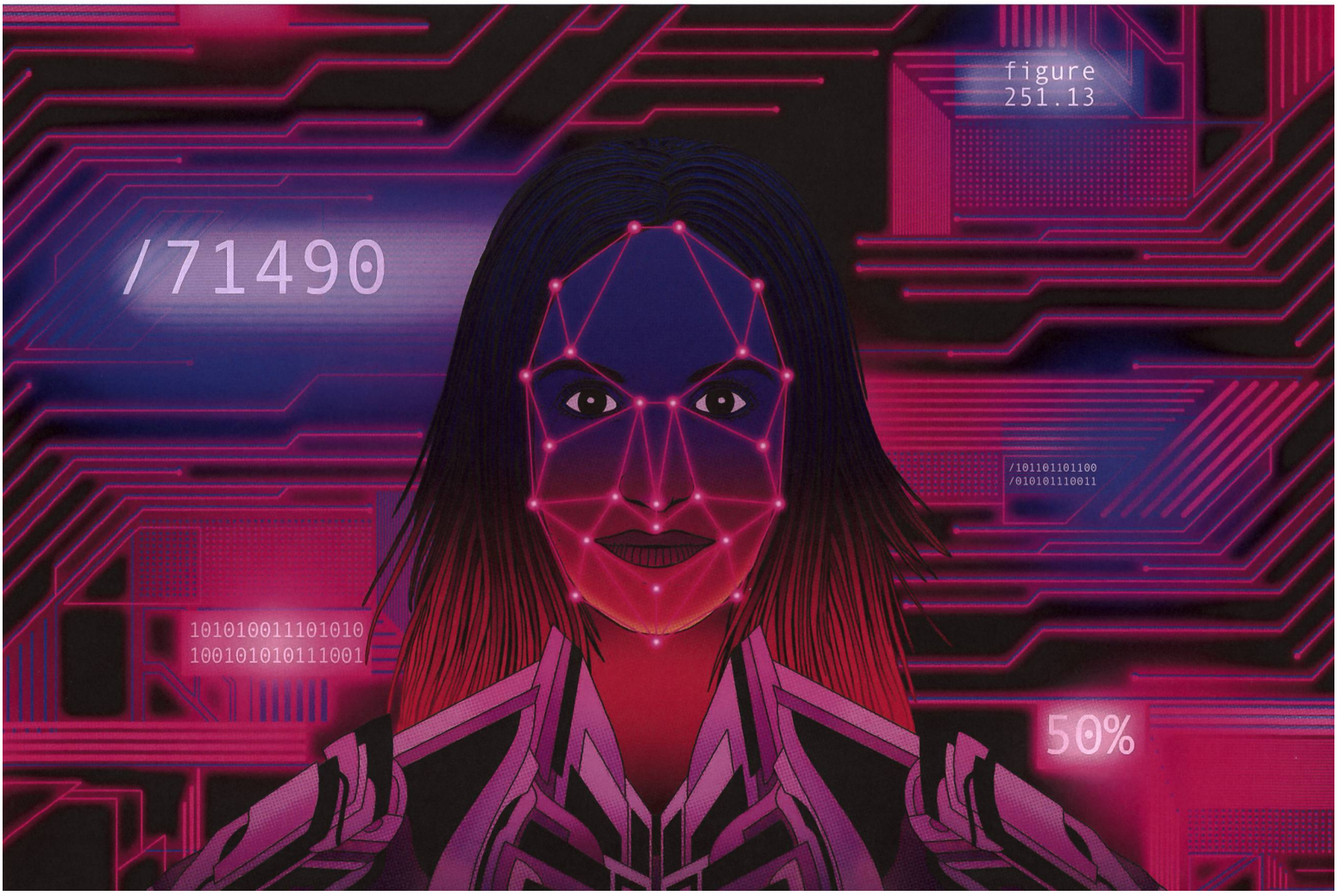
Seit mehreren Jahren ist das National Institute for Standards and Technology (NIST) in den Vereinigten Staaten eine treibende Kraft in der Entwicklung und im Vergleich von Lösungen zur automatischen Gesichtserkennung. Regelmässig werden in Face Recognition Vendor Tests (FRVT) viele kommerzielle und akademische Algorithmen in verschiedenen Einsatzgebieten auf Herz und Nieren getestet. In der letzten Reihe der FRVT wurden fast 400 Algorithmen aus aller Welt untersucht, darunter auch Lösungen von vier Schweizer und sechs deutschen Firmen. Dabei ist die staatliche Intelligence Advanced Research Projects Activity (IARPA) eine grosse Geldgeberin. Sie finanziert viele Projekte, die die Gesichtserkennung immer weiter vorantreiben.

Die Forschung an der Gesichtserkennung ist weder neu noch ist sie abgeschlossen. Erste kommerzielle Produkte zur Gesichtserkennung gibt es ungefähr seit Mitte der 90er-Jahre, die ersten Versuche der automatischen Gesichtserkennung in der Wissenschaft gehen jedoch bis in

die 1980er-Jahre zurück. Die Forschenden versuchten damals, die Abstände von unterschiedlichen Gesichtslandmarken in zwei Bildern miteinander zu vergleichen. Da die automatischen Algorithmen zur Landmarkendetektion noch in den Kinderschuhen steckten und die Vergleichsalgorithmen noch nicht sehr zuverlässig waren, bedeutete dies, dass die Landmarken per Hand positioniert werden mussten. Den Ansatz der Distanzabmessung hat die Forschung dann in den 1990er-Jahren aufgegeben.

Seither versucht man eher, Texturmerkmale aus den Eingabebildern zu extrahieren. Diese werden grob in zwei Klassen eingeteilt: Globale Merkmale schauen das Gesicht als Ganzes an und extrahieren Texturmerkmale aus dem gesamten Bild. Ein sehr bekannter Vertreter dieses Ansatzes ist der sogenannte Eigengesichtsansatz, der die einzelnen Bildpunkte als unabhängig betrachtet und mit Hilfe einer Eigenwertanalyse einen sogenannten Gesichtsraum (Face Space) erzeugt, in dem Bilder miteinander verglichen werden können. Dieser Ansatz ist zwar schnell, aber kleinste Abweichungen in der Ausrichtung des Bildes können die Erkennung erschweren. Lokale Merkmale zerschneiden das Eingabebild in viele kleine Teile, aus denen jeweils Merkmale extrahiert werden, die am Ende wieder zu einem Ganzen zusammengefügt werden. Solche lokalen Texturmerkmale könnten zum Beispiel implizit die Grösse, Form und Position von Teilen des Gesichts kodieren, wie zum Beispiel jene einer Augenbraue, ohne diese Kenngrössen explizit zu berechnen. Ein bekannter Ansatz, der sogenannte elastische Graphenvergleich (Elastic Graph Matching), detektiert Landmarken im Gesicht und extrahiert dort lokale Texturmerkmale, die der Bildverarbeitung im Gehirn nachempfunden sind.

In den letzten Jahren hat die Forschung zur Gesichtserkennung neuen Aufwind bekommen. Anhand von riesigen Mengen an Gesichtsbildern, die zu Eingabebildern normiert wurden, lernt ein Deep-Learning-Verfahren automatisch, welche Merkmale aus dem Bild extrahiert werden müssen, um eine Identifikation vornehmen zu können. Dazu müssen die riesigen Datenmengen mit Namen versehen werden. Zudem sollen viele Bilder einer Person in möglichst vielen Variationen hinsichtlich Beleuchtung, Gesichtsausdruck oder Kopfrotation vorhanden sein. Solche Datensätze existieren meist von prominenten Persönlichkeiten, weil sie besonders oft fotografiert und die Bilder veröffentlicht werden. Jedoch sind nicht alle Urheberrechtsfragen für die Nutzung solcher Datensätze geklärt.



Generell gilt: je grösser und variantenreicher der Datensatz zum Training des Deep-Learning-Systems, desto besser funktioniert dieses.

Mit Hilfe eines solchen Systems wird nun eine Galerie von Templates angelegt, die die zu erkennenden Personen beinhaltet. Dabei besteht ein Template aus möglichst vielen verschiedenen Variationen von Gesichtsbildern einer Person. Dies ermöglicht es, später Gesichter mit einer der enthaltenen Variationen besser zu erkennen. Manche kommerzielle Algorithmen, zum Beispiel zur Entsperrung des Smartphones, fordern die Nutzer daher auf, ihr Gesicht aus mehreren Perspektiven und mit oder ohne Brille aufzunehmen. Durch neueste Entwicklungen in der Technologie reicht es aber auch aus, ein einzelnes frontales Bild zum Erstellen des Templates aufzunehmen.

«Fehlerhafte Identifikation hat schon mehrfach zu Festnahmen von Unschuldigen geführt.»

Manuel Günther

Auch Ansätze zur Verwendung von 3D-Daten zur Gesichtserkennung sollten nicht unerwähnt bleiben. Dabei werden Merkmale aus 3D-Scans von Gesichtern extrahiert und miteinander verglichen. Allerdings sind die Datenmengen im Bereich 3D noch weit kleiner als bei 2D-Bilddaten – die Deep-Learning-Verfahren haben wenig Material, das sie als Trainingsbasis verwenden können. 3D-Algorithmen sind dementsprechend momentan noch sehr rar.

Ob beispielsweise bei Face ID, dem Programm zur Entsperrung der neueren iPhones, tatsächlich schon 3D-Algorithmen zum Einsatz kommen, ist unklar. Apple schreibt dazu auf seiner Website: «Die TrueDepth-Kamera erfasst genaue Gesichtsdaten, indem Tausende unsichtbarer Punkte projiziert und analysiert werden, um daraus eine Tiefenkarte des Gesichts zu erstellen. Zudem zeichnet sie ein Infrarotbild deines Gesichts auf.» Weil Apple die genaue Vorgehensweise aber nicht veröffentlicht und es keine wissenschaftlichen Publikationen dazu gibt, bleibt einiges unklar.

Probleme und versuchte Austricksereien

Mit Hilfe des Deep Learning sind automatische Gesichtserkennungsalgorithmen mittlerweile viel besser als menschliche Experten, wenn es um den Vergleich zweier Gesichtsbilder in Passfotoqualität geht: Für die E-Gates werden mittlerweile Schwellenwerte verwendet, die es im Mittel nur einer von einer Million Personen erlauben, mit einem fremden Ausweis zu reisen. Wenn die Bildqualität allerdings nicht so hoch ist, wie das gerade bei Überwachungskameras oft der Fall ist, können schnell einmal Probleme auftauchen. Zum Beispiel werden in einigen Staaten der USA und im Vereinigten Königreich Polizisten mit Bodycams ausgestattet, die alle Personen automatisch mit einer Verbrecherdatenbank abgleichen. Fehlerhafte Identifikation hat dort schon mehrfach zu Festnahmen Unschuldiger geführt. Unbescholtene Bürger werden also einer Straftat verdächtigt, die sie nicht begangen haben.

Ein weiteres Problem besteht in der Auswahl der Daten zum Training des Deep Learning: Aktuelle Untersuchungen haben gezeigt, dass Männer bisher besser erkannt werden als Frauen und dass die Fehlerrate bei gewissen ethnischen Minderheiten weitaus höher ist – nur weil diese Gruppen nicht ausreichend im Trainingsdatensatz enthalten sind. Ein grosses Feld der Biometrie befasst sich momentan auch mit der Detektion von Angriffen auf Gesichtserkennungssysteme: Es soll verhindert werden, dass sich jemand einfach ein Bild einer anderen Person ausdruckt oder ein Video auf dem Smartphone abspielt und dieses zur Authentifikation verwendet. Solche Attacken können leicht mit Hilfe von 3D- und Materialerkennung verhindert werden. Neuere Attacken – bekannt etwa aus der Filmreihe «Mission Impossible» – beinhalten auch die Erstellung von dreidimensionalen Gesichtsmasken oder die Verwendung von Make-up mit dem Ziel, das Gesichtserkennungssystem auszutricksen. Wer zur Gesichtserkennung forscht, soll sich dieser Risiken im voraus schon bewusst sein. ‹

Manuel Günther
ist Assistenzprofessor für künstliche Intelligenz und maschinelles Lernen an der Universität Zürich.