

Täuschend echt: Die Erzeugung von digitalen Bildfälschungen und ihre Erkennung

CHRISTIAN RIESS

Abstract: Bilder und Videos spielen seit jeher eine tragende Rolle in der Kommunikation. Die Bilder zu Ereignissen wie der Mondlandung, dem Mauerfall, oder dem 11. September 2001 beeinflussen stark das kollektive Gedächtnis. Ein visuelles Dokument stellt jedoch nicht automatisch ein tatsächliches Ereignis dar. Film- und Videospielproduktionen der letzten Jahre führen eindrucksvoll vor Augen, dass eine visuelle Realität jenseits der tatsächlichen physikalischen Welt zunehmend einfacher und glaubwürdiger erschaffen werden kann. Führt dies automatisch dazu, dass keinem Bild mehr geglaubt werden darf? In dieser Arbeit werden Methoden aus der digitalen Bildforensik vorgestellt, mit denen ein Bild oder Video auf Authentizität und Ursprung geprüft werden kann. Hierdurch erhalten Analysten einen digitalen Werkzeugkasten, um in Streitfällen Bild- und Videomanipulationen aufdecken zu können.

Bilder sind Nachrichten

Bilder sind eine sehr elementare und doch auch gleichzeitig komplexe Form der Kommunikation. Ein Betrachter muss nicht lesen können oder andere Formen der Ausbildung mitbringen, um die Botschaft eines Bildes verstehen zu können. Durch diese visuelle Kommunikation können schnell und effektiv auch Emotionen mitgeteilt werden, was über andere Medien deutlich schwieriger und langwieriger ist, insbesondere die Schriftform. Hierdurch spielen Bilder seit ehedem eine herausragende Rolle, beispielsweise zur Dokumentation historischer Ereignisse, politischer und religiöser Kommunikation oder zur Illustration phantastischer Welten. Insbesondere bleiben Bilder oftmals dauerhaft im Gedächtnis haften und erzielen so eine besondere Langzeitwirkung auf den Betrachter. Man denke etwa an bekannte Beispiele aus der jüngeren Geschichte wie die Dokumentation der Mondlandung von Neil Armstrong und Edwin ‚Buzz‘ Aldrin im Jahr 1969, die von Millionen weltweit auf dem Fernseher verfolgt wurde. Oder an die Bilder von dem Fall der Berliner Mauer 1989, oder vom Einschlag der Flugzeuge in das World Trade Center am 11. September 2001. Neben diesen weltpolitisch sehr bedeutenden Ereignissen spielen Bilder auch in der

alltäglichen Kommunikation eine wichtige Rolle. Viele Nachrichtenmagazine illustrieren mit Bildern Kurzfassungen der jeweiligen Nachricht, und die Werbeindustrie arbeitet fast ausschließlich mit visueller Kommunikation. Durch das Internet sind die Grenzen zwischen textueller und visueller Kommunikation noch weiter verschoben worden – im Gegensatz zu herkömmlichen Büchern oder Tageszeitungen ist nur ein geringer Anteil der Kommunikation in den sozialen Medien ausschließlich textbasiert. Die *Online*-Plattform *Twitter* beschränkt einen Beitrag auf 280 Buchstaben oder Zeichen, erlaubt es aber, ein Bild oder Video an den Beitrag anzuhängen. In Beiträgen mit Bildern oder Videos beinhaltet dann der Textbeitrag oftmals nur einen kurzen persönlichen Kommentar zu der Bildbotschaft. Die *Online*-Plattform *Instagram* rückt in den von Benutzern erstellten Beiträgen das Bild komplett in den Vordergrund der Nachricht: Ein Benutzerbeitrag besteht im Wesentlichen aus einem Bild, das mit einem zusätzlichen kurzen Kommentar versehen werden kann.

Wie entsteht eine Bildbotschaft?

Der Verfasser einer Nachricht verfolgt immer einen Zweck, der sich typischerweise im Inhalt der Nachricht wiederfindet. Im Fall von Bildern oder Videos wird die Nachricht durch eine Reihe von Entscheidungen des Verfassers geformt und gegebenenfalls hervorgehoben und verstärkt.

Unter den vielfältigen Möglichkeiten, eine Bildbotschaft zu erstellen, kann die Wahl des Ausschnitts einer Szene entscheidend sein. Werden beispielsweise in einer Aufnahme aus einem Kriegsgebiet zerstörte Gebäude gezeigt, so schließt die Bildbotschaft automatisch die Verwüstungen der Kampfhandlungen mit ein. Als Gegenbeispiel einer Aufnahme an dem selben Ort könnte beispielsweise ein noch erhaltenes Gebäude mit der Fahne der siegreichen Fraktion gezeigt werden, um den Erfolg dieser Kriegspartei zu betonen, ohne im Bild die Zerstörungen zu thematisieren.

Ein Bildredakteur hat die ‚Freiheit‘, bestimmte Akteure gezielt in den Blickpunkt zu rücken. Um bei dem Beispiel der Kriegsberichterstattung zu bleiben: Es ist denkbar, eine Szene in Nahaufnahme ohne die umgebenden Ruinen zu zeigen, in der ein Soldat einem Kind Süßigkeiten gibt. Hier kann die Bildbotschaft darauf abzielen, dass der Verfasser die Soldaten als der Bevölkerung wohlgesonnene Befreier darstellt. Falls nur wenige Meter neben dieser Szene, aber außerhalb des Bildes, eine Kolonne von Kriegsgefangenen vorbeimarschiert, dann wäre dies ein Motiv mit einer gänzlich anderen Botschaft.

Der Bildredakteur hat überdies die ‚Freiheit‘, einen konkreten Aufnahmepunkt auszuwählen. Beispielsweise werden politische Meldungen oftmals mit dem Gesicht eines Politikers illustriert, das je nach Situation eine glückliche, betrübte oder wütende Miene zeigt. Selbst wenn das Bild zu der Zeit und an dem Ort aufgenommen wurde, auf die sich die Meldung bezieht, dann beruht solch ein Bild doch oft auf einem Schnappschuss, auf dem just in diesem Augenblick die Person das Gesicht so verzog, dass es als Standbild offenbar eine Emotion transportiert, die dem Inhalt der Meldung entspricht.

Diese und weitere Entscheidungen zusammengefasst formen die Bildbotschaft, nämlich indem eine Szene so komponiert wird, dass der Betrachter die Bildsprache automatisch interpretieren und verstehen kann.

Was ist eine Bildmanipulation im technischen Sinn?

Der Verfasser einer Nachricht verfolgt immer einen bestimmten Zweck. Hinzu kommt, dass eine Nachricht stets aus der Perspektive des Verfassers heraus konzipiert wird. Das führt dazu, dass – selbst wenn der Verfasser versucht, einen Sachverhalt objektiv darzustellen – diese Darstellung zwangsläufig von dem Erfahrungshintergrund des Verfassers eingefärbt wird.

In vielen Fällen ist eine objektive Darstellung von Ereignissen aber gar nicht das Ziel der Dokumentation, und die Grenzen zur Lüge sind naturgemäß fließend. Analog zu geschriebenen oder gesprochenen Nachrichten kann auch eine Bildnachricht eine Lüge enthalten. Hierbei möchte ich den Begriff ‚Lüge‘ eingrenzen auf eine bewusste Täuschungsabsicht des Verfassers der Nachricht. Die Motive für diese Täuschungsabsicht können vielfältig sein und sind im Wesentlichen unabhängig von dem Umstand, dass die Täuschung mit einem Bild oder Video umgesetzt wird. Ebenso vielfältig sind die Möglichkeiten, die Täuschungsabsicht konkret mit einem Bild oder Video zu realisieren.

Eine Palette von Möglichkeiten wurde bereits im vorangegangenen Abschnitt zur Entstehung einer Bildbotschaft vorgestellt: Durch die Wahl des Orts, des Zeitpunkts, und des konkreten Ausschnitts eines Motivs wird die Bildbotschaft substanziell geformt. Bewusste Betonungen oder Weglassungen können die handwerklichen Bausteine für die Umsetzung einer Lüge mit Bildmaterial sein.

Ein bekanntes Beispiel ist ein Bild aus einem Kriegsgebiet, in dem ein Kriegsgefangener von einer Seite von einem anderen Soldaten mit der

Waffe bedroht, von der anderen Seite von einem weiteren Soldaten mit einer Wasserflasche versorgt wird. Durch die Wahl eines Ausschnitts der Szene in dem letztendlichen Bildmotiv war der Kriegsgefangene entweder nur mit dem ihn bedrohenden Soldaten oder dem ihn versorgenden Soldaten im Bild zu sehen. Beide Ausschnitte zeigen jeweils nur einen Teil der Realität, der die Bildbotschaft in die eine oder andere Richtung polarisiert.

Derartige Manipulationen lassen sich in der Regel durch klassische Recherchearbeit aufdecken, durch dritte Beobachter vor Ort und durch allgemeines Kontextwissen zu der abgebildeten Situation.

Es ist allerdings auch denkbar, dass der Verfasser einer Bildbotschaft keinen Zugang zu einer geeigneten realen Szene hat (oder dass gar keine reale Szene existiert), in der er durch geeignete Auswahl des Aufnahmeorts, des Aufnahmezeitpunkts, des Bildausschnitts und des Moments der Aufnahme die beabsichtigte Bildbotschaft erstellen könnte. In diesem Fall besteht für einen Verfasser mit Täuschungsabsicht die Möglichkeit, durch technische Eingriffe in ein bestehendes Bild eine passende Bildbotschaft zu erzeugen. Solch ein Eingriff kann beispielsweise darin bestehen, mit einem Bildbearbeitungsprogramm Objekte und Personen zu entfernen oder aus anderen Szenen dem bearbeiteten Bild hinzuzufügen. In der Bildforensik wird solch ein Eingriff als Bildmanipulation bezeichnet.

Derartige technische Eingriffe in ein bestehendes Bild führen dazu, dass das Bild kein ‚Kamera-Original‘ mehr ist, das heißt, dass das Bild nicht mehr in dieser Form aus einer Kamera stammt. Diese Eingriffe hinterlassen oftmals Spuren, die sich mit einer technischen Analyse finden lassen. Das Forschungsfeld der Bild- und Videoforensik (oftmals auch summarisch als ‚Multimediaforensik‘ bezeichnet) benutzt Methoden aus der Mathematik, der Informatik, der Elektrotechnik und der Physik, um derartige Spuren aufzuzeigen. Praktisch gesehen werden Algorithmen entwickelt, um Indizien von Eingriffen in Bildern und Videos zu sammeln.

Hierbei ist es wichtig, sich die Möglichkeiten und Grenzen einer algorithmischen Prüfung bewusst zu machen. Mit technischen Mitteln ist es lediglich möglich aufzuzeigen, dass etwas an einem Bild oder Video verändert wurde, sowie in günstigen Fällen, an welchen Stellen und eventuell mit welchen Methoden. Es wäre allerdings voreilig, dies gleichzusetzen mit der Aufdeckung einer Lüge: Grundsätzlich muss ein Eingriff nicht mit einer Täuschungsabsicht verbunden sein. Das vielleicht populärste Beispiel für einen Eingriff in Bilddaten ohne Täuschungsabsicht ist die Entfernung von roten Augen in Fotografien. Ebenso können komplexere Eingriffe gänzlich ohne Täuschungsabsicht vorgenommen werden. Beispielsweise speisen

sich viele sogenannte Memes, ein Genre von satirischen Collagen der Internet-Subkultur, im Wesentlichen aus dem Zusammenschnitt bekannter Bildelemente in neuem Kontext. Hier liegt *per se* keine Täuschungsabsicht vor, da dem Zielpublikum die Quellen und Stilelemente dieser Collagen bekannt sind. Darüber hinaus bemühen sich die Verfasser der Collagen oftmals nicht darum, eine visuell realistische Kombination der Bildelemente zu erreichen. Stattdessen dient häufig der rohe Zusammenschnitt in seinem prototypischen Stadium dem Kommunikationsziel, die künstliche Kombination der beiden Bildquellen zu betonen.

Das Vorliegen eines technischen Eingriffs in ein Bild ist also kein hinreichendes Merkmal, um von einer Täuschungsabsicht zu sprechen. Da ein Algorithmus aber nur technische Eingriffe feststellen kann, ist es einem Algorithmus selbst nicht möglich, eine Täuschungsabsicht nachzuweisen. Dies führt in letzter Konsequenz dazu, dass es kein vollautomatisiertes System zur Manipulationserkennung geben kann.

Diese Schlussfolgerung basiert auf einem pathologischen, seit Gründung der Forschungsrichtung der ‚Künstlichen Intelligenz‘ bestehenden Problem, der sogenannten ‚Semantischen Lücke‘: Es ist zwar grundsätzlich möglich, Muster zu erkennen, aber es ist nicht möglich, diese so zu verstehen, wie dies dem Menschen möglich ist. Dies lässt sich am Beispiel von Witzen erklären. Einem Algorithmus ist es nicht möglich, einen Witz zu verstehen. Die algorithmischen Möglichkeiten beschränken sich darauf, entweder alle bekannten Witze auswendig zu lernen (und bei Bedarf auch wiederzugeben), oder bestimmte Strukturelemente und Auftrittswahrscheinlichkeiten von Witzen zu gewichten, beispielsweise an Hand des Vorkommens des Wortes ‚Fritzchen‘, oder des Erscheinens der fraglichen Textpassage in einem bekannten Satiremagazin. Auf Bilder bezogen kann dieses Beispiel auf die satirischen Internet-Memes übertragen werden: Es ist grundsätzlich nicht möglich, ein Meme als solches zu erkennen, sondern es können lediglich Indizien dafür gesammelt werden, dass es sich um ein Meme handelt, beispielsweise über die publizierte Website und die Wiedererkennung bekannter Text- und Bildelemente.

Für die Anwendung von bild- und videoforensischen Methoden hat dies die weitreichende Konsequenz, dass die technischen Prüfmethode lediglich Werkzeuge sein können, die Indizien für Eingriffe in ein Bild oder Video bereitstellen. Für die letztendliche Beurteilung, ob es sich bei einem Eingriff um eine Bildmanipulation handelt, muss allerdings immer ein menschlicher Analyst diese Indizien bewerten, ähnlich wie ein Polizist und gegebenenfalls ein Richter die Ergebnisse einer forensischen Untersuchung am Tatort zu einem Gesamturteil zusammensetzt.

Ein Rundgang durch die Geschichte der Bildmanipulationen

Mit der Erkenntnis, dass Bildmanipulationen eine Täuschungsabsicht beinhalten, liegt der Verdacht nahe, dass die Geschichte der Bildmanipulationen so alt ist wie bildliche Abbildungen.

Ein historisches Beispiel für eine gefärbte Darstellung von Ereignissen ist der berühmte ‚Teppich von Bayeux‘ (11. Jh.), der die Geschichte der Invasion von Wilhelm dem Eroberer in England zeigt. Die Abbildungen auf dem Teppich dokumentieren ein historisches Ereignis, sie sind aber auch eine politische Rechtfertigung der Invasion durch die Darstellung der normannischen Sichtweise der Vorgeschichte. Das Sprichwort ‚Geschichte wird von Siegern geschrieben‘ passt in diesem Fall ausgezeichnet.

Ein historisches Beispiel für eine Bildmanipulation, das unsere Definition für einen technischen Eingriff in das eigentliche Bild erfüllt, ist das Gemälde *„Jacques Marquet, Baron de Montbreton de Norvins“* (1811) von Jean Auguste Dominique Ingres. Dieses Bild enthielt in seiner ursprünglichen Fassung im Hintergrund eine Büste, die mutmaßlich ein Abbild von Napoleons Sohn war. Nach Napoleons Sturz 1814 wurde dieser Teil des Bildes offenbar nachträglich übermalt, wohl um der veränderten politischen Lage Rechnung zu tragen.¹

Seit Erfindung der Fotografie wurde der technische Teil der Bildentstehung schlagartig vereinfacht: Ein Foto erstellt in Sekundenbruchteilen ein Spiegelbild der Szene. Hierdurch gewinnt das Bild eine neue Rolle als Mittel der Dokumentation, zeigt doch die Abbildung eine zeitliche und räumliche Momentaufnahme der realen Welt. Diese dramatische Vereinfachung der visuellen Kommunikation ermöglichte die modernen Formen visueller Kommunikation, in der öffentliche – aber auch private – Ereignisse effizient aufgezeichnet und zeitnah verbreitet werden können.

Die klassische, analoge Fotografie ist ein chemischer Prozess, in dem sich das einfallende Licht auf einen lichtempfindlichen Film einbrennt. Dieser wird in einem mehrstufigen Verfahren zu einem dauerhaften, lichtunempfindlichen Negativ entwickelt, mit dessen Übertragung auf Fotopapier das eigentliche Foto entsteht.

In dem Entwicklungsprozess wird stets das Bild als ganzes bearbeitet. Dennoch gibt es eine ganze Reihe von historischen Beispielen, in denen analoge Bilder retuschiert wurden, ganz ähnlich zu dem oben genannten Beispiel des Gemäldes von Ingres. Eine schöne Auswahl von Fälschungen

¹ Vgl. Gary Tinterow, Philip Conisbee and Hans Naef, ed., *Portraits by Ingres: Image of an Epoch* (New York: The Metropolitan Museum of Art, 1999), 140–143.

von Analogfotos hat eine Forschergruppe um Hany Farid auf einer Website gesammelt.² Diese dokumentieren insbesondere das Hinzufügen und Entfernen von Personen auf Bildern. Beispielsweise wurde auf einer Aufnahme von Matthew Brady von 1865, die Generäle der US-Nordstaaten zeigt, der General Francis P. Blair nachträglich hinzugefügt. Für das Entfernen von Personen aus Bildern gibt es ebenfalls viele Beispiele. Insbesondere in der Sowjetunion, aber auch in anderen Staaten, wurden häufig Personen nachträglich aus Fotos entfernt, nachdem sie bei Machthabern in Ungnade gefallen waren. Neben Hinzufügungen und Entfernungen gibt es auch Fälle von Komposition analoger Bilder. Ein bekanntes Foto von US-Präsident Abraham Lincoln ist beispielsweise eine Zusammensetzung: Lincolns Kopf wurde auf den Körper des Südstaatenpolitikers John Calhoun montiert.

Die Manipulation von analogen Fotos ist ein aufwendiger, manueller Prozess. Obwohl es eine Vielzahl von Beispielen historischer Bildmanipulationen gibt, darf vermutet werden, dass sich Retuschen oder Kompositionen in historischen Bildern auf Fälle beschränken, in denen der Zeitaufwand und die Kosten für die Durchführung des Eingriffs durch andere Rahmenbedingungen gerechtfertigt wurden.

Die Reduktion von Kosten ist eine wesentliche Triebfeder des technischen Fortschritts im Allgemeinen und der Digitalisierung im Speziellen. Die Auswirkungen der fortschreitenden Automatisierung können besonders eindrucksvoll am Beispiel von Bild- und Videobearbeitungsmethoden nachvollzogen werden.

Die Unterhaltungsindustrie hat hierbei ein besonderes Interesse daran, den Stand der Kunst voranzutreiben. Das Geschäftsmodell beruht zu einem guten Teil darauf, dem Publikum Kuriositäten und eine Realität außerhalb des gewohnten Umfelds zu präsentieren. Dies lässt sich am Beispiel von Filmproduktionen aus den Hollywood-Studios nachvollziehen. Die Monumentalfilme der fünfziger und sechziger Jahre wie *Ben Hur* (1959), *El Cid* (1961) und *Cleopatra* (1963) stachen sowohl durch die epische Präsentation der jeweiligen Geschichten heraus als insbesondere auch durch exorbitante Produktionskosten durch Massenszenen, aufwendige Kostüme und Kulissen.³

Die großen Produktionsstudios sind zwar grundsätzlich im Stande, außergewöhnlich anspruchsvolle, komplexe und kostspielige Projekte zu

² Vgl. Hany Farid, „Photo Tampering throughout History“. (online). (Für diese und alle folgenden Angaben von Internetquellen in den Fußnoten wird auf die vollständige Nennung im Literaturverzeichnis verwiesen.)

³ Vgl. William Wyler, *Ben Hur* (USA, 1959); Anthony Mann, *El Cid* (Italien, 1961); Joseph L. Mankiewicz, *Cleopatra* (USA, 1963).

realisieren, in denen historische oder fantastische Welten künstlich erzeugt werden. Dennoch ist der Preis einer Produktion eine für den Unternehmenserfolg kritische Größe, und preiswertere Produktionen senken erheblich das Risiko eines Filmprojekts. Die seit den neunziger Jahren einsetzende Digitalisierung eröffnet hier die Möglichkeit, entweder Filme preiswerter zu produzieren oder zu einem vergleichbaren Preis dem Publikum wesentlich aufwendigere Effekte zu präsentieren. Im Zuge dieser Entwicklung setzten einige Unterhaltungsproduktionen der letzten 25 Jahre die Maßstäbe des technisch Machbaren.

Ein bekanntes Beispiel ist der Film *Jurassic Park* von 1993, in dem Dinosaurier mit Hilfe von Computeranimationen realistisch dargestellt wurden, ein für das Publikum zu der damaligen Zeit völlig neuartiges Spektakel, weswegen in Deutschland in vielen Kinos strikte Alterskontrollen durchgeführt wurden.⁴ Trotz der bahnbrechenden Effekte war der Film mit Produktionskosten von circa 63 Millionen Dollar recht preiswert und ausgesprochen profitabel für das Filmstudio.

Ein weiteres bekanntes Beispiel ist der Film *The Matrix* von 1999.⁵ In der möglicherweise bekanntesten Szene des Films weicht der Held in Zeitlupe den Projektilen einer Pistole aus, während die Kamera um die Szene herumfährt. Auch dieser Effekt baute auf der engen Verzahnung von traditioneller Aufnahmetechnik mit moderner Computeranimation auf.

Ein kurz darauf vorgestelltes Beispiel kommt nicht aus den großen Produktionsstudios, aber durchaus aus der Unterhaltungsindustrie. Während des Finales der American Football-Meisterschaft Superbowl im Jahr 2001 kam das sogenannte ‚EyeVision-System‘ einer Forschungsgruppe rund um Takeo Kanade zum Einsatz. Hierbei handelte es sich um ein komplexes Netzwerk von Kameras, die das Spielfeld aus vielen Positionen gleichzeitig beobachteten. Das System machte es möglich, den Effekt aus dem Film *The Matrix* live auf eine reale Spielszene anzuwenden. So konnte beispielsweise eine Szene in der Wiederholung kurz angehalten werden, die Kameraperspektive rotierte für den Betrachter flüssig und ohne sichtbare Schnitte um den Spieler, und der Spielzug konnte nunmehr von der neuen Kameraperspektive aus weiter abgespielt werden. Die wesentliche technische Neuerung bestand darin, dass das System während des Spiels zuverlässig und in Echtzeit, das heißt, während des Spiels, funktionieren musste.

⁴ Vgl. Steven Spielberg, *Jurassic Park* (USA, 1993).

⁵ Vgl. Lilly Wachowski (as Andy Wachowski) and Lana Wachowski (as Larry Wachowski), *The Matrix* (USA, 1999).

Kurz darauf wurde mit dem Charakter des Gollum in der Filmproduktion *Der Herr der Ringe* 2001 ein komplett Computer-animiertes menschenähnliches Wesen gezeigt.⁶ Zwar war es noch ein ‚menschlicher‘ Schauspieler, der seine Mimik veränderte und der gestikulierte, aber das eigentliche Erscheinungsbild von Gollum wurde komplett im Computer gerechnet.

Nach diesen technischen Durchbrüchen rund um die Jahrtausendwende hat sich gezeigt, dass es grundsätzlich möglich ist, beliebige Inhalte künstlich zu erzeugen. Viele nachfolgende technische Arbeiten konzentrieren sich darauf, virtuelle Inhalte mit geringerem Aufwand zu realisieren, entweder in Bezug auf den händischen Aufwand oder die benötigte Rechenzeit.

Seit ungefähr 2010 hat sich diese Entwicklung substanziell beschleunigt durch zwei ungefähr zeitgleiche Durchbrüche. Einerseits wurden Graphikarten deutlich schneller, wodurch die verfügbare Rechenkraft für spezielle Rechenoperationen vervielfacht wurde. Andererseits wurden neue Methoden im maschinellen Lernen entwickelt, die von diesen speziellen Rechenoperationen in bisher nicht dagewesener Effektivität Gebrauch machen. Die hieraus entstandenen Methoden werden typischerweise unter dem Schlagwort ‚*Deep Learning*‘ geführt.

Aus dieser Entwicklung entstand unter anderem in zeitlich kürzester Folge eine Vielzahl von Methoden, um automatisiert Videos zu editieren. Besondere Aufmerksamkeit über die Grenzen der wissenschaftlichen Gemeinschaft hinaus haben zum Beispiel die Methoden ‚*Face2Face*‘ und ‚*DeepFakes*‘ erfahren. Beide Methoden zielen darauf ab, Gesichter in Videos zu verändern.⁷

Face2Face ist eine Methode aus der Computergraphik, um die Mimik eines Schauspielers auf eine Zielperson in einem anderen Video zu übertragen. Das Verfahren unterschied sich zum Zeitpunkt der Veröffentlichung 2016 von vergleichbaren Ansätzen dadurch, dass es in Echtzeit lief, das heißt, während das Video der Zielperson spielte.

DeepFakes ist eine Methode aus dem maschinellen Lernen. Sie überträgt auch ein Gesicht von einer Person auf eine andere, hierbei werden allerdings die Charakteristika von der Quellperson übertragen, das heißt, Form und Hautfarbe, während die Mimik der Zielperson beibehalten wird.

⁶ Vgl. Peter Jackson, *Der Herr der Ringe* (USA und Neuseeland, 2001).

⁷ Vgl. Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt and Matthias Nießner, „Face2Face: Real-Time Face Capture and Reenactment of RGB Videos“, IEEE Conference on Computer Vision and Pattern Recognition, o. O. 2016, 2387-2395 und Anonymous, „GifFakes“. (online).

Zweifelhaften Ruhm haben *DeepFakes* erfahren, als im Internet Videos von Dritten veröffentlicht wurden, die Gesichter bekannter Schauspielerinnen auf Porno-Darstellerinnen überblendet hatten.

Während die bahnbrechenden Effekte großer Filmproduktionen einen erheblichen Einsatz von Geld und Arbeitskraft voraussetzten, können viele Methoden aus dem maschinellen Lernen mit vergleichsweise geringem Ressourceneinsatz verwendet werden. Die Ergebnisse erreichen visuell in vielen Fällen nicht die Qualität der sehr teuren Produktionen, allerdings sind sie mit technischem Sachverstand oftmals schon auf einem einzelnen Heimrechner erzielbar. Hiermit wurde durch die fortschreitende Automatisierung eine substanzielle Reduktion des Aufwands erreicht, um bestimmte Formen von Bild- und Videomanipulationen durchzuführen, und es ist im Allgemeinen zu erwarten, dass diese technische Entwicklung sich in den nächsten Jahren fortsetzt.

Beispiele für technische Prüfverfahren

Algorithmische Methoden zur Prüfung von Bild- und Videomaterial können entweder gezielt nach Spuren eines bekannten Eingriffs forschen oder ‚blind‘ nach Irregularitäten in dem Medium suchen. Die meisten Methoden verarbeiten Bilder auf eine andere Art als das menschliche Auge. Die meisten Prüfverfahren sind statistische Verfahren, mit denen Beziehungen von benachbarten Bildpunkten analysiert werden. Ein Eingriff in das Bild führt zu statistischen Abweichungen zwischen benachbarten Bildpunkten, die durch die Analyse hervorgehoben werden. Unterschiede zwischen benachbarten Bildpunkten sind für das menschliche Auge kaum wahrnehmbar und daher auch für die Entwicklung von visuell glaubwürdigen Manipulationsmethoden ohne Bedeutung. Daher ist die zunehmende visuelle Qualität der Manipulationsmethoden für die Prüfverfahren kaum von Belang.

Viele forensische Verfahren nehmen die Bildentstehung in der Kamera als Ansatzpunkt für die Prüfung. Dies ist schematisch in Abbildung 1 illustriert. Die abgebildete Szene muss physikalisch konsistent sein, und die Kamera-linse, der Sensor und die Kamerasoftware verändern die Abbildung der Szene auf jeweils charakteristische Weise. Beispielsweise erzeugen viele Linsen typische Farbverzeichnungen, die in dem Bild validiert werden können. Das Ausgabebild einer Kamera ist typischerweise in dem sogenannten ‚JPEG-Verfahren‘ codiert, welches ebenfalls leichte Bildfehler hinzufügt, die gezielt validiert werden können.

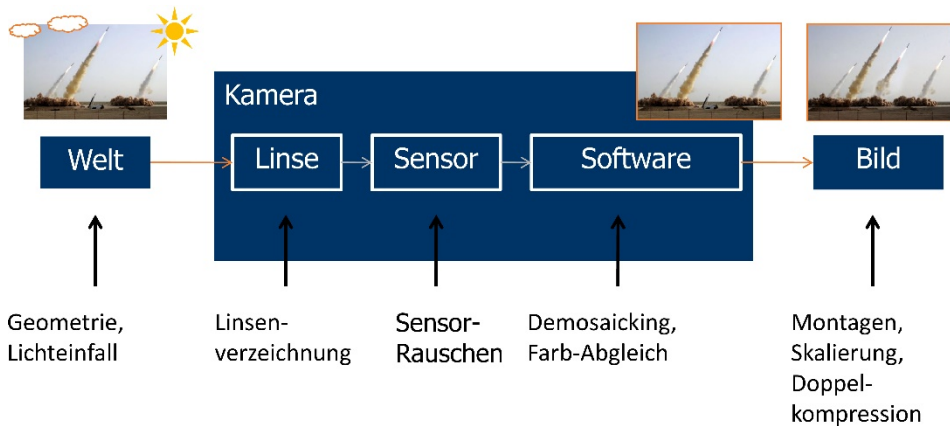


Abbildung 1: Schematische Darstellung der Bildentstehung in der Kamera. Jeder Schritt bietet Ansätze zu der Konsistenzprüfung eines digitalen Bilds.

Relevante Faktoren für den Erfolg eines Prüfverfahrens sind der technische Zustand und die Historie eines digitalen Bilds. Beispielsweise sind Bilder von niedriger Qualität, also stark verkleinerte und stark komprimierte Bilder, grundsätzlich schwieriger zu analysieren, da hierdurch die statistisch relevanten Nachbarschaftsbeziehungen leicht zerstört werden. Aus diesem Grund ist es eine der laufenden Herausforderungen der Forschung, mit Bildern aus sozialen Netzwerken zu arbeiten, da diese oftmals von dem Netzbetreiber stark verkleinert und komprimiert werden.

Im Folgenden werden exemplarisch einige klassische Prüfverfahren vorgestellt, die gezielt nach Spuren einer bestimmten Form von Eingriff suchen.

Ein beliebtes Werkzeug in Bildverarbeitungsprogrammen ist der Kopierstempel. Mit diesem Werkzeug kann eine Bildregion innerhalb eines Bildes vervielfältigt werden, was auch als ‚Copy-Move-Manipulation‘ bezeichnet wird. So ist es beispielsweise möglich, ein den Bildeindruck störendes Flugzeug am Himmel zu entfernen, indem ein Stück freien Himmels über das Flugzeug gelegt wird. Ebenfalls beliebt ist dieses Werkzeug, um visuell besonders unstrukturierte Bildelemente zu vergrößern, beispielsweise indem dichte Vegetation erweitert wird.

Aus algorithmischer Sicht ist das Finden von *Copy-Move-Manipulationen* ein Suchproblem.⁸ Die Detektionsmethode sucht nach Regionen im Bild, die besonders ähnlich zueinander sind. Wenn dies einen größeren zusammenhängenden Bereich betrifft, dann wird davon ausgegangen, dass die Selbstähnlichkeit nicht zufällig ist, und die betreffenden Bildregionen werden für den Analysten markiert zur einfacheren Beurteilung.

Dies ist insbesondere dann hilfreich, wenn die kopierte Region visuell unauffällig ist, wie beispielsweise eine Kopie in der relativ uniformen Himmelsregion. Da die Bildpunkte selbst in visuell uniformen Bereichen selten exakt gleich sind, sondern stets leicht variieren, ist es für ein Suchverfahren oftmals möglich, diese für das Auge unsichtbaren leichten Variationen zu finden. Sobald der Analyst auf die betreffenden Quell- und Zielregionen der mutmaßlichen Kopie von dem Algorithmus aufmerksam gemacht wurde, können diese Abweichungen manuell überprüft werden, um Fehlerkennungen auszuschließen.

Ein weiteres Verfahren zielt auf die Erkennung von vergrößerten oder rotierten Bildregionen ab.⁹ Wenn es sich bei einem Bild um eine aus mehreren Quellbildern komponierte Collage handelt, dann ist es häufig der Fall, dass zur Erstellung der Collage ein eingefügtes Bildelement vergrößert oder rotiert werden muss. In Bildverarbeitungsprogrammen sind beide Arbeitsschritte als sogenannte ‚*Resampling*-Verfahren‘ implementiert. Hierbei werden neue Bildpunkte aus der Nachbarschaft der bestehenden Bildpunkte erzeugt. Das einfachste Beispiel ist eine Vergrößerung um den Faktor zwei: Hier muss zwischen je zwei bestehenden Bildpunkten ein neuer Bildpunkt eingefügt werden, um die Bildgröße zu verdoppeln. Eine einfache Möglichkeit, die Farbe des neuen Bildpunktes zu wählen, ist es, den Durchschnitt der Nachbarn zu berechnen. Dies ist Beispiel für eine *Resampling*-Operation.

Resampling führt zu statistischen Abhängigkeiten zwischen benachbarten Bildpunkten. In dem obigen Beispiel ist der eingefügte Bildpunkt komplett durch seine Nachbarn bestimmt. Diese Abhängigkeiten können geschätzt und visualisiert werden. Der erste Ansatz hierzu wurde 2004 von Alin Popescu und Hany Farid präsentiert und in vielen Nachfolgearbeiten sowohl vereinfacht als auch verbessert.

⁸ Vgl. Vincent Christlein, Christian Riess, Johannes Jordan, Corinna Riess and Elli Angelopoulou, „An Evaluation of Popular Copy-Move Forgery Detection Approaches“, *IEEE Transactions on Information Forensics and Security* 7.6 (2012): 1841-1854.

⁹ Vgl. Alin Popescu and Hany Farid, „Exposing Digital Forgeries by Detecting Traces of Resampling“, *IEEE Transactions on Signal Processing* 53.2 (2005): 758-767.

Die Methode erzeugt ein Grauwertbild, das *p-Map* („probability map“) genannt wird. Die *p-Map* bildet die geschätzte statistische Abhängigkeit zwischen benachbarten Pixeln über Helligkeitsstufen ab. Wenn ein Bild aus mehreren Komponenten zusammengesetzt ist, dann zeigt sich ein fremdes Bildelement in der *p-Map* oftmals in sehr hellen oder sehr dunklen Bereichen oder in einem von der Umgebung abweichenden Gittermuster.

Dieses grundsätzlich sehr starke Verfahren ist allerdings äußerst anfällig für nachträgliche globale Veränderungen im Bild. Wird das Bild als ganzes zum Beispiel verkleinert, dann verdecken die Spuren dieser Verkleinerung die Spuren einer möglicherweise davor durchgeführten Collage von Bildelementen.

Die Familie der sogenannten physikbasierten Ansätze prüft physikalische Rahmenbedingungen der abgebildeten Szene, wie zum Beispiel die Richtung des Lichteinfalls auf Objekten, die Konsistenz von optischen Fluchtpunkten in der Szene, die Richtung des Schattenwurfs und die Konsistenz der reflektierten Farbe des Lichts. Unter der Annahme, dass ein Bild eine reale Szene präsentiert, müssen die in der Abbildung gezeigten Objekte den Gesetzen der Physik in den genannten Punkten gehorchen. Für ein künstlich zusammengesetztes Bild ist diese Bedingung nicht in jedem Fall erfüllt.

Als konkretes Beispiel wird die Prüfung der Richtung des Lichteinfalls herausgegriffen.¹⁰ Besonders gut einsetzbar ist dieses Verfahren für Außenaufnahmen, in denen die Objekte der Szene von der Sonne bestrahlt werden. Auf Grund der großen Entfernung zwischen Sonne und Erde ist die Richtung des Lichteinfalls auf den Objekten so ähnlich, dass sie unter üblichen Kameraaufnahmen faktisch identisch ist. Für die menschliche Wahrnehmung ist es jedoch schwierig, die Richtung des Lichteinfalls exakt zu bestimmen, weshalb ein mathematisches Verfahren zur Berechnung der Lichtrichtung hier eine wesentliche quantitative Unterstützung leisten kann.

Zur Schätzung der Richtung des Lichteinfalls kann im einfachsten Fall die Verteilung der Helligkeit entlang der Kontur eines Objekts betrachtet werden. Dies liefert über ein physikalisches Modell die Projektion der Lichtrichtung in der Bildebene. Über den Vergleich des Lichteinfalls auf mehreren Objekten kann die Konsistenz der Szene geprüft werden.

¹⁰ Vgl. Micah Johnson and Hany Farid, „Exposing Digital Forgeries in Complex Lighting Environments“, *IEEE Transactions on Information Forensics and Security* 2.3 (2007): 450-461.

Kompliziertere Varianten dieses Verfahrens setzen voraus, dass die Geometrie der zu prüfenden Objekte bekannt ist. Beispielsweise lässt sich aus Abbildungen von Gesichtern relativ zuverlässig ein dreidimensionales Modell des Gesichts schätzen. Mit dieser geometrischen Zusatzinformation ist die Schätzung nicht auf die Bildebene beschränkt, sondern kann im dreidimensionalen Raum durchgeführt werden, so dass auch Lichtquellen, die hinter dem Fotografen platziert sind, unterschieden werden können. Ein besonderer Vorteil derartiger physikalischer Ansätze ist, dass sie kaum betroffen sind von den Schwächen der statistischen Verfahren: Verkleinerte oder stark komprimierte Bilder sind kein Hinderungsgrund für die Anwendung der Methode. Eine Schwäche dieser Ansätze ist jedoch, dass sie nur auf bestimmte Szenen anwendbar sind, in denen die zu Grunde liegenden physikalischen Annahmen getroffen werden können. Beispielsweise ist die Schätzung der Lichtrichtung in Innenräumen typischerweise viel weniger aussagekräftig, da hier oftmals eine Vielzahl von sehr lokalen Lichtquellen vorliegt, so dass auch in Originalbildern unterschiedliche Objekte oft aus unterschiedlichen Richtungen beleuchtet werden.

Ein weiteres sehr bekanntes forensisches Verfahren nutzt Erkenntnisse aus der Elektrotechnik aus. Die Fertigung der mikroskopisch kleinen lichtempfindlichen Sensorzellen einer Kamera unterliegt minimalen Schwankungen, was dazu führt, dass sich jede Sensorzelle in der Fläche der lichtempfindlichen Region leicht unterscheidet. Auf größere Flächen treffen statistisch gesehen mehr Photonen auf als auf kleinere Flächen, so dass diese Bildpunkte systematisch etwas heller erscheinen. Dieser Effekt ist als *Photo-Response Non-Uniformity (PRNU)* bekannt. *PRNU*-Variationen sind so schwach, dass sie mit dem Auge nicht wahrnehmbar sind, aber sie sind stark genug, um aus dem digitalen Bild extrahiert werden zu können. Da diese Unterschiede tatsächlich für jeden Bildsensor spezifisch sind, kann ein Bild eindeutig einem Aufnahmegerät zugeordnet werden,¹¹ analog dazu wie eine Pistolenkugel in einer ballistischen Untersuchung einer Pistole zugeordnet werden kann.

Eine Voraussetzung für den Einsatz der Methode ist allerdings das Vorliegen von mehreren Vergleichsbildern aus einer digitalen Kamera. Bei der Prüfung werden die Helligkeitsunterschiede der Bildpunkte mit den Helligkeitsunterschieden in den Vergleichsbildern abgeglichen. Diese Prüfung wird allerdings substanziell verkompliziert, wenn das zu prüfende

¹¹ Vgl. Jan Lukas, Jessica Fridrich and Miroslav Goljan, „Digital Camera Identification from Sensor Pattern Noise“, *IEEE Transactions on Information Forensics and Security* 1.2 (2006): 205-214.

Bild nur in einer verkleinerten oder beschnittenen Form vorliegt, denn in diesem Fall muss zuerst eine Passung des zu untersuchenden Bilds mit den Vergleichsbildern gefunden werden, was je nach Degradation des Bildes oftmals ein sehr aufwendiger Prozess oder sogar gänzlich unmöglich ist.

Den oben diskutierten Videomanipulations-Methoden mittels moderner Methoden der Computergraphik oder des maschinellen Lernens, beispielsweise *Face2Face* oder *DeepFakes*, kann ebenfalls mit statistischen Werkzeugen aus dem maschinellen Lernen in einer gezielten Analyse begegnet werden. Hierfür ist es erforderlich, einem maschinellen Lernverfahren Beispiele für Originalvideos und mit der jeweiligen Methode manipulierte Videos zu präsentieren. Das Lernverfahren sucht daraufhin charakteristische Unterschiede in der Nachbarschaft der Bildpunkte automatisch heraus, um in einem nächsten Schritt selbstständig diese Form der Manipulation erkennen zu können. Ein Beispiel dafür ist unsere kürzlich vorgestellte *FaceForensics*-Methode.¹² Gezielte Prüfverfahren dieser Art arbeiten sehr zuverlässig, auch auf Videos von schlechter Qualität, sofern geeignete Beispiele vorliegen. Die Verfahren eignen sich jedoch nur zur Erkennung einer spezifischen, bekannten Manipulationsmethode, und es ist aktuell zum Beispiel nicht möglich, mit einem Verfahren, das auf *Face2Face* eingestellt wurde, ein *DeepFake*-Video zu erkennen. Diese Einschränkung zu mindern oder gänzlich aufzulösen, ist derzeit eine der wichtigen Forschungsfragen in der Multimediaforensik.

Eine weitere Familie von Verfahren umgeht grundsätzlich die Einschränkung, dass gezielte Prüfverfahren nur Eingriffe einer bestimmten Art erkennen können. Die sogenannten ‚blinden‘ forensischen Methoden konzentrieren sich auf die Modellierung eines konsistenten Originalbilds. Jegliche Abweichung hiervon wird dem Analysten präsentiert. Damit sind blinde Ansätze komplett unabhängig von der Art des Eingriffs. Diese Flexibilität wird allerdings damit erkauft, dass diese Methoden meist weniger spezifisch sind als Verfahren, die gezielt nach einer bestimmten Art von Bildfehler suchen.

In einer praktischen Analyse werden meist erst die sogenannten ‚EXIF-Metadaten‘ eines Bildes untersucht. Hierbei handelt es sich um zusätzliche Informationen, typischerweise von dem Aufnahmegerät, die neben dem eigentlichen Bildinhalt mit in der Datei gespeichert werden. Häufige zusätzliche Informationen sind der Name des Geräts, Datum und

¹² Vgl. Andreas Rössler, Davide Cozzolino and Luisa Verdoliva, „FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces“, arXiv preprint 1803.09179, 2018.

Uhrzeit der Aufnahme, in Smartphone-Bildern manchmal auch die GPS-Koordinaten der Aufnahme. Des Weiteren werden oftmals auch die optischen und algorithmischen Einstellungen der Kamera in den Metadaten gespeichert. Diese Informationen können einen ersten Anhaltspunkt für die Aufnahmesituation bereitstellen. Für einen Manipulator mit Täuschungsabsicht ist es sehr einfach, die Metadaten zu editieren, weswegen der Beweiswert der Metadaten oftmals als gering eingeschätzt wird. Allerdings ist es umgekehrt so, dass inkonsistente Metadaten aus der Geschichte des jeweiligen Bildes heraus erklärbar sein müssen, und in diesem Sinn die Metadaten durchaus interessante Hinweise liefern können.

Auch die Nachbarschaften von Bildpunkten können auf Inkonsistenzen hin untersucht werden, ohne eine bestimmte Art von Manipulation anzunehmen. Ein hierzu vor kurzem vorgestelltes Verfahren ist *NoisePrint*.¹³ Dieses basiert darauf, kleine Variationen von benachbarten Bildpunkten in einem statistischen Modell zusammenzufassen. Das Bild wird dabei in viele kleine Teilregionen unterteilt, und jede Region wird unabhängig in einem Modell beschrieben. In einem zweiten automatischen Schritt wird die Ähnlichkeit dieser Modelle untersucht. Finden sich in einem Bild mehrere unterschiedliche Statistiken, dann werden diese Abweichungen in der Ausgabe hervorgehoben.

Diese sehr allgemeine Form einer statistischen Bildbetrachtung fasst in Regionen, die nicht bearbeitet wurden, die Einflüsse verschiedener Bearbeitungsschritte der Kamera und der Bildcodierung in einem Modell zusammen. Die Berechnung des Abstands zweier Modelle wird nicht durch die Präsentation von manipulativen Eingriffen angereichert, sondern im Wesentlichen durch Betrachtung von Bildern aus unterschiedlichen Kameras parametrisiert. In praktischen Experimenten zeigt sich jedoch, dass dies oftmals genügt, um Bildstörungen aus einem manipulativen Eingriff zuverlässig als Abweichungen von dem gelernten Kameramodell zu erkennen.

¹³ Vgl. Davide Cozzolino and Luisa Verdoliva, „NoisePrint: a CNN-Based Camera Model Fingerprint“, arXiv preprint 1808.08396, 2018.

Fazit

Durch den rapiden Fortschritt in der digitalen Automatisierung wird auch die Erzeugung visuell glaubwürdiger, künstlicher Bilder und Videos zunehmend einfacher. Hierdurch wird die Erstellung von Bild- und Video-manipulationen gewissermaßen ‚demokratisiert‘, und in letzter Zeit ist häufig in den Medien die Frage aufgeworfen worden, ob es möglich ist, einem Bild oder Video noch zu vertrauen.

In diesem Artikel wurden exemplarisch einige Methoden aus dem Bereich der Multimediaforensik aufgezeigt, mit denen die Prüfung von Bildern und Videos an Hand technischer Merkmale durchgeführt werden kann. Obwohl der Fortschritt in der digitalen Automatisierung auch die Forschung in der Multimediaforensik laufend vor neue Herausforderungen stellt, ist es insgesamt durchaus jetzt und auch in Zukunft möglich, ein Bild oder Video auf Ursprung und Echtheit hin zu überprüfen.

Allerdings können diese Methoden immer nur technische Indizien für einen Eingriff in ein digitales Medium liefern. Auf Grund der sogenannten ‚Semantischen Lücke‘ in der künstlichen Intelligenz sind Maschinen ohne Bewusstsein grundsätzlich nicht in der Lage, die Absicht hinter einem Eingriff zu erkennen. Damit ist es einer Maschine nicht möglich zu ermitteln, ob ein technischer Eingriff mit einer Täuschungsabsicht vorgenommen wurde. Hieraus folgt, dass zum Zweck der Manipulationserkennung stets ein Experte die von der Maschine bereitgestellten Indizien sichten und mit dem vorliegenden Fall abgleichen muss, um zu ermitteln, ob es sich bei dem Eingriff tatsächlich um eine bössartige Manipulation handelt.

Literaturverzeichnis

- Christlein, Vincent, Christian Riess, Johannes Jordan, Corinna Riess and Elli Angelopoulou. „An Evaluation of Popular Copy-Move Forgery Detection Approaches“. *IEEE Transactions on Information Forensics and Security* 7.6 (2012): 1841-1854.
- Cozzolino, Davide and Luisa Verdoliva. „NoisePrint: a CNN-Based Camera Model Fingerprint“. arXiv preprint 1808.08396, 2018.
- Johnson, Micah and Hany Farid. „Exposing Digital Forgeries in Complex Lighting Environments“. *IEEE Transactions on Information Forensics and Security* 2.3 (2007): 450-461.
- Lukas, Jan, Jessica Fridrich and Miroslav Goljan. „Digital Camera Identification from Sensor Pattern Noise“. *IEEE Transactions on Information Forensics and Security* 1.2 (2006): 205-214.
- Popescu, Alin and Hany Farid. „Exposing Digital Forgeries by Detecting Traces of Resampling“. *IEEE Transactions on Signal Processing* 53.2 (2005): 758-767.
- Thies, Justus, Michael Zollhöfer, Marc Stamminger, Christian Theobalt and Matthias Nießner. „Face2Face: Real-Time Face Capture and Reenactment of RGB Videos“. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2387-2395.
- Tinterow, Gary, Philip Conisbee and Hans Naef. Ed. *Portraits by Ingres: Image of an Epoch*. New York: The Metropolitan Museum of Art, 1999. 140-143.
- Rössler, Andreas, Davide Cozzolino and Luisa Verdoliva. „FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces“. arXiv preprint 1803.09179, 2018.

Internetquellen

- Anonymous. „GifFakes“, <http://www.reddit.com/r/GifFakes/>, [zuletzt besucht 03. Dezember 2018].
- Farid, Hany. „Photo Tampering throughout History“, <http://pth.izitru.com>, [zuletzt besucht 03. Dezember 2018].

Filme

Jackson, Peter. *Der Herr der Ringe*. USA und Neuseeland, 2001.

Mankiewicz, Joseph L. *Cleopatra*. USA, 1963.

Mann, Anthony. *El Cid*. Italien, 1961.

Spielberg, Steven. *Jurassic Park*. USA, 1993.

Wachowski, Lilly (as Andy Wachowski) and Lana Wachowski (as Larry Wachowski). *The Matrix*. USA, 1999.

Wyler, William. *Ben Hur*. USA, 1959.