Digitale Forensik
– Spuren in Digitalfotos –

Matthias Kirchner
Technische Universität Dresden
s9296871@mail.inf.tu-dresden.de

Mit der mehr und mehr digitalisierten Fototechnik ist es heute ohne Vorwissen nahezu jedem möglich, Bilder zu manipulieren. Bekanntgewordene Fälle in den Medien haben auch die Öffentlichkeit für dieses Thema sensibilisiert. Verfahren der digitalen Bildforensik bieten die Möglichkeit, die Authentizität eines Bildes auch ohne Zugriff auf das Original zu überprüfen. Mit diesem Artikel sollen stellvertretend zwei Ansätze vorgestellt werden.

1 Digitale Bilder – Ein Abbild der Wirklichkeit?

Das Bestreben Fotos zu manipulieren oder zu fälschen ist in etwa so alt wie die Fotografie selbst. Bilder stellen ein effektives, leicht verständliches und einprägsames Kommunikationsmittel dar: Es bedarf keiner besonderen Sprache um eine Fotografie zu verstehen. Das erkannten auch und gerade Politiker sehr schnell. Besonders bekannt sind die von Stalin initiierten Bildfälschungen. Mit der Zeit unliebsam gewordene Parteigenossen verschwanden nicht nur im wirklichen Leben, sondern auch auf Fotografien, die Stalin mit den betreffenden Personen abbildeten. In der heutigen multimedialen Gesellschaft ist es für Personen im öffentlichen Blickfeld umso wichtiger, im wahrsten Sinne des Wortes ein gutes und authentisches Bild abzugeben. Wohl aus diesem Grund wurde etwa in Zeiten von Arbeitsplatzabbau und steigenden Managergehalten in der öffentlichen Kritik bei einem repräsentativen Foto des Siemens-Chefs Klaus Kleinfeld eine Rolex-Armbanduhr wegretuschiert.

Mit der immer stärkeren Digitalisierung der Fototechnik ist es heute praktisch jedem möglich, Bilder zu manipulieren. Ausgefeilte und verhältnismäßig leicht bedienbare Fotobearbeitungssoftware erlaubt das Erstellen überzeugender Fälschungen mit relativ wenig Aufwand und kaum Vorwissen. Die Internetseite `www.worth1000.com` bietet beispielsweise in Form von Wettbewerben ein Forum zum Publizieren selbst erstellter digitaler Manipulationen. Ein weiteres Beispiel ist die Seite `www.fakeorfoto.com`, die mit der Frage spielt, ob mit dem Computer generierte Bilder von tatsächlichen Fotografien unterscheidbar sind.

Insbesondere bei dem „normalen“ Endnutzer besteht sicherlich ein Interesse an geringfügigen verschönernden Manipulationen der selbst aufgenommenen Bilder. Dies kann bei einem maßvollen Einsatz durchaus sinnvoll sein und zeigt sich auch daran, dass mehr und mehr solcher (allerdings nicht immer sinnvollen) Funktionalität direkt in die Kameras eingebaut wird [7].

Demgegenüber ist es mit Sicherheit als problematisch zu bewerten, wenn mit Hilfe von Bildmanipulationen bestimmte Aussagen verstärkt oder gar verändert werden sollen. Ein recht aktuelles Beispiel ist die in Abbildung 1 gezeigte Fälschung aus dem Libanon-Krieg. Die Ausmaße eines israelischen Luftangriffes wurden auf dem oberen Bild (mit zugegebenermaßen einfachen Mitteln) verschärft dargestellt. Nach Bekanntwerden der Fälschung wurde der für die Nachrichtenagentur Reuters arbeitende Fotograf entlassen. Wie viele andere Beispiele zeigen, wäre es fahrlässig zu glauben, dass es sich bei Manipulationen an Bildern aus den täglichen Nachrichten um Einzelfälle handelt. In letzter

Abbildung 1: Manipulation eines Bildes aus dem Libanon-Krieg.

Zeit sind durch den Fall des südkoreanischen Klonforschers Hwang Woo-Suk auch Bildfälschungen in wissenschaftlichen Veröffentlichungen verstärkt in das öffentliche Blickfeld geraten. Nachdem bekannt geworden war, dass es sich bei dessen Publikation im Magazin *Science* zu Fortschritten in der Stammzellenforschung um eine Fälschung handelte, trat der angesehene Professor von allen Ämtern zurück. Die gefälschten Ergebnisse wurden dabei maßgeblich durch manipulierte Abbildungen unterstützt. Dass es sich auch in diesem Bereich bei Bildmanipulationen nicht um Ausnahmen handelt, zeigt, dass Schätzungen zufolge bei einem Fünftel aller akzeptierten Artikel des *Journal of Cell Biology* Abbildungen aufgrund von unzulässigen Manipulationen nachgefordert werden müssen [5].

2 Digitale Bildforensik

Zum Schutz der Authentizität von digitalen Bildern wird gerne auf *digitale Wasserzeichen* verwiesen. Diese werden als in der Regel äußerlich nicht wahrnehmbares Signal in das Bild eingebettet. Bei einer Manipulation des Bildes wird auch das eingebrachte Wasserzeichen verändert. Somit kann die Echtheit des Mediums dadurch überprüft werden, ob das ausgelesene Wasserzeichen mit dem eingebetteten übereinstimmt. Je nach Szenario und zulässigen Manipulationen am Medium spricht man von robusten, fragilen oder semifragilen Wasserzeichen. Voraussetzung dafür ist jedoch, dass das Wasserzeichen im unmittelbaren Entstehungsprozess des Mediums eingebracht wird. Dazu wird z.B. in [1] eine so genannte *Secure Digital Camera* vorgeschlagen. Diese bettet ein Wasserzeichen aus biometrischen Daten des Urhebers, einer Signatur des Bildes und Informationen über die Kamera im Entstehungsprozess der Fotografie ein. Die Einschränkung auf speziell ausgerüstete Aufnahmegeräte lässt den Einsatz von Wasserzeichen als Verfahren zur Überprüfung der Echtheit von digitalen Bildern in der Praxis allerdings als nur bedingt geeignet erscheinen. Ein weiterer nicht zu vernachlässigender Punkt ist die Annahme zur Sicherheit der Wasserzeichen-Verfahren und die Frage, ob ein digitales Wasserzeichen aus dem Trägermedium entfernt und nach einer Manipulation wiederum eingefügt werden kann (z.B. [2, 8]).

Der Begriff der *digitalen Bildforensik* umfasst dagegen ein vergleichsweise neues Forschungsgebiet im Bereich der digitalen bzw. digitalisierten Bilder. Er subsumiert Verfahren zum Nachweis von Manipulationen an oder der Fälschung von solchen Mediendaten. Der Kernpunkt dabei ist, dass das Original zu keinem Zeitpunkt als bekannt vorausgesetzt wird. Stattdessen basieren die Verfahren auf statistischen Analysen des (potentiell gefälschten) Bildes auf Grundlage von Modellen der Digitalisierungstechnik oder des Bildinhaltes. Die Detektion von Fälschungen mit Hilfe von Modellen der Digitalisierungstechnik beruht dabei auf der Annahme, dass sich bestimmte Charakteristika des Entstehungsprozesses eines digitalen Bildes in diesem nachweisen lassen. Modelle des Bildinhaltes arbeiten hingegen mit der Tatsache, dass es sich bei einem digitalen Bild um ein Abbild der „Wirklichkeit“ handelt und versuchen diese entsprechend zu modellieren. Unter der Annahme, dass Manipulationen am Bild auch die dem verwendeten Modell zugrunde liegenden statistischen Eigenschaften des Bildes

ändern, kann dessen Echtheit auch ohne Zugriff auf das Original oder dem aktiven Einbetten eines Wasserzeichens untersucht werden. Verfahren der Bild-Forensik werden daher auch als *blind* und *passiv* bezeichnet [4].

Grundsätzlich beschäftigt sich die digitale Bildforensik mit der

- ▷ Identifikation des Bildursprunges sowie der
- ▷ Detektion von Bildmanipulationen.

Die Frage nach dem Bildursprung beinhaltet dabei neben der Unterscheidung zwischen echten und computergenerierten Bildern auch die Frage nach dem konkreten Aufnahmegerät. Vornehmlich in den letzten drei bis vier Jahren wurden in der Literatur zu jedem dieser beiden Punkte verschiedenartige Ansätze präsentiert. Stellvertretend sollen im Folgenden je ein Verfahren zur Digitalkamera-Identifikation und zur Detektion von Manipulationen an digitalen Bildern näher vorgestellt werden.

3 Digitalkamera-Identifikation mittels Sensorrauschen

Ein sehr zuverlässiges Verfahren zur Identifikation der Digitalkamera, mit der ein Bild aufgenommen wurde, basiert auf dem Rauschen der Bildsensoren [3]. Als Sensoren dienen bei heutigen Kameras meist CCD-Arrays. Jedes Pixel entspricht dabei einem Halbleiterbauelement (CCD - *Charge Coupled Device*), welches eintreffendes Licht in elektrische Signale wandelt. Dabei wird dem resultierenden Signal zwangsläufig Rauschen hinzugefügt, welches grob in einen statischen und einen dynamischen Anteil unterschieden werden kann. Einen großen Anteil am dynamischen Rauschen hat das so genannte Schrotrauschen. Es ist u.a. temperaturabhängig und spiegelt die Tatsache wider, dass der Stromfluss an Halbleiterelementen ein stochastischer Prozess ist. Der statische Anteil (*Pattern Noise*) wird dagegen in jedem mit einer Kamera aufgenommenen Bild in verhältnismäßig gleichem Maße auftreten. Das statische Rauschen setzt sich aus Dunkelströmen und Pixel-zu-Pixel-Ungleichheiten zusammen. Ersterer Anteil wird in der englischen Fachliteratur als *Fixed Pattern Noise* bezeichnet und hat additiven Charakter. Die Pixel-zu-Pixel-Ungleichheiten (*Photo Response Non-Uniformity*) bezeichnen geringfügige im Herstellungsprozess eingefügte Unterschiede der einzelnen Bauelemente. Diese hinterlassen in jedem Bild ein charakteristisches Rauschmuster der Kamera. Dieses kameraspezifische Rauschen ist hochfrequenter Natur und wird als normalverteilt angenommen. Zur Identifikation der verwendeten Kamera wird es mit einem geeigneten Rauschfilter aus dem Bild extrahiert und dessen Korrelation mit bekannten Referenzmustern bestimmt. Bezeichnet X ein zu untersuchendes Bild, so ist das mit dem Filter F extrahierte Rauschmuster durch die Gleichung $X - F(X)$ beschrieben. Für dieses kann dann die Korrelation ρ mit jedem der bekannten Referenzmuster P_i bestimmt werden, wobei das Muster P_i der Kamera C_i zugeordnet ist:

$$\begin{aligned}\rho_i(X) &= corr(X - F(X), P_i) \\ &= \frac{(X - F(X) - E[X - F(X)]) \cdot (P_i - E[P_i])}{\|X - F(X) - E[X - F(X)]\| \cdot \|P_i - E[P_i]\|}\end{aligned}$$

Aus diesem Schema ergeben sich zwei Möglichkeiten der Identifikation:

- ▷ Aus einer Gruppe von Kameras soll die bestimmt werden, mit der am wahrscheinlichsten das vorliegende Bild aufgenommen wurde. Dazu ist lediglich die Kamera zu wählen, deren Referenzmuster am stärksten mit dem extrahierten Muster korreliert.
- ▷ Es soll die Aussage bewertet werden, dass eine spezifische Kamera das vorliegende Bild aufgenommen hat. Dazu ist die Definition von Grenzwerten für ρ nötig, die eine zuverlässige Identifikation erlauben.

Offensichtlich ist eine absolute Aussage zum Ursprung eines Bildes deutlich schwerer.

Bisher unbeantwortet geblieben ist die Frage, auf welche Weise die augenscheinlich wichtigen Referenzmuster der einzelnen Kameras erhalten werden können. Die Autoren in [3] schlagen vor, jeweils den Mittelwert aus einer hinreichend großen Anzahl von extrahierten Rauschmustern zu nutzen. Somit ist es für dieses Verfahren letztlich nicht notwendig, in Besitz der zu untersuchenden Kameras zu sein. In der Praxis sollte die Anzahl der verwendeten Bilder größer als 50 sein.

Mit ihrem Verfahren konnten Lukáš et al. sehr zuverlässig die Digitalkamera bestimmen, mit der ein Bild gemacht wurde. Als Testdatensatz wurden Bilder von insgesamt 9

Kameras verwendet, darunter auch zwei Kameras des gleichen Types. Wie gezeigt werden konnte, ist eine korrekte Zuordnung auch nach einer JPEG-Kompression des Bildes noch möglich. Da die vorgestellte Methode auf lokalen Eigenschaften des Sensorfeldes beruht, stellen jedoch geometrische Transformationen (z.B. Rotation, Abschneiden von Bildbereichen) ein Problem für die Identifikation dar. Die Autoren weisen deshalb darauf hin, dass eine belastbare Aussage zum Ursprung des Bildes in jedem Fall von mehreren Verfahren gestützt werden sollte.

4 Detektion von Re-Sampling

Das Erstellen überzeugender Bildmanipulationen schließt in vielen Fällen das Skalieren, Rotieren oder Verzerren des ganzen Bildes oder einzelner Bildteile ein. Diese unter dem Begriff der geometrischen Transformation zusammengefassten Operationen beinhalten im Allgemeinen eine Umtastung (*Re-Sampling*) auf ein neues Bildgitter und somit einen Interpolationsschritt. Interpolation beschreibt den Prozess der Bestimmung von Funktionswerten zwischen den bekannten Abtastwerten einer Funktion und hinterlässt – wie in [6] gezeigt wurde – im resultierenden Signal nachweisbare Spuren.

Zur Veranschaulichung soll der eindimensionale Fall betrachtet werden. Die Veränderung der Abtastrate um den Faktor p/q eines Signals $\mathbf{x}$ mit m Abtastwerten auf ein Signal $\mathbf{y}$ mit n Abtastwerten erfolgt in drei Schritten. Zunächst wird die Anzahl der Abtastwerte auf pm erhöht. Dabei entspricht jedes p-te Sample einem Abtastwert aus dem Originalsignal; alle anderen Samples haben den Wert 0. Die Punkte des so erhaltenen Signals werden anschließend interpoliert, was als Faltung mit einem Tiefpassfilter beschrieben werden kann. Das resultierende Signal mit n Abtastwerten ergibt sich indem man jeden q-ten Abtastwert der interpolierten Sequenz übernimmt. Da alle drei Schritte linearer Natur sind, können diese in Form einer Matrixmultiplikation zusammengefasst werden,

$$\mathbf{y} = \mathbf{R}_{p/q}\mathbf{x},$$

wobei die $n \times m$ Matrix $\mathbf{R}_{p/q}$ die charakteristische Re-Sampling-Matrix bezeichnet. Im Falle der Verdoppelung der Abtastrate bei linearer Interpolation hat sie z.B. die Form

$$\mathbf{R}_{2/1} = \begin{bmatrix} 1 & 0 & 0 & \\ 0.5 & 0.5 & 0 & \\ 0 & 1 & 0 & \\ 0 & 0.5 & 0.5 & \\ 0 & 0 & 1 & \\ & & & \ddots \end{bmatrix},$$

sodass sich für die Abtastwerte des neuen Signals $\mathbf{y}$ gilt:

$$\begin{aligned} y_{2i-1} &= x_i \\ y_{2i} &= 0.5x_i + 0.5x_{i+1}. \end{aligned}$$

Setzt man diese beiden Gleichungen ineinander ein, folgt

$$y_{2i} = 0.5y_{2i-1} + 0.5y_{2i+1}.$$

Offensichtlich ist jedes Sample mit einem geraden Index ein Linearkombination seiner beiden Nachbarn, woraus sich ein periodisches Korrelationsmuster ergibt. Ähnliche Zusammenhänge lassen sich auch für beliebige andere Faktoren p/q aufstellen, sodass allgemein zur Detektion von Re-Sampling gefragt werden kann: Ist ein Abtastwert eine Linearkombination seiner $2N$ Nachbarn?

$$y_i \stackrel{?}{=} \sum_{k=-N}^{N} \alpha_k y_{i+k}$$

Dabei beschreibt $\boldsymbol{\alpha}$ die skalaren Gewichte. Diese sind ebenso wie das verwendete Interpolationsverfahren im Allgemeinen nicht bekannt. Zur Schätzung nutzen Popescu und Farid daher das Expectation/Maximization(EM)-Verfahren. Dabei handelt es sich um ein iteratives Verfahren, welches im so genannten E-Schritt die Wahrscheinlichkeit dafür schätzt, dass ein Abtastwert mit seinen Nachbarn korreliert. Im M-Schritt wird mit den im E-Schritt bestimmten Wahrscheinlichkeiten ein neuer Schätzwert für $\boldsymbol{\alpha}$ berechnet.

Ähnlich wie im eindimensionalen Fall ergeben sich auch bei geometrisch transformierten Bildern periodische Korrelationen zwischen benachbarten Pixeln. Dies ist beispielhaft für ein Bild der Frauenkirche Dresden in Abbildung 2 dargestellt. Dieses wurde um 5 % und 10 % vergrößert. Die mittlere Spalte zeigt die

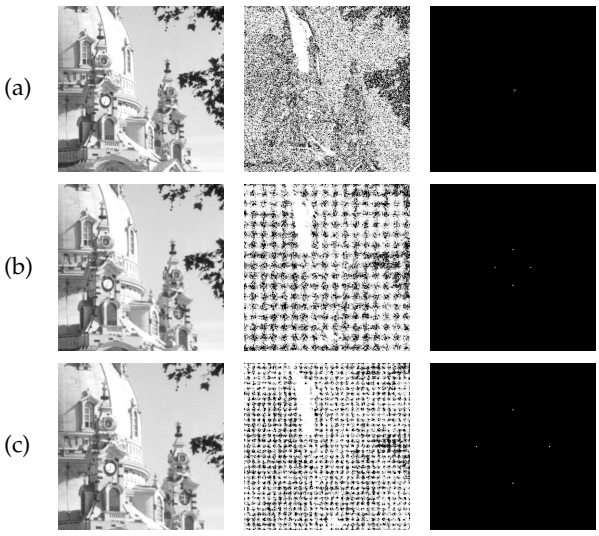

Abbildung 2: Detektion von Re-Sampling nach Vergrößerung des Originalbildes um (b) 5 % und (c) 10 %. Die mittlere Spalte zeigt jeweils die Wahrscheinlichkeiten dafür, dass ein Pixel mit seinen Nachbarn korreliert. Im Falle von Re-Sampling haben diese eine periodische Struktur und dementsprechend charakteristische Peaks im Betragsspektrum (rechte Spalte). Im Originalbild (a) sind keine Artefakte nachweisbar.

ermittelten Wahrscheinlichkeiten dafür, dass ein Pixel mit seinen Nachbarn korreliert. Diese haben eine klare periodische Struktur, welche sich in deren Betragsspektrum in Form von charakteristischen hochfrequenten Peaks zeigt, rechte Spalte. Für das Originalbild (erste Zeile) lassen sich keinerlei periodischen Artefakte feststellen.

Wie die in [6] vorgestellten Ergebnisse zeigen, können mit diesem Verfahren sehr zuverlässig verschiedenste Formen von geometrischen Transformationen detektiert werden. Erwartungsgemäß nimmt die Zuverlässigkeit bei einer starken Verkleinerung des Bildes ab. Problematisch gestaltet sich die Detektion in JPEG-komprimierten Bildern, da durch die Kompression bedingte periodische (Block-)Artefakte im Bild die Spuren der Umtastung zerstören können. Auch hier gilt jedoch, dass eine aussagekräftige Detektion von Bildmanipulation immer auf mehreren Analysen fußen sollte.

5 Zusammenfassung

Mit den beiden vorgestellten Verfahren sollte ein Einblick in das verhältnismäßig neue Gebiet der digitalen Bildforensik gegeben werden. Insbesondere für die Detektion von Manipulationen existiert (bisher) kein allgemein anwendbares Verfahren. Vielmehr wurden zahlreiche Methoden entwickelt, die auf spezielle Schritte beim Erstellen einer Fälschung ausgerichtet sind.[1] Diese sind häufig noch nicht ausgereift genug, um es mit raffinierten Manipulationen aufzunehmen. Durch einen kombinierten Einsatz mehrerer Verfahren ist das unbemerkte Erstellen von überzeugenden Manipulationen jedoch bereits als ungleich schwerer zu beurteilen.

Literatur

[1] BLYTHE, P. und J. FRIDRICH: *Secure Digital Camera*. In: *Digital Forensic Research Workshop, Baltimore*, 2004.

[2] CRAVER, S.A., M. WU, B. LIU, A. STUBBLEFIELD, B. SWARTZLANDER, D.S. DEAN und E. FELTEN: *Reading between the Lines: Lessons from the SDMI Challenge*. In: *10th USENIX Security Symposium*, 2001.

[3] LUKÁŠ, J., J. FRIDRICH und M. GOLJAN: *Digital Camera Identification from Sensor Noise*. IEEE Transactions on Information Security and Forensics, 1(2):205–214, 2006.

[4] NG, T.-T., S.-F. CHANG, C.-Y. LIN und Q. SUN: *Passive-blind Image Forensics*. In: ZENG, W., H.YU und C.-Y. LIN (Herausgeber): *Multimedia Security Technologies for Digital Rights*. Academic Press, 2006.

[5] PEARSON, H.: *Image Manipulation: CSI: Cell Biology*. Nature, 434(7036):952–953, 2005.

[6] POPESCU, A.C. und H. FARID: *Exposing Digital Forgeries by Detecting Traces of Re-sampling*. IEEE Transactions on Signal Processing, 53(2):758–767, 2005.

[7] SCHMUNDT, H.: *Obskure Kameras*. Der Spiegel, 39:234 – 237, 2006.

[8] WESTFELD, A.: *Lessons from the BOWS Contest*. In: *Proceedings of ACM MM-SEC*, 2006.

[1] Ein wirklich umfassender Literaturüberblick, auf den an dieser Stelle verwiesen werden könnte, ist leider noch nicht erstellt worden. Traditionell entstammen aber viele Ansätze den Arbeitsgruppen von Hany Farid (`http://www.cs.dartmouth.edu/farid/`) und Jessica Fridrich (`http://www.ws.binghamton.edu/fridrich/`).