Der Einfluss gleichgewichteter Fusion in der Mikrofonforensik unter beispielhafter Nutzung von zwei Klassifikatoren

Christian Krätzer, Jana Dittmann

Fakultät für Informatik, AG Multimedia and Security
Otto-von-Guericke-Universität Magdeburg
Universitätsplatz 2, 39106 Magdeburg
{kraetzer,dittmann}@iti.cs.uni-magdeburg.de

Abstract: Für den beispielhaft gewählten Anwendungsbereich der Mikrofonforensik wird in diesem Beitrag gezeigt, dass bisherig verwendete statistische Mustererkennungsverfahren zur Mikrofonerkennung von der Informationsfusion profitieren können. Dies wird im Beitrag mit Ergebnissen für Fusionen auf Match-, Rank- und Decision-Level sowie der Nutzung zweier Mehrklassenklassifikatoren (einem Entscheidungsbaumverfahren und logistischen Regressionsmodellen) belegt. Durch die Fusion konnten die erzielten Klassifikationsgenauigkeiten in der überwachten Klassifikation auf den zwei hier genutzten Testmengen (eine mit vier und die anderer mit sieben Mikrofonen) zum Teil bis auf 100% erhöht werden. Damit wird der Wert der Mikrofonforensik als Methode der Quellverifikation z.B. für die Langzeitarchivierung, weitergesteigert.

1 Einleitung

Im Vergleich zur Kameraforensik sind die Audio- und die Mikrofonforensik bisher ein relativ schwach beforschtes Gebiet. Felder auf denen Medienforensikansätze bereits wesentlich weiter gediegen sind als in der Mikrofonforensik sind u.a. das bereits erwähnte Feld der Kameraforensik ([FFG08], [GFF09]), die Scanner- und Druckerforensik [KMC08] oder die Identifikation von Grafiktablets [OVD07]. Die meisten dieser Medienforensikansätze bedienen sich dabei der statistischen Mustererkennung als Mechanismus für die Quellenidentifikation/-verifikation. Die primären Ziel einer Mikrofonforensik sollten es also sein ein Mikrofon als Quelle einer Aufnahme eindeutig zu identifizieren bzw. zu verifizieren das eine Audiodatei nur aus einer Quelle stammt (d.h. nicht zusammenmontiert wurde). Neben diesen primären zielen, die z.B. in der Aufbereitung von Audiomaterial für die Verwendung vor Gericht oder in der Ingest-Phase sicherer digitaler Langzeitarchivierung verfolgt werden, könnte ein zuverlässiger Mikrofonforensik auch auf anderen Anwendungsfeldern Verwendung finden.

So basieren bisherige Audioforensikmethoden wie z.B. die Schussgeräuscherkennung [Ma07] zumeist auf starren Testaufbauten. Dadurch sind solche Ansätze fest an einen Sensor (Mikrofon) gebunden und sind kaum von einem System auf ein anderes übertragbar ohne den Sensor mit auszutauschen. Mikrofonforensik, wie sie in diesem Beitrag betrachtet wird, würde zum einen dazu beitragen ein Ähnlichkeitsmaß zwischen Mikrofonen zu definieren und darüber den Austausch von Sensoren ohne kompletten Vertrauensverlust ermöglichen. Zum anderen wäre es mit einem zuverlässigen Mikrofonforensikansatz möglich den Einfluss eines Mikrofons auf ein Signal zu interpolieren und damit zu minimieren.

Dieser Beitrag greift einen existierenden Mikrofonforensikansatz aus [KOD07] auf und erweitert ihn durch den Einsatz von Informationsfusion. Dazu wird ein state-of-the-art fünf Ebenen Fusionsmodell [RNJ06] aus der Biometrie herangezogen, welches eine Informationsfusion auf Signal-, Feature-, Match-, Rank- und Decision-Level erlaubt. Das grundlegende Modell dazu ist in Abbildung 1 zu sehen.

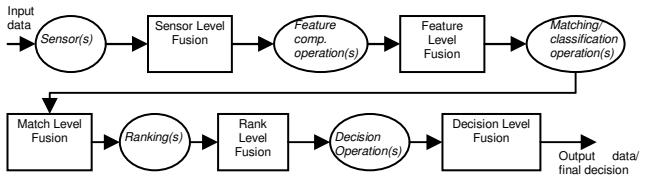

Abbildung 1: Der verwendete Ansatz zur Informationsfusion (nach [RNJ06])

Die Fusionsbetrachtungen in diesem Beitrag beschränken sich auf drei der fünf möglichen Ebenen, nämlich auf Match-, Rank- und Decision-Level Fusion. Sensor-Level Fusion wird ausgeschlossen da hier gerade der Einfluss den Sensors (Mikrofon) auf das Signal evaluiert werden soll, Feature-Level Fusion wird ausgeschlossen, da diese für den verwendeten Merkmalsextraktor und ein ähnliches Anwendungsfeld bereits in vorangegangenen Publikationen (z.B. [KD08b]) durchgeführt wurde.
Den Autoren sind bisher zwei Publikationen zur Mikrofonforensik bekannt. In [KOD07] wird eines der beiden hier verwendeten Mikrofon-Testsets schon einmal ohne Fusion evaluiert, was dort zu einer maximalen Klassifikationsgenauigkeit von 75,99% führt. Das zweite hier getestete Testset wird bereits in [BKD09] mit einem Frequenzraum-Merkmalsextraktor und einer maximalen Klassifikationsgenauigkeit von 93.5% getestet. Für den Vergleich zu den existierenden Resultaten im Bereich der Mikrofonforensik sei an dieser Stelle vorweggenommen das durch die hier vorgestellten Fusionsoperationen in der Mikrofonforensik auf den gleichen Testmengen die Klassifikationsgenauigkeiten z.T. bis auf 100% erhöht werden konnte. Ließen sich diese Testergebnisse generalisieren – was mit deutlich größeren Testmengen verifiziert werden müsste – so wäre mit der Mikrofonforensik ein wertvoller Mechanismus zur Quellenverifikation gegeben.

Dieser Beitrag ist wie folgt aufgebaut: Abschnitt 2 beschreibt das grundlegende Konzept. inklusive der Formalisierung der verwendeten Fusionskonzepte. Kapitel 3 fasst den Testaufbau zusammen, während Kapitel 4 die Testergebnisse darstellt. Kapitel 5 fasst die gewonnenen Erkenntnisse noch einmal zusammen und gibt einen Ausblick auf weiterführende Arbeiten.

2 Grundlegendes Konzept

In diesem Abschnitt werden die zugrunde liegenden Signalverarbeitungsoperationen wie Signalerzeugung, Merkmalsextraktion und Klassifikation beschrieben. Anschließend folgt eine Formalisierung des Fusionskonzeptes mit den verwendeten Fusionsfunktionen.

2.1 Grundlegende Signalverarbeitungsschritte

In der Signalerzeugung werden Audioaufnahmen analog zum Vorgehen in [KOD07] erzeugt. Dies geschieht unter der Verwendung von zehn Referenzsignalen bei der Aufnahme an zehn unterschiedlichen Aufnahmepositionen (R01, R02, …, R10), um den Einfluss von Umgebungslärm zu minimieren. Für eine detailliertere Beschreibung der Prozedur siehe [KOD07]. Um die Generalisierbarkeit der Aussagen zu erhöhen, wurden die Aufnahmen zeitversetzt mit zwei unterschiedlichen Mengen von Mikrofonen durchgeführt. Die resultierenden Aufnahmen werden im Folgenden als Aufnahmemenge 1 und 2 (*RS1* und *RS2*) bezeichnet. Aus den Mengen wird jeweils eine feste Teilmenge[1] für die Tests reserviert (20%) und der Rest für das Training verwendet.

Aufnahmemenge 1 (*RS1*): Die Mikrofon und Mikrofonvorverstärker Kombinationen die für die Erzeugung dieser Aufnahmemenge genutzt wurden sind in Tabelle 1 identifiziert. Die Menge der dabei verwendeten Mikrofone wird im Folgenden als M_{RS1} bezeichnet.

Mikrofon (M_i)	**Hersteller & Modell**	**Vorverstärker**
M_1	AKG SE 300 B (CK93)	Millenium Mic 1
M_2	TerraTec HeadsetMaster	Sound Blaster USB
M_3	Shure SM58	Sound Blaster USB
M_4	T.bone MB45	Millenium Mic 1

Tabelle 1: M_{RS1} – die Menge der zur Erzeugung von Aufnahmemenge 1 genutzten Mikrofone

Aufnahmemenge 2 (*RS2*): Die Mikrofon und Mikrofonvorverstärker Kombinationen die für die Erzeugung von *RS2* genutzt wurden sind in Tabelle 2 identifiziert. Während gleiche Kombinationen aus Mikrofon und Vorverstärker (z.B. M_2 in *RS1* und *RS2*) ihre Kennung behalten, erhalten neue Kombinationen dabei eine neue Kennung[2]. Die Menge der in *RS2* verwendeten Mikrofone wird im Folgenden als M_{RS2} bezeichnet. Erwähnenswert ist der Fakt das M_7 und M_8 relativ gute Ergebnisse zeigen, obwohl sie sich bei gleichem Mikrofonkorpus nur durch den Richtaufsatz unterscheiden.

---

[1] Zwei Dateien pro Aufnahmeposition und Mikrofon, also insgesamt 20% des aufgezeichneten Materials.

[2] Bei allen hier gemachten Betrachtungen bleibt der genaue Einfluss des Mikrofonvorverstärkers unbetrachtet. Bei der obigen Beschreibung der Testmengen wird die Kombination aus Mikrofon und Vorverstärker als eine Einheit betrachtet.

Mikrofone (M_i)	Hersteller & Modell	Vorverstärker
M_2	TerraTec HeadsetMaster	Sound Blaster USB
M_5	PUX 70TX-M1	Sound Blaster USB
M_3	Shure SM58	Sound Blaster USB
M_6	T.bone MB45	Sound Blaster USB
M_7	AKG SE 300 B (CK93)	Sound Blaster USB
M_8	AKG SE 300 B (CK98)	Sound Blaster USB
M_9	AKG SC600	Sound Blaster USB

Tabelle 2: M_{RS2} – die Menge der zur Erzeugung von Aufnahmemenge 2 genutzten Mikrofone

Für die Merkmalsextraktion wird hier das AMSL Audio Steganalysis Toolset (AAST [KD08a] Version 1.04 build 20071005) verwendet. Dieser Merkmalsextraktor berechnet für Fenster eines Audiosignals sieben Zeitbereichmerkmale, 56 Cepstral-Domain Merkmale und 35 Frequenzraummerkmale. Das Resultat ist ein 98 dimensionaler Merkmalsvektor pro Fenster (frame) einer Datei.

Im Rahmen dieses Beitrages wird zur Klassifikation eine Fusion auf Basis der zwei Multi-Klassen-Klassifikatoren SimpleLogistics (*SL*; generiert lineare logistisch Regressionsmodelle; Landwehr et al. [LHF03]) und J48 (ein C4.5 Entscheidungsbaum [Qu93]) verwendet. Diese werden auf Basis von Vortests auf kleineren Testmengen ausgewählt, wo sie aus der Menge aller in der Data Mining Suite Weka [WF05] enthaltenen Klassifikatoren die besten Ergebnisse (bezüglich erzielter Klassifikationsgenauigkeit und benötigter Rechenzeit) lieferten. Bei beiden Klassifikatoren wird die Weka Implementation mit ihren Standardparametern verwendet.

2.2 Fusionsschritte und Fusionsoperatoren

Die folgenden Abschnitte enthalten eine Beschreibung der Fusionsschritte sowie eine Formalisierung der Fusionsoperatoren/-funktionen. Die Betrachtungen enthalten darüber hinaus Aussagen zur Messung der Klassifikationsgenauigkeiten.

2.2.1 Fusionsschritte

Der zugrundeliegende Klassifikationsprozess für die Mikrofonforensik ohne Fusion ist in Abbildung 2 dargestellt. Für die Menge aller für das Training aufgenommenen Dateien (F^{Train}) wird pro Datei eine feste Anzahl (l_{max}) von fensterweise bestimmten Merkmalsvektoren berechnet. Alle diese Merkmalsvektoren (mit den entsprechenden Informationen zur Klassenzugehörigkeit) werden dann genutzt um das Modell für den jeweiligen Klassifikator zu trainieren. In der Testphase wird auf den für das Testen reservierten Dateien (F^{Test}; $F^{Test} \cap F^{Train} = \{\}$) dieselbe Merkmalsextraktion durchgeführt wie für das Training, unter Nutzung desselben Wertes für l_{max}. Anschließend wird mit dem vorher bestimmten Modell die Klassifikation der Testvektoren durchgeführt.

Da in dieser überwachten Klassifikation die Informationen über die wahre Klassenzugehörigkeit eines Merkmalsvektors verfügbar ist, kann hier auch die Klassifikationsgenauigkeit bestimmt werden (siehe Abschnitt 2.2.2).

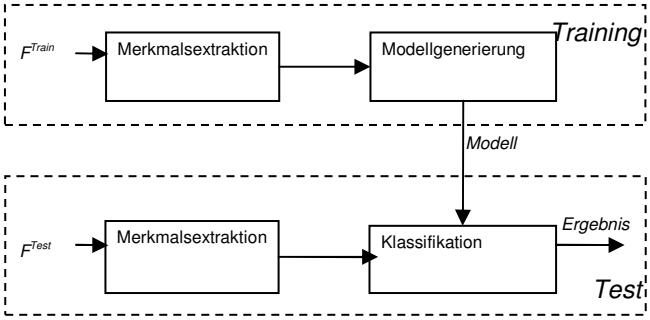

Abbildung 2: Training und Test für einen Klassifikator ohne Fusion

Für die Betrachtungen des Einflusses von Match-, Rank- and Decision-Level Fusion werden die in Abschnitt 2.2.2 beschriebenen Fusionsfunktionen und –operatoren auf die Ausgabe des Klassifikators angewandt, wie in Abbildung 3 dargestellt. Die Ausgabe ist dabei (abhängig von verwendeten Fusionslevel) entweder die Entscheidungen der beiden Klassifikatoren für einen Merkmalsvektor des Testmaterials (Match-Level), je eine Rangmatrix (Rank-Level) oder die zwei individuell getroffenen Entscheidungen bezüglich der Klassenzugehörigkeit einer Testdatei (Decision-Level).

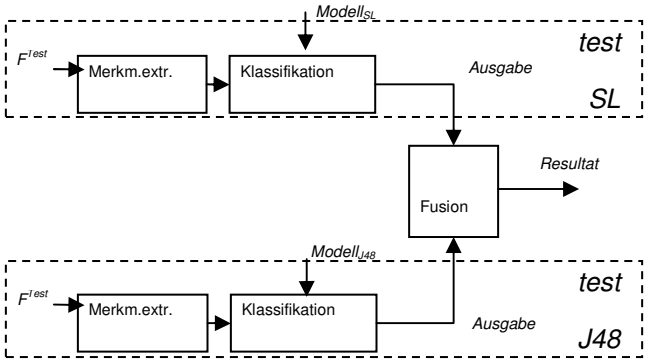

Abbildung 3: Test mit Fusion zweier Klassifikatoren

2.2.2 Fusionsfunktionen / -operatoren

Für die Beschreibung der Fusion müssen an dieser Stelle zunächst einige **Grundlegende Definitionen** getroffen werden. So seien M_{RS1} und M_{RS2} die zwei Mengen der für die Aufnahmen in *RS1* und *RS2* (siehe Abschnitt 2.1; $M_{RS1} = \{M_1, M_2, M_3, M_4\}$ und $M_{RS2} = \{M_2, M_5, M_3, M_6, M_7, M_8, M_9\}$) verwendeten Mikrofone.

Die Menge M aller getesteten Mikrofone ist dann definiert als $M = M_{RS1} \cup M_{RS2}$ und M_i ist ein Mikrofon in dieser Menge ($M_i \in M$). Die Anzahl der Mikrofone in M ist $n=|M|$, des Weiteren gilt $n_{RS1} = |M_{RS1}| = 4$ und $n_{RS2} = |M_{RS2}| = 7$.

Bezüglich des **Testmaterial**s werden folgende Definitionen getroffen: f_j ist eine Datei in der Menge der für die Testphase aufgenommenen Dateien (F^{Test}): $f_j \in F^{Test}$ ($1 \le j \le j_{max}$ mit $j_{max} = |F^{Test}|$; $j_{max}=2 \cdot n_{RS1}$ oder $j_{max}=2 \cdot n_{RS2}$ für eine Aufnahmeposition[3] und *RS1* oder *RS2*). Für jede Datei wird eine festgelegte Anzahl von Merkmalsvektoren $f_{j,l}$ ($1 \le l \le l_{max}$; hier wird konsequent eine Fenstergröße von 1024 Samplewerten pro Merkmalsvektor genutzt, d.h. $l_{max}=200$) für die Auswertungen betrachtet. Die Identifikationsfunktion $IDF(f_j)$ bestimmt das korrekte M_i Mikrofon für eine Datei f_j als: $M_{i,f_j} = IDF(f_j) \in M$ während $ID(f_{j,l})$ das korrekte Mikrofon für einen Merkmalsvektor $f_{j,l}$ bestimmt als $M_{i,f_{j,l}} = ID(f_{j,l}) \in M$.

Die **Entscheidungen der genutzten Klassifikatoren bezüglich der Klassenzugehörigkeit eines Merkmalsvektors** $f_{j,l}$ geschieht durch die Klassifikationsfunktionen *SL*() und *J48*() mit den zugehörigen Modellen $model_{SL}$ und $model_{J48}$ (generiert durch das trainieren des Klassifikators auf dem Trainingsmaterial F^{Train}) als: $M_{i,f_{j,l},SL}=SL(f_{j,l},model_{SL})$ und $M_{i,f_{j,l},J48}=J48(f_{j,l},model_{J48})$. Die **Entscheidungen der genutzten Klassifikatoren bezüglich der Zugehörigkeit einer ganzen Datei** wird durch die modifizierten Klassifikationsfunktionen *SLM*() und *J48M*() getroffen, welche jeweils eine Mehrheitsentscheidung über alle l_{max} Merkmalsvektoren von f_j bilden: $M_{i,f_j,SL}=SLM(f_j,model_{SL})$ und $M_{i,f_j,J48}=J48M(f_j,model_{J48})$. Die Ausgabe aller Klassifikationsfunktionen ist jeweils ein Element aus M.

Die Funktion *verify*() transferiert die Entscheidung eines Klassifikators für einen Merkmalsvektor $f_{j,l}$ auf natürliche Zahlen im Bereich [0;1]. Dies geschieht durch den Vergleich der Entscheidung mit der wahren Mikrofonklasse (am Beispiel *SL*):

$$verify(f_{j,l}) = \begin{cases} 1 & if \quad M_{i,f_{j,l},SL}=M_{i,f_{j,l}} \\ 0 & else \end{cases} ; 1 \le j \le j_{max}; 1 \le l \le l_{max} \qquad (1)$$

Die Klassifikationsgenauigkeit für eine Datei kann ohne Fusion bestimmt werden als:

$$accuracy_{f_j} = \frac{1}{l_{max}} \sum_{l=1}^{l_{max}} verify(f_{j,l}) \qquad (2)$$

Die **Match-Level Fusion (ML)** wird durch den Vergleich der Entscheidungen beider Klassifikatoren für jeden Merkmalsvektor $f_{j,l}$ in einer Datei f_j realisiert.

---

[3] Der Faktor 2 in der Berechnung ergibt sich aus dem Umstand dass pro Mikrofon und Aufnahmeposition zwei der aufgenommenen zehn Dateien für das Testen reserviert werden (siehe Abschnitte 2.1 und 3).

Wenn beide Entscheidungen übereinstimmen, so liefert die Fusion diese als Konsens zurück, ansonsten liefert sie "*uncertain*":

$$fusion_{ML}(f_{j,l}) = \begin{cases} M_{i,f_{j,l},SL} & if \quad M_{i,f_{j,l},SL} = M_{i,f_{j,l},J48} \\ "uncertain" & else \end{cases} \quad \text{mit } 1 \le j \le j_{max};\ 1 \le l \le l_{max} \quad (3)$$

Für die Bestimmung der Genauigkeit der Klassifikation wird das Ergebnis wieder auf natürliche Zahlen im Bereich [0;1] transferiert:

$$verify_{ML}(f_{j,l}) = \begin{cases} 1 & if \quad fusion_{ML}(f_{j,l}) = M_{i,f_{j,l}} \\ 0 & else \end{cases} \quad \text{mit } 1 \le j \le j_{max};\ 1 \le l \le l_{max} \quad (4)$$

Dann kann die Klassifikationsgenauigkeit pro Datei f_j berechnet werden als:

$$accuracy_{ML,f_j} = \frac{1}{l_{max}} \sum_{l=1}^{l_{max}} verify_{ML}(f_{j,l}) \quad (5)$$

Die **Rank-Level Fusion (RL)** geschieht in drei Schritten: Zuerst werden die Konfusionsmatrizen C_{SL,f_j} und C_{J48,f_j} (Größe: $n_{RS1} \times n_{RS1}$ für *RS1* und $n_{RS2} \times n_{RS2}$ für *RS2*) der beiden Klassifikatoren für jede Datei f_j erstellt[4]. Dies geschieht elementweise für die Elemente der Matrizen $c_{x,y}^{SL,f_j}$ und $c_{x,y}^{J48,f_j}$ als:

$$c_{x,M_i}^{SL,f_j} = count(M_{i,f_{j,l},SL} = M_x) \quad \text{mit } 1 \le l \le l_{max} \quad (6)$$

$$c_{x,M_i}^{J48,f_j} = count(M_{i,f_{j,l},J48} = M_x) \quad \text{mit } 1 \le l \le l_{max} \quad (7)$$

Die Funktion *count*() zählt dabei für alle Merkmalsvektoren $f_{j,l}$ in einer Datei f_j, wie oft ein Mikrofon als ein jeweils anderes Mikrofon klassifiziert wird. Somit enthalten die Hauptdiagonalenelemente der Konfusionsmatrizen die Fälle wo M_i als M_i richtig klassifiziert wurde.

Im zweiten Schritt wird pro Datei für jede Konfusionsmatrix eine Rangmatrix erzeugt. Dies geschieht unter Nutzung der Funktion *ranking*(), die zeilenweise Ränge zuweist. Diese Ränge sind hier im Bereich [1;n_{RS1}] oder [1;n_{RS2}], wobei "1" der beste erreichbare Rang ist.

---

[4] Bei blinden Tests in der Praxis stehen die Konfusionsmatrizen natürlich nicht zur Verfügung. Daher würde dort, unter der Grundannahme dass die Komplette Datei mit (nur) einem Mikrofon aufgenommen wurde, ein Ranking über der Testmenge gegen alle Mikrofone in der Trainingsmenge durchgeführt. Dies entspräche im Prinzip einem Zeilenvektor eines überwachten Tests, der dann mit den gleichgearteten Vektoren der anderen Klassifikatoren fusioniert würde. Als Entscheidung wird auch hier der beste Rang zurückgeben.

Im Optimalfall gilt $c_{M_i,M_i}^{SL,f_j} = \max(c_{M_x,M_i}^{SL,f_j})$ wodurch dem Hauptdiagonalenelement in der jeweilig betrachteten Zeile der beste Rang ("1") zugewiesen wird. Die daraus resultierenden Rangmatrizen werden im Folgenden als RM_{J48,f_j} and RM_{SL,f_j} identifiziert. Im dritten Schritt werden die zwei entstandenen Rangmatrizen pro Datei f_j fusioniert, wie in Gleichung (8) gezeigt. Dabei werden beide Rangmatrizen addiert und alle Elemente der entstehenden fusionierten Rangmatrix FRM_{f_j} mit der Anzahl der beteiligten Klassifikatoren (hier zwei) normiert, wodurch die Werte wieder in die Bereiche [1,n_{RS1}] bzw. [1,n_{RS2}] zurückskaliert werden.

$$FRM_{f_j} = \frac{1}{2}\left(RM_{SL,f_j} + RM_{J48,f_j}\right) \tag{8}$$

Um die Bestimmung der Klassifikationsgenauigkeit zu vereinfachen werden die Resultate zeilenweise (also pro Mikrofon M_i) auf natürliche Zahlen im Bereich [0;1] transferiert. Dies geschieht unter Betrachtung der einzelnen Elemente $frm_{x,y}^{f_j}$ in der Zeile der Fusionierten Rangmatrix (hier am Beispiel *RS1*):

$$verify_{RL}(M_i) = \begin{cases} 1 & \text{if} \quad frm_{M_i,M_i}^{f_j} = min(fmr_{M_1,M_i}^{f_j}, \ldots, fmr_{M_4,M_i}^{f_j}) \\ 0 & \text{else} \end{cases} \tag{9}$$

Die Klassifikationsgenauigkeit kann dann pro Datei f_j bestimmt werden als (Bsp.: *RS1*):

$$accuracy_{RL,f_j} = \frac{1}{n_{RS1}} \sum_{M_i} verify_{RL}(M_i) \tag{10}$$

Die hier verwendete **Decision-Level Fusion (DL)** ist der oben vorgestellten Match-Level Fusion sehr ähnlich. Die beiden zu fusionierenden Klassifikatorentscheidungen werden hier mittels *SLM*() und *J48M*() pro Datei generiert. Wenn beide Klassifikatoren dasselbe Ergebnis liefern, dann wird dieser Konsens als gemeinsame Entscheidung zurückgeliefert. Wird kein Konsens erreicht, so wird "*uncertain"* zurückgegeben:

$$fusion_{DL}(f_j) = \begin{cases} M_{i,f_j,SL} & \text{if} \quad M_{i,f_j,SL} = M_{i,f_j,J48} \\ "uncertain" & \text{else} \end{cases} \quad \text{mit } 1 \leq j \leq j_{max} \tag{11}$$

Im Gegensatz zur Match-Level Fusion wird hier unter Nutzung von $IDF(f_j)$ eine Entscheidung pro Datei f_j getroffen wird anstelle einer Entscheidung mittels $ID(f_{j,l})$ pro Merkmalsvektor $f_{j,l}$. Die entsprechende Vereinfachung der Ergebnisdarstellung pro Datei f_j um die Berechnung der Klassifikationsgenauigkeit zu erleichtern sieht wie folgt aus:

$$verify_{DL}(f_j) = \begin{cases} 1 & \text{if} \quad fusion_{DL}(f_j) = M_i \\ 0 & \text{else} \end{cases} \quad \text{mit } 1 \leq j \leq j_{max} \tag{12}$$

Im Anschluss kann dann somit Klassifikationsgenauigkeit über alle j_{max} Dateien eines Testsets bestimmt werden als:

$$accuracy_{DL} = \frac{1}{j_{max}} \sum_{f_j} verify_{DL}(f_j) \tag{13}$$

Zusammenfassung der Rückgabewerte der verschiedenen Fusionen: Sei M die Menge aller Mikrophone in der Evaluierung und $f_{j,l}$ ein (als Merkmalsvektor beschriebener) frame einer Audiodatei f_j, dann liefert die hier beschriebene **Match-Level Fusion** für jeden $f_{j,l}$ ein Resultat aus $\{M\} \cup \{uncertain\}$ – siehe Gleichung (3). Die R**ang-Level Fusion** liefert für eine Datei oder eine Menge von Dateien eine fusionierte Rangmatrix (Gleichung (8)) und die D**ecision-Level Fusion** liefert eine Entscheidung aus $\{M\} \cup \{uncertain\}$ (siehe Gleichung (11)) für jede Datei in der Testmenge.

3 Testparametrisierung

Die Evaluierungen in diesem Beitrag beschäftigt sich mit zwei Fragen: Was ist der erzielbare Nutzen von Fusion in der Mikrofonforensik? und: Welche der vorgestellten Fusionsstrategien zeigt den größten Nutzen? Dabei enthält dieser Abschnitt des Beitrages die notwendigen Informationen zur Parametrisierung der hier durchgeführten Signalverarbeitungsschritte.

Wie bereits in Abschnitt 2.1 beschrieben erfolgt die Signalerzeugung per Audioaufnahmen mit zwei Mengen an Mikrofonen an zehn unterschiedlichen Aufnahmepositionen (siehe zu Details zur Aufnahmeprozedur [KOD07]). Die aufgenommenen Signale werden als PCM kodierte WAV Dateien (mono, 44,1kHz Abtastrate und 16Bit Quantisierung) abgespeichert. Es entstehen pro Mikrofon und Aufnahmeposition zehn Dateien (eine pro aufgenommenem Referenzsignal) von denen acht für das Training und die restlichen zwei für die Tests reserviert werden ($j_{max}=2 \cdot n_{RS1}$ bzw. $j_{max}=2 \cdot n_{RS2}$; mit $n_{RS1}=4$ und $n_{RS2}=7$). Es soll an dieser Stelle noch einmal festgestellt werden dass die Referenzdateien nur dazu verwendet werden den Umgebungseinfluss der Aufnahmepositionen zu minimieren und sie nicht in die Klassifikation einfließen.

Außerdem wird in diesem Beitrag der erwähnte Einfluss der Aufnahmeumgebungen nicht weiter betrachtet – dies bleibt weiterführender Forschung vorbehalten – hier werden dir Identifier der Aufnahmeumgebungen nur weitergeführt um eine möglichst übersichtliche Präsentation der Testergebnisse zu ermöglichen.

In den hier durchgeführten Tests werden pro Datei $l_{max}=200$ aufeinanderfolgende, nicht überlappende Fenster a 1024 Audiosamples für die fensterbasierte Merkmalsextraktion genutzt. Der durch die Nutzung von AAST (siehe Abschnitt 2.1) entstehende Merkmalsvektor ist dabei 98-dimensional und enthält zusätzlich noch ein Klassenidentifier für die Verifikation der erzielten Genauigkeiten in diesem Ansatz der überwachten Klassifikation.

Insgesamt werden pro Aufnahmeposition und Mikrofon ca. 50 Sekunden Audiomaterial für Training und Test verwendet. Bei dem hier verwendeten Verhältnis von 80% Training zu 20% Test entspricht dies 6400 pro Aufnahmeposition für das Trainieren und 1600 für das Testen. Als Gesamtsumme über beide Aufnahmemengen werden somit 176000 Merkmalsvektoren für das Training und 44000 für das Testen betrachtet.

4 Testergebnisse

Tabelle 3 und Tabelle 4 fassen an dieser Stelle vorwegnehmend die erzielten Testergebnisse mit und ohne Fusion zusammen. Um ein gewisses Maß an Granularität zu gewährleisten werden an dieser Stelle die Ergebnisse pro Aufnahmeposition angegeben. In den folgenden Sektionen werden diese Ergebnisse ausgewertet und der Einfluss der betrachteten Fusionsstrategien auf die erzielten Klassifikationsgenauigkeiten herausgearbeitet.

	R01	**R02**	**R03**	**R04**	**R05**	**R06**	**R07**	**R08**	**R09**	**R10**	**Durchschn.**
ohne Fusion SL	78,8%	85,8%	73,6%	88,5%	82,9%	85,7%	81,7%	85,3%	85,4%	90,9%	83,8%
ohne Fusion J48	83,1%	93,1%	73,3%	90,9%	86,9%	84,9%	90,0%	91,9%	82,6%	95,0%	87,2%
ML	69,4%	82,4%	59,3%	83,5%	76,5%	77,6%	74,8%	80,6%	75,8%	88,1%	76,8%
RL	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
DL	100,0%	100,0%	75,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	97,5%

Tabelle 3: Zusammenfassung der Testergebnisse für *RS1*

	R01	**R02**	**R03**	**R04**	**R05**	**R06**	**R07**	**R08**	**R09**	**R10**	**Durchschn.**
ohne Fusion SL	74,6%	70,8%	80,4%	77,4%	74,1%	81,1%	74,9%	77,9%	77,8%	77,3%	76,6%
ohne Fusion J48	76,6%	74,6%	87,1%	79,6%	74,8%	83,5%	75,4%	80,1%	84,6%	80,3%	79,7%
ML	64,9%	60,9%	74,1%	67,2%	62,0%	72,2%	63,5%	68,5%	68,8%	67,4%	66,9%
RL	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
DL	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	85,7%	100,0%	100,0%	85,7%	97,1%

Tabelle 4: Zusammenfassung der Testergebnisse für *RS2*

4.1 Klassifikationsgenauigkeiten ohne Fusion

Wie in Tabelle 3 und Tabelle 4 gezeigt liegen die durchschnittlich erzielten Klassifikationsgenauigkeiten für *RS1* in den Bereichen 73.6%-90.9% für SimpleLogistics und 73.3%-95.0% für *J48*. Bei *RS2* sind die erzielten Ergebnisse etwas niedriger (70.8%-81.1% und 74.6%-87.1%). Der Unterschied mag sich darin begründen das *RS2* mehr Mikrofone enthält und diese Mikrofone sich z.T. auch noch ähnlicher sind als in *RS1* (z.B. haben M_7 and M_8 denselben Korpus). Der Klassifikator *J48* erzielt auf *RS1* ein 3,4% besseres Resultat als SimpleLogistics und ist auch im Falle von *RS2* 2,9% besser.

	SL		***J48***	
	Durchschnitt	**Max.**	**Durchschnitt**	**Max.**
M_1	80,1%	88,3% (R08)	80,6%	85,5% (R10)
M_2	97,8%	100% (R02)	98,1%	100% (R02)
M_3	84,3%	93,8% (R10)	89,8%	97,5% (R04)
M_4	73,2%	90,3% (R04)	80,3%	99,5% (R10)

Tabelle 5: Durchschnittliche und maximale Klassifikationsgenauigkeiten über alle 10 Aufnahmepositionen in *RS1* (bestes Resultat Grau markiert)

Werden die Ergebnisse mikrofonweise betrachtet, so zeigt Tabelle 5 für *RS1* das beste Ergebnis für M_2. Dies resultiert vermutlich aus dem Umstand dass dies das niederwertigste Mikrofon in dieser Testmenge ist, und somit über einen charakteristischeren Frequenzgang verfügt.

Führt man dieselben Betrachtungen für *RS2*, so zeigt Tabelle 6 das auch hier die Mikrofone mit der geringsten Qualität (M_2 und M_5 ein Headset und ein minderwertiges Sprachmikrofon) besser erkannt werden. Allerdings werden auch sehr gute Ergebnisse für das höchstwertige Mikrofon in dieser Menge erzielt (M_9, ein hochqualitatives Studiomikrofon), was sich vermutlich auch wieder damit begründen lässt dass der Frequenzgang dieses Mikrofons sehr charakteristisch ist. Interessant ist der Fakt das M_7 und M_8 relativ gute Ergebnisse zeigen, obwohl es sich dabei um denselben Mikrofonkorpus handelt, der nur mit anderen Richtaufsätzen versehen wurde. Dies lässt vermuten dass jedes Mikrofon über ein charakteristisches Eigenrauschen (bestimmt durch die serientypische Umwandlungstechnologie und einen mikrofonspezifischen, alterungsabhängigen Zustand der Membran) verfügt. Detailliertere Untersuchungen mit gleichartigen Mikrofonen wären hier notwendig um diese Frage hinreichend zu beantworten.

	SL		*J48*	
	Durchschnitt	**Max.**	**Durchschnitt**	**Max.**
M_2	92,9%	97,8% (R10)	94,4%	98,8% (R09)
M_5	97,7%	100% (R07)	76,1%	99,3% (R07)
M_3	70,5%	90% (R03)	69,6%	97,8% (R03)
M_6	61,3%	71,5% (R06)	66,9%	95,5% (R03)
M_7	56,7%	69,5% (R09)	65,0%	74,5% (R04)
M_8	67,9%	77,3% (R07)	91,5%	73,3% (R08)
M_9	89,4%	100% (R10)	94,4%	98,8% (R10)

Tabelle 6: Durchschnittliche und maximale Klassifikationsgenauigkeiten über alle 10 Aufnahmepositionen in *RS2* (bestes Resultat Grau markiert)

4.2 Einfluss der Fusion auf die Klassifikationsgenauigkeiten

Vorwegnehmend sei an dieser Stelle gesagt dass durch Fusion in zwei von drei Fällen ein z.T. ein deutlich besseres Ergebnis erzielt werden konnte als ohne Fusion. Im Falle der Rank-Level Fusion stieg das Ergebnis auf 100% Genauigkeit an, während das durchschnittliche Ergebnis für die Decision-Level Fusion noch über 97% lag. Die Match-Level Fusion verschlechterte generell das Ergebnis.

Für die drei hier betrachteten Fusionsebenen erzielte die **Match-Level Fu**sion mit Abstand das schlechteste Klassifikationsergebnis (76,8% (*RS1*) und 66,9% (*RS2*); siehe Tabelle 3 und Tabelle 4). Dieses Ergebnis ist offenkundig, da hier nur eine Übereinstimmung beider Klassifikatoren zu einer Zuordnung führt. Tabelle 7 zeigt für den schlechtesten Fall für *RS1* (Aufnahmeposition R03) die Konfusionsmatrix ohne die "*uncertain*"-Symbole, welche aus der Anwendung von Gleichung (3) resultieren. Im Ergebnis ist der Wert für M_4 mit 31,8% nur geringfügig größer als der Wert für raten wäre (hier 25%). Demzufolge wäre hier eine geringe Konfidenz für diesen Fusionsansatz anzunehmen.

	M_1	M_2	M_3	M_4
M_1	**69,8%**	0,0%	0,0%	1,0%
M_2	0,0%	**86,0%**	0,5%	0,3%
M_3	2,3%	0,0%	**49,5%**	11,0%
M_4	2,3%	0,3%	6,0%	**31,8%**

Tabelle 7: Konfusionsmatix für den fusionierten Output für *RS1* in R03 (ohne "*uncertain*"-Symbole), das schlechteste Ergebnis ist in Grau markiert

Tabelle 8 zeigt für den schlechtesten Testfall für *RS2* (Aufnahmeposition R02; ohne "*uncertain*"-Symbole) die Konfusionsmatrix. Hier liegt das schlechteste Ergebnis (M_3 mit 34,8%) um ca. 21,5% über dem Wert für raten (14,3%).

	M_2	M_5	M_3	M_6	M_7	M_8	M_9
M_2	**90,3%**	0,0%	0,0%	0,0%	0,0%	0,3%	0,0%
M_5	0,0%	**93,8%**	1,8%	2,5%	7,5%	4,5%	3,8%
M_3	0,0%	0,0%	**34,8%**	10,8%	1,3%	2,0%	0,0%
M_6	0,0%	0,0%	10,3%	**41,3%**	1,0%	4,8%	0,0%
M_7	0,0%	0,0%	0,3%	0,3%	**44,5%**	2,3%	0,0%
M_8	0,0%	0,0%	0,0%	0,0%	0,0%	**38,5%**	0,0%
M_9	2,0%	0,0%	0,5%	0,3%	0,8%	0,5%	**83,3%**

Tabelle 8: Konfusionsmatix für den fusionierten Output für *RS2* in R02 (ohne "*uncertain*"-Symbole), das schlechteste Ergebnis ist in Grau markiert

Der Abfall der durchschnittlichen Klassifikationsgenauigkeiten (im Mittel von 76,8% für *RS1* auf 66.9% für *RS2*) lässt sich wahrscheinlich mit der erhöhten Anzahl der Mikrofone erklären. Allerdings könnte auch die höhere Ähnlichkeit der Mikrofone (besonders zwischen M_7 und M_8) in *RS2* einen Einfluss auf das Ergebnis haben – weitere Tests wären hier von Nöten um diese Frage zu klären.

Generell zeigt die **Rank-Level Fusion** die besten Testergebnisse. Für alle Aufnahmepositionen wurde hier eine Genauigkeit von 100% mit dem oben beschriebenen Ansatz erzielt (siehe Tabelle 3 und Tabelle 4). Tabelle 9 zeigt die fusionierte Rangmatrix für den schlechtesten Testfall in *RS1*.

Die Hauptdiagonalenelemente haben zwar alle den besten (kleinsten) Zeilenrang aber der minimale Abstand in einer Zeile ist nur 0,75 (M_3).

	M_1	M_2	M_3	M_4
M_1	**1**	3,75	3,5	3,25
M_2	4	**1**	3,25	3,75
M_3	2,75	2,5	**1**	1,75
M_4	2,25	2,75	2,25	**1,25**

Tabelle 9: Fusionierte Rangmatrix mit minimalen Abstand in *RS1* (R03)

Auf die Präsentation einer beispielhaften fusionierten Rangmatrix für in *RS2* wird an dieser Stelle aus Platzgründen verzichtet. Auch hier haben selbst im schlechtesten Testfall (R07) alle Hauptdiagonalenelemente den besten (kleinsten) Zeilenrang. Gleichfalls ist der minimale Abstand in einer Zeile ist nur 0,75 (in R07 für M_3, M_6, M_7 and M_9).Mit einer durchschnittlichen Klassifikationsgenauigkeit von 97% zeigt die D**ecision-Level Fusion** das zweitbeste Ergebnis nach der Rank-Level Fusion (siehe Tabelle 3 und Tabelle 4). Für *RS1* wird genau ein Mikrofon an genau einer Position falsch klassifiziert. Für RS2 passierte dies an zwei Aufnahmepositionen.

5 Zusammenfassung

Die Ergebnisse in diesem Beitrag geben erste Antworten auf die beiden Fragen: Was ist der erzielbare Nutzen von Fusion in der Mikrofonforensik? und: Welche der vorgestellten Fusionsstrategien zeigte den größten Nutzen? Wobei hier der Fokus ausschließlich auf dem Einfluss auf die Klassifikationsgenauigkeit lag.

Wie in den Testergebnissen gezeigt wurde liegt der Nutzen darin das ausgewählte Fusionsstrategien das Klassifikationsergebnis bis auf 100% erhöhen können. Die Antwort auf die zweite Frage liegt in der Identifikation der hier genutzten Rank-Level Fusionsstrategie als derjenigen Strategie die für das Problem der Mikrofonforensik, wie es hier betrachtet wurde, am besten geeignet erscheint.

Vergleicht man die hier erzielten Ergebnisse mit Bisherigen, so konnten wir durch den Einsatz von Fusion eine Erhöhung der Klassifikationsergebnisse von 75,99% für *RS1* (Kraetzer et al. [KOD07]) bzw. 93,5% für *RS2* (Buchholz et al. [BKD09] unter Nutzung von ausschließlich Frequenzraummerkmalen) auf 100% bei gleichen Testmengengrößen erzielen.

Betreffend abzuleitendem Forschungsbedarf, der auf den hier präsentierten Ergebnissen basiert, wird zeitnah in einer erweiterten Version dieses Beitrages sowohl eine Konfidenzabschätzung für die Fusionen (Betrachtung der Stabilität der fusionierten Entscheidung, bzw. Abstand der Entscheidung zur Entscheidungsschwelle in der Fusion) als auch eine Komplexitätsbetrachtung (d.H. Kosten der Fusion) vorgenommen. Ein weiterer wichtiger Punkt wäre die Evaluierung anderer Fusionsoperatoren (vor allem im Bereich der Match-Level Fusion).

Aus den hier vorgestellten Ergebnissen ergeben sich wichtige Fragen, wie: Welche Hardwareeigenschaften eines Mikrofons werden bei der Klassifikation ausgewertet? Oder: Sind die Merkmale auch geeignet Geräte gleichen Typs zu unterscheiden? Daher wären wichtige Arbeiten für die weitere Forschung auf diesem Gebiet unter anderem: die Erstellung und Evaluierung größerer Testmengen (vor allem mir mehr Mikrofonen), die Untersuchung der Möglichkeit Mikrofone derselben Baureihe statistisch zu trennen, sowie der Einfluss der einzelnen physikalischen Baugruppen sowie der Aufnamelokalität/-position auf die Performanz.

6 Acknowledgements

The work in this paper has been supported in part by the European Commission through the FP7 ICT Programme under Contract FP7-ICT-216736 SHAMAN. The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose.

7 Literaturverzeichnis

[KOD07] C. Kraetzer, A. Oermann, J. Dittmann and A. Lang: *Digital Audio Forensics: A First Practical Evaluation on Microphone and Environment Classification*. Proc. ACM Multimedia and Security Workshop 2007, ACM, 2007.

[RNJ06] A.A. Ross, K. Nandakumar, and A.K. Jain: *Handbook of Multibiometrics*. International Series on Biometrics. Springer Verlag, 2006.

[LHF03] N. Landwehr, M. Hall, and E. Frank: *Logistic Model Trees*, Proc.ECML'03, 2003

[Qu93] R. Quinlan: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[KD08a] C. Kraetzer and J. Dittmann: *Impact of Feature Selection in Classification for Hidden Channel Detection on the Example of Audio Data Hiding*. Proc. ACM Multimedia and Security Workshop 2008.

[MA07] R. C. Maher: *Acoustical Characterization of Gunshots*. Proc. Of. SAFE 2007, USA, 2007.

[FFG08] T. Filler, J. Fridrich, M. Goljan: *Using Sensor Pattern Noise for Camera Model Identification*. In Proc. ICIP08, 2008.

[GFF09] M. Goljan, J. Fridrich, and T. Filler: *Camera Identification-Large Scale Test*. Proc. SPIE, Electronic Imaging, Security and Forensics of Multimedia Contents XI, San Jose, USA, 2009.

[KMC08] N. Khanna, A.K. Mikkilineni, G.T. Chiu, J.P. Allebach, E.J. Delp: *Survey of Scanner and Printer Forensics at Purdue University*. IWCF08. LNCS vol. 5158, Springer, 2008.

[OVD07] A. Oermann, C. Vielhauer, J. Dittmann: *Sensometrics: Identifying Pen Digitizers by Statistical Multimedia Signal Processing*. In: SPIE Multimedia on Mobile Devices'07, USA, 2007.

[KD08b] C. Kraetzer, J. Dittmann: *Impact of Feature Selection in Classification for Hidden Channel Detection on the Example of Audio Data Hiding*. Proc. ACM Multimedia and Security Workshop, ACM, 2008.

[BKD09] R. Buchholz, C. Kraetzer, J. Dittmann: *Microphone Classification Using Fourier Coefficients*. Accepted for Information Hiding 2009.