



Automatische Bestimmung von 3D-Szenen aus 2D-Filmen

3D-Filme sind zurzeit in den USA der Renner. Die New York Times überschlug sich fast in ihren Ankündigungen zu „Beowulf“ und meinte gar, dass es nicht mehr lange dauern wird, bis alles in 3D erstellt wird. Und selbst große Produktionshäuser in den USA wollen angeblich nur noch in 3D aufnehmen. Im Zeichen von Home-HDTV will man wieder im Filmgeschäft einen Tick voraus sein. Aber man sollte wohl solche Ankündigungen sehr gelassen sehen – solche 3D-Hypes kommen in Wellen immer wieder. Ein Normalzustand wird das aus vielen Gründen noch sehr lange Zeit nicht werden. Aber Film ist auch – manche sagen nur – Geschäft und so möchte man natürlich auch ältere bestehende 2D-Kassenschlager – Blockbuster – als 3D-Filme in Zweitverwertung wieder anbieten können. Das aber wird sich finanziell nur rechnen wenn es gelingt, die dritte Dimension automatisch aus den

Filmbildern direkt ableiten zu können.

Nun haben sich im Rahmen ihrer Dissertationen zwei junge Wissenschaftler, Dr.-Ing. Matthias Kunter und Sebastian Knorr, TU Berlin, Fachgebiet Nachrichtenübertragung (**Bild 1**), dieser Thematik angenommen und sind bereits mit gutem Erfolg dabei, eine solche automatische Bestimmung des zugehörigen Zweitbildes für die Stereowirkung zu errechnen. Insbesondere im eigentlich schwierigsten Fall, nämlich wenn sich die Kamera bewegt, funktioniert das auch schon beeindruckend.

Technik

Für einen Betrachter, der im Zug fährt und aus dem Fenster schaut, rasen die Bäume des Vordergrunds scheinbar in Blitzeschnelle am Fenster vorbei, während sich ein weit entfernter Hintergrund kaum bzw. nur sehr

langsam zu ändern scheint. Jeder von uns kennt diesen Effekt. Würde man zwei dieser Momente als einzelne Bilder sehen, so ließe sich aus dieser sogenannten Bewegungsparallaxe die räumliche Anordnung der einzelnen Bildelemente ermitteln. Auf dieser Kenntnis basiert das entwickelte Verfahren, um aus bewegten 2D-Bildern automatisch 3D-Bilder zu generieren. Denn ähnliches wie eben bei der Zugfahrt beschrieben passiert auch, wenn eine Film- oder Videokamera bewegt wird. Die Unterschiede zwischen den zeitlich versetzten Bildern machen es möglich, für jeden einzelnen Bildpunkt ein in der räumlichen Parallaxe abweichendes zweites Bild zu erzeugen. Was sich hier allerdings von der Idee her sehr einfach anhört ist in der Realisierung höchst komplex.

Der theoretisch einfachste Fall wäre es, wenn die Kamera, die ja mit konstanter Bildfrequenz aufnimmt, mit einer vorbe-

rechneten Geschwindigkeit konstant parallel zur Szene starr aufnehmen würde. Dann nämlich wären im Prinzip gar keine Rechenoperationen nötig, man könnte einfach das zweite Bild nebeneinander horizontal etwas versetzt anordnen und eine dreidimensionale Wirkung erreichen. Dieser Fall ist aber in der Praxis vollkommen unrealistisch. Und so sieht das Verfahren auch vor, das die Kamerabewegung ungleichförmig und unabhängig von der Richtung sein kann. Die Kamera kann schwenken, sich neigen, sogar „wackeln“, langsam wie oft bei einer mit der Hand geführten Kamera „schweben“ und es darf gezoomt werden, denn auch dabei ist ja eine Bewegung des Inhalts der aufgenommenen Bilder vorhanden. Bloß eines darf die Kamera nicht, unverrückbar fest auf einem Stativ sitzen und starr die Szene filmen (für diesen Fall sind andere Algorithmen entwickelt worden).

Das Verfahren gliedert sich in zwei Bereiche. Im ersten Bereich steht die Analyse des Bildes bzw. der Bilder und im zweiten Bereich die Synthese eines neuen Bildes.

Die Analyse wird als „Structure-from-Motion“ bezeichnet. Das heißt, man sucht im Bild mit mehreren Erkennungs- und Punktverfolgungsalgorithmen nach gut strukturierten Kanten und ihre Lage im Bild und vergleicht die Abweichungen dieser Kanten mit dem Nachbar-



Bild 1. Dr. Matthias Kunter (links) und Sebastian Knorr von der TU Berlin, Fachgebiet Nachrichtenübertragung, die beiden Entwickler des Systems „3D aus 2D“

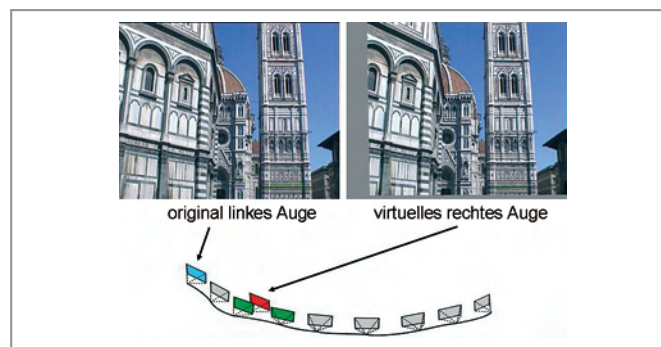


Bild 2. Originalbild für das linke Auge. Für das „Zweitaugenbild“ wurde, ausgehend vom ersten zeitlich dem Augenabstand entsprechenden grünen Bild, das virtuelle Bild für das rechte Auge berechnet

bild. Daraus kann man als erstes die Bewegung der Kamera im dreidimensionalen Raum ermitteln. **Bild 2** zeigt das in relativ einfacher Form. In einer Fahrscene erkennt man erst einmal die zeitliche und räumliche Lage der einzelnen Szenenbilder. Zu jedem einzelnen Szenenbild gilt es nun, das Bild zu finden, das das andere Auge eines Betrachters (bzw. die Kamera) sehen würde, ausgehend von der Augenparallaxe, die als Normalbetrachter mit 64 mm anzunehmen ist. Und das ist keineswegs zwingend unbedingt das nächste Bild, sondern kann auch das zweite, dritte oder x-te Bild sein, abhängig von der Geschwindigkeit der sich bewegenden Kamera. Die Software kann die Tiefenstaffelungen bzw. die Lageabweichungen der einzelnen Ermittlungspunkte nur relativ erfassen (**Bild 3**). Deshalb erfolgt für jede Szene gewissermaßen als Grundmaß die manuelle Einschätzung und Parametereingabe der realen Handlungsentfernung zum Kameraobjektiv anhand eines Keyframes am Beginn der Szene. Damit ist die Software in der Lage, die nötige Abweichung der zu betrachtenden Einzelpunkte zu ermitteln und das Bild zu finden, bei dem sich die Standard-Augenparallaxe (64 mm) ergibt. Der geschätzte Wert ist relativ unkritisch. Abweichungen vom tatsächlichen Wert erhöhen oder verringern nur die Tiefenwirkung und werden deshalb auch bewusst manchmal abweichend gewählt, um die Effektwirkung zu steigern oder zu reduzieren. Hat man das am nächsten kommende Bild ermittelt (im Bild 2 sind die beiden dem realen Zeitpunkt am nächsten liegenden Bilder grün gekennzeichnet) so wird dieses Bild mit den relativ einfachen sich aus dem Strahlengesetz ergebenden perspektivischen Änderungen, die sich aus der örtlichen Änderung der Kamera ergeben, verzerrt bzw. entzerrt und es ergibt sich als Konsequenz das zweite (3D-)Bild zu jedem einzelnen rea-

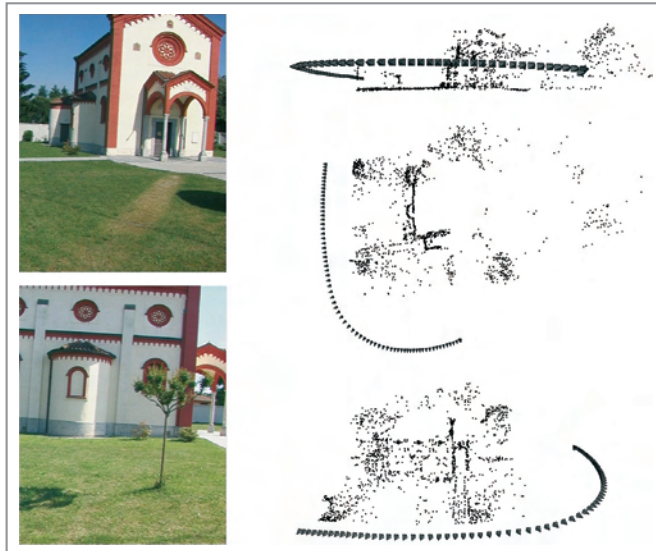


Bild 3. Zwei Originalbilder (Anfangs- und Endbild einer Szene) mit den automatisch ermittelten Erkennungspunkten und der daraus abgeleiteten Kamerabewegung (rechts)

len Bild. Da aber bei einer örtlichen Kameraänderung auch Teile des ersten Bildes beim Bild für das zweite Auge möglicherweise nicht mehr sichtbar sind (siehe grauer linker Streifen im Bild 2 des virtuellen Bildes), werden diese Teile aus Teilen der vorhergehenden Bilder rekonstruiert und hinzugefügt.

Das bedeutet, dass die Berechnung des Bildes für das zweite Auge kein Renderingprozess im klassischen Sinne darstellt, sondern dass das Bild überwiegend „nur“ geometrisch geändert wird, was den Rechenprozess deutlich reduziert. Es können deshalb auch normale Standard-PCs verwendet werden, und es sind prinzipiell keine Super-Hochleistungsrechner nötig. Der besondere Vorteil dieser Methode ist natürlich, dass nach Festlegung gewisser Grundparameter

die Erzeugung des zweiten Bildes, das die Stereowirkung ergibt, vollautomatisch ablaufen kann und ein hohes Maß an Realistik ergibt. Die gesamte Systemkette zeigt **Bild 4**.

Die beiden Wissenschaftler zeigten eindrucksvolle Beispiele dafür, wie insbesondere bei Landschaftsaufnahmen aus dem Hubschrauber und ähnlichen Sujets eine geradezu überwältigend realistische dreidimensionale Wiedergabe erreicht wird.

Das Verfahren der 3D-Simulation aus 2D-Bildern mit bewegter Kamera hat allerdings starke Limitierungen, sodass es mit weiteren Maßnahmen ergänzt werden muss, an deren Realisierung man zurzeit ebenfalls arbeitet. Denn Szenen mit statischen Motiven und bewegter Kamera gibt es nach einer internen überschlagsmäßigen Wertung nur bei

2 % bis 5 % aller Szenen in einem Spielfilm. Bewegt sich auch das Motiv selbst, wird es schwieriger, denn zwischen dem realen Bild und dem aus einem nachfolgenden Bild ermittelten „Zweitaugenbild“ liegt ja eine zeitliche Differenz. Und bei zum Beispiel schnell und deutlich erkennbaren, bewegenden Personen im Vordergrund würde das geometrisch veränderte „Zweitaugenbild“ die Bewegung in einem anderen zeitlichen Moment darstellen (zum Beispiel Bild real „Arm unten“, Zweitaugenbild „Arm oben“). Für diesen Fall ist man zurzeit dabei, mit verschiedenen Erkennungsmethoden die Personen elektronisch aus dem ersten Bild „herauszuschneiden“ und sie in das „Zweitaugenbild“ wieder einzusetzen. Hier sind noch Einschränkungen in der Anwendung gegeben, ebenso wie im Falle einer vollkommen unbewegten Kamera. Letzteres sehen die beiden Wissenschaftler allerdings weniger kompliziert, denn es gibt verschiedene Methoden, bei einem statischen Kamerabild die relative Lage von Elementen zu bestimmen, dazu gehören Texturbestimmungsverfahren, Ableitungen aus der (Tiefen-)Schärfe des Bildes und ähnlichem. Sebastian Knorr entwickelt ausgehend von seiner zurzeit gerade laufenden Dissertation ein Verfahren auf Basis der Fluchtpunkt- und der Interpositionsbestimmung (Vordergrund verdeckt Hintergrund) auch diesen ei-

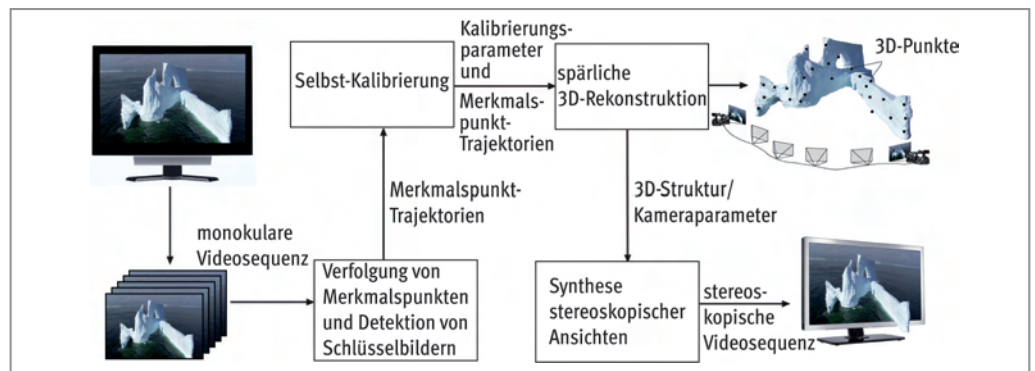


Bild 4. Einzelschritte des Gesamtsystems bestehend aus Analyse- und Sythese-Elementen

gentlich im filmischen Bereich häufigeren Fall abzudecken.

Vermarktung

Das Verfahren wurde patentiert, und die beiden Wissenschaftler haben zur Vermarktung dieses (und anderer) Verfahren die Firma imcube Media (zusammen mit Prof. Dr. Thomas Sikora von der TU Berlin und den Juristen Prof. Dr. Walter Rust) als Spin-off gegründet. Sie wollen die 3D-Erstellung von 2D-Bildsequenzen als professionellen Dienstleistungsservice anbieten. Es gibt in den USA zwei Firmen, die ähnliches machen, allerdings mit extrem hohen Personalaufwand, der seinen Niederschlag in exorbitanten Preisen findet.

Der zurzeit aufgebaute Prototyp bietet die automatische Generierung der „Zweitaugenbilder“ bei entsprechend geeigneten Szenen in etwa 300- bis 900-facher Realzeit (eine Sekunde Film etwa 5 bis 15 Minuten). Das hört sich enorm an, allerdings sind sich die beiden Wissenschaftler einig, dass man schon in wenigen Monaten durch einfache Programmoptimierung etwa 10- bis 20-facher Realzeit bei der Bearbeitung mit einem Standard-PC erreichen wird, und man muss das sicher auch im Vergleich sehen. Die Bearbeitung einer Szene bei dem besagten amerikanischen Unternehmen dauert mit dem hohen Personalaufwand etwa 4 bis 5 Tage.

Im übrigen wurden die beiden Wissenschaftler im März 2008 für ihre Geschäftsidee im Bereich der 3D-Technik mit dem mit 25.000 Euro dotierten Gründerpreis „Mit Multimedia erfolgreich starten“ auf der CeBit ausgezeichnet. Zusätzlich wurde die Ausgründung aus der TU Berlin mit einem Sonderpreis in Höhe von 5.000 Euro ausgezeichnet, den die Loewe AG zu dem Schwerpunktthema „Multimedia in der Unterhaltungselektronik“ ausgelobt hatte.

Norbert Bolewski

Optischer Datenspeicher der 4. Generation

Nachtrag zum gleichnamigen Artikel in der FKT 11/2007

Im Rahmen des deutsch-französischen Projektes 4GOOD („4th-Generation, Omni-Purpose Optical Disc-system“) werden Basistechnologien für die 4. Generation optischer Discs entwickelt. In der FKT 11/2007 erschien ein Artikel hierüber, und auf der Jahrestagung der FKTG im Mai 2008 wurde nun über die aktuellen Entwicklungen in 4GOOD berichtet. Nachfolgend letzte Neuigkeiten über das Projekt, das vom BMWi bzw. in Frankreich vom MINÉFI gefördert wird.

Schlüsselkomponenten für die optische Datenspeicherung

Optische Disc

Für die SuperRENS-Schicht wurden Phasenwechsel- und Halbleitermaterialien untersucht. Phasenwechselmaterialien wie AIST (AgInSbTe) ändern unter

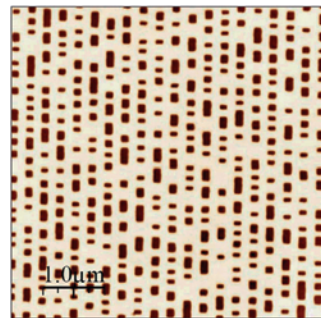


Bild 1. Ausschnittbild des Substrats einer Testdisc mit einem Spurabstand von 260 nm (mittels Atomic Force Microscope aufgenommen)

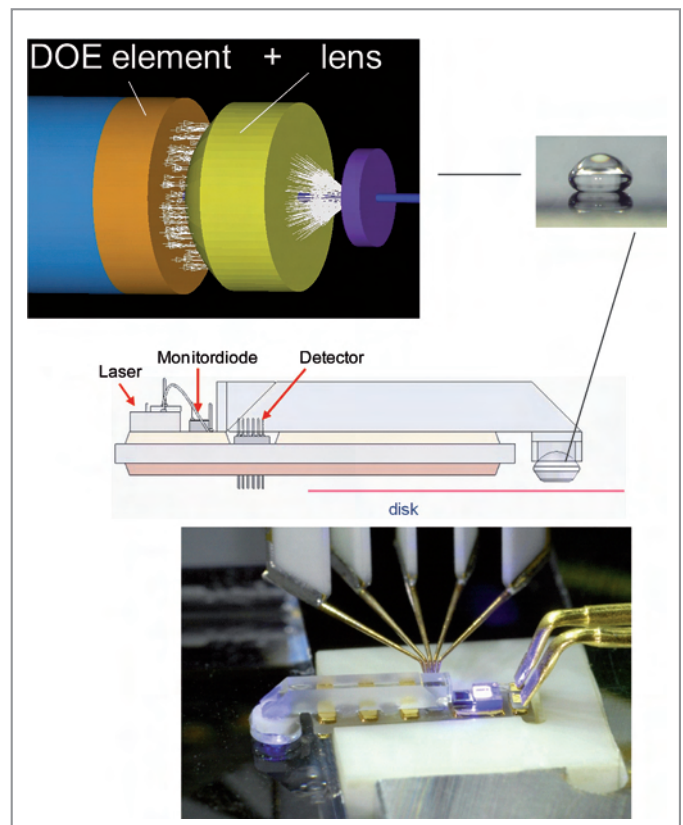
Bild 2. Optische Einheit des Laufwerks (Foto: Fraunhofer IOF, Jena) zusammen mit DOE (Diffractive Optical Element) und Miniaturlinse zur Fokussierung des Laserstrahls

Einwirkung des Laserstrahls ihren Zustand zwischen amorph und kristallin und werden „durchsichtig“, während bei Halbleitermaterialien wie InSb lokal und reversibel freie Ladungsträger angeregt werden, damit eine Metallisierung eintritt und die Reflektivität verändert wird – ohne Änderung der atomaren Struktur. In beiden Fällen kann die darunterliegende Datenschicht innerhalb einer kleinen Apertur gelesen werden. Phasenwechselmaterialien sind intensiv untersucht worden, und das IPI der RWTH Aachen hat eine „Schatzkarte“ vorgelegt, die es erlaubt, aus der Zusammensetzung optische Eigenschaften wie den Kontrast vorherzusagen.

InSb bietet gegenüber AIST den Vorteil, dass dasselbe CNR bei niedrigerer Laserleistung erzielt wird, außerdem bietet InSb eine höhere Signalbandbreite. In

Langlebigkeitstests zeigte sich, dass selbst nach 12.000 Lesezyklen die Bitfehlerrate nach der Signaldetektion und vor der Fehlerkorrektur so gering ist, dass die Daten einwandfrei gelesen werden können.

Um die Speicherdichte gegenüber der Blu-ray Disc zu erhöhen, wurde neben der minimalen Mark-Länge (und damit Bitlänge) der Spurabstand variiert (**Bild 1**). Bei einem Spurabstand von 320 nm wurde eine minimale Mark-Länge von 60 nm realisiert und damit die 2,5-fache Speicherdichte einer Blu-ray Disc erreicht. Durch Kombination mit dem verringerten Spurabstand von 260 nm wird sich die 3-fache Speicherdichte erzielen lassen. Derzeit werden Ansätze zur weiteren Verringerung des Spurabstandes untersucht, um letztlich einen Faktor 4 zu erreichen.



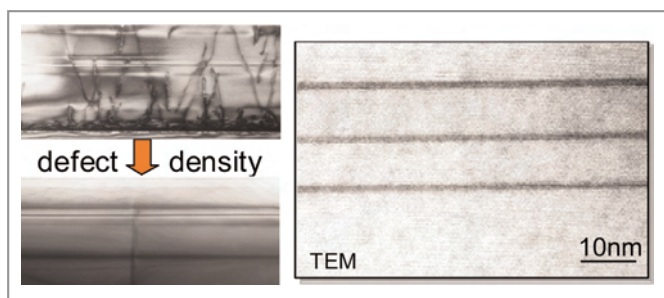


Bild 3. Verringerung der Defektdichte im Schichtensystem der blauviolett emittierenden Laserdiode (links) und TEM-Bild der lichterzeugenden Quantentröge (rechts)

Konzept des Miniaturlaufwerks

Der Schwingarmaktuator für das Miniaturlaufwerk ist leichter als ein konventioneller Aktuator, hat einen niedrigeren Energiebedarf und eine hohe Stabilität. Es treten keine thermischen Einflüsse von Seiten der Linse auf. Die Komponenten des Pick-ups am Ende des Schwingarms wurden inzwischen realisiert und auf dem gemeinsamen Substrat montiert und vermessen (**Bild 2**). Servo-Verfahren für die Fokussierung und die Spurnachführung wurden ebenfalls entwickelt.

Um einen konventionellen und den Schwingarmaktuator zu testen, wurde ein optisches Laufwerk als Tester für 12- und für 3-cm-Discs mit zahlreichen Messmöglichkeiten entwickelt.

Blauvioletter Halbleiterlaser zum Auslesen der Daten

Die Defektdichte innerhalb des Schichtensystems der Laserdiode konnte, wie gefordert, um etwa drei Größenordnungen gesenkt werden (**Bild 3 links**). Die in der aktiven Zone im Zentrum der Wellenleiterstruktur liegenden lichterzeugenden Quantentröge ha-

ben gleichbleibende Schichtdicken von 2 nm (**Bild 3 rechts**). Die Laserdiode mit reduziertem Leistungsbedarf ist verfügbar.

Miniaturlinse zur Fokussierung des Laserstrahls

Für die Linse zur Fokussierung des Laserstrahls wurde von Schott ein hybrides Design aus einem DOE (Diffractive Optical Element) und einer Miniaturlinse gewählt (s. Bild 2). Das DOE bewirkt eine Vorformung des Lichtstrahls, sodass die Anforderungen an die Oberflächengenauigkeit der Linse entspannter sind. Nach Optimierung der Presswerkzeuge und des Blankpressprozesses liegt die Linse – mit weniger als 1 mm Durchmesser – nunmehr mit einer praxistauglichen Formtoleranz von 18 nm RMS vor, und das DOE ist ebenfalls verfügbar. Das Licht, das die Linse verlässt, weist einen Wellenfrontfehler von $\leq \lambda/30$ auf.

Beschleunigungssensoren für den Laufwerksschutz

Für den Beschleunigungssensor fiel die Entscheidung für ein hybrides Design aus einem MEMS-Chip (mikroelektromechanisches

System) als eigentliches Sensorelement und einem ASIC mit der Auswerteelektronik. Beide Bauelemente sind entworfen und hergestellt worden, die Integration zu einem gemeinsamen Bauelement läuft derzeit.

Signalverarbeitung für die Datenaufzeichnung und -wiedergabe

Die verwendeten Verfahren der Signalverarbeitung wurden an die Eigenschaften des optischen Kanals angepasst. Bei der Bit-Detektion und der RLL-Demodulation kommen „Soft Decision“-Verfahren zum Einsatz. Die Fehlerkorrektur (FEC) ist zweistufig: Der innere Decoder (Turbo Product Code oder Low-Density Parity-Check Code) mit Soft-Decision dient der Korrektur von Einzelfehlern, der äußere Reed-Solomon-Decoder der von Bündelfehlern. Die Coderate wird insgesamt ähnlich der dritten Generation sein, bei der Aufteilung zwischen innerem und äußeren Code ist ein Trade-off zu wählen.

Die Leistung der Signalverarbeitung zeigt **Bild 4** anhand eines Beispiels. Selbst die kurzen sogenannten 2T-Symbole, die im Kanalausgangssignal nur als geringfügige Signaländerungen zu erkennen sind, werden hier fehlerfrei detektiert.

Ausblick

Die Projektergebnisse werden auf der IFA 2008 (29. August bis 3. September 2008) am Stand des BMWi präsentiert.

HD-Satelliten-Ü-Wagen für Media Broadcast

ND SatCom hat für die Media Broadcast GmbH den Bau und die Systemintegration von neun redundanten Übertragungsfahrzeugen durchgeführt. Drei Fahrzeuge (Sprinter 518 CDI) sind mit A/V-Produktionssequenzequipment ausgerüstet und sechs (Sprinter 515 CDI) sollen als Uplinkfahrzeuge eingesetzt werden. Es wurden Systeme beauftragt, die universell und flexibel in verschiedenen Anwendungsszenarien einsetzbar sind.

Die SNG-Technik besteht jeweils aus einem 1,9 m (diag.) Antennenspiegel (SkyRAY MAS 1900), komplett redundantem Sendezug mit 400 W oder 750 W Sendeverstärker (750 W HPA bei Uplinkfahrzeugen) sowie HD/SD-fähigen Encodern (E5788 DVBS2) und IRDs (RX1290) von Tandberg. Alle Geräte lassen sich über eine Bediensoftware oder per Hand steuern. Klimatisiert wird über eine Raumklimaanlage und Gestellklimageräte. Das Fahrzeug lässt sich wahlweise über externes Netz oder über Generator betreiben (Unterflurgenerator und Fischer Panda 14 kVA).

Alle Fahrzeuge sind mit HD/SD-Testgeneratoren und Waveform-Monitoren (Tektronix) und einer Signal-Processing-Plattform (Lynx) ausgestattet. Sony-Monitoren (LMD-9050), 19-inch-Monitoren der „Grey Owl Serie“ von Tamuz und ein Multiviewer von Zandar ermöglichen die Darstellung von sechs bzw. zwölf Videosignalen. Geschaltet werden die Signale mit einem „Sublime“-System (16x16) von Network. Für die Kommunikation werden analoge 4- bzw. 8-Kanal-Intercom-Stationen (4-Draht) von Zähl eingesetzt.

Als Videomischer wird ein 8-Kanal-HD/SD-Mischer von Panasonic (AV-HS400) mit eingebauten Frame-Synchronisierern und Up-Convertern genutzt. Der Ton läuft über ein digitales Mischpult (Yamaha 01V) mit bis zu zwei drahtlosen Mikrofonstrecken. Vier elektronisch und mechanisch vorbereitete, flexibel bestückbare Kamerazüge sind vorhanden.

Das Uplink-Konzept arbeitet mit SDI-Signalen und embedded Audio. Eingangsseitig werden alle gängigen Signalquellen (HD/SD-SDI, FBAS, AES/EBU und analog Audio) mithilfe von Konvertern und Embeddern zusammengefasst. Über den Waveform-Monitor kann das gewünschte Audiosignal analog auf die Abhörlautsprecher gegeben werden. Vier zusätzliche IRDs können einbaubar, um insgesamt acht Downlinks zu realisieren.

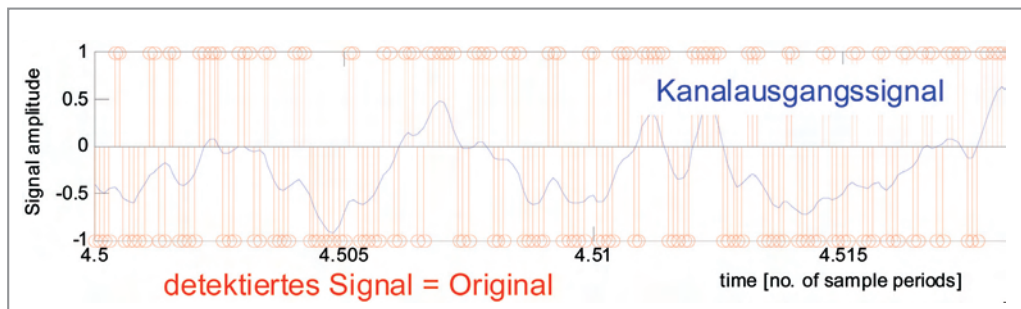


Bild 4. Beispiel eines Kanalausgangssignals und des zurückgewonnenen Original-NRZI-Signals nach Bit-Detektion (min. Mark-Länge 60 nm, Laser-Leseleistung 2 mW)