Abschlußbericht für die Deutsche Forschungsgemeinschaft

1 Allgemeine Angaben

Abschlußbericht zum DFG-finanzierten Forschungsvorhaben Gi 198/5-1.

1.1 Antragsteller

Bernd Girod, Prof. Dr.-Ing.
Ordentlicher Universitätsprofessor
1.12.1957, deutsch
Lehrstuhl für Nachrichtentechnik I:
Signalverarbeitung und Multimedia-Kommunikation
Universität Erlangen-Nürnberg

dienstliche Adresse:	Cauerstraße 7 91058 Erlangen Tel. 09131-85-7100 FAX 09131-85-8849
private Adresse:	Am Köhler-Hof 6 91080 Spardorf Tel. 09131-50089

Lebenslauf und Schriftenverzeichnis liegen der Deutschen Forschungsgemeinschaft vor.

1.2 Thema

Bewegungskompensierte Langzeitprädiktion zur Übertragung von Bewegtbildsequenzen bei niedrigen Bitraten

1.3 Kennwort

Bewegungskompensierte Langzeitprädiktion

1.4 Fachgebiet und Arbeitsrichtung

Elektrotechnik, Bildkommunikation

1.5 Antragszeitraum, Förderungszeitraum

Berichtszeitraum: 15.03.1997 - 31.08.1998, 01.09.1998 - 14.03.2000
Förderungszeitraum: 15.03.1997 - 14.03.1999, 15.03.1999 - 14.03.2000

1.6 Veröffentlichungen aus dem Vorhaben

Beiträge in Fachzeitschriften

[1] G. J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression", *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[2] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, Feb. 1999.

[3] B. Girod, "Efficiency Analysis of Multi-Hypothesis Motion-Compensated Prediction", *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, Feb. 2000.

[4] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-Resilient Video Transmission Using Long-Term Memory Motion-Compensated Prediction", *IEEE Journal on Selected Areas in Communications*, June 2000.

Beiträge auf internationalen Konferenzen

[5] T. Wiegand, M. Flierl, and B. Girod, "Entropy-Constrained Design of Quadtree Video Coding Schemes", in *Proceedings of the IEE International Conference on Processing and its Applications*, Dublin, Ireland, July 1997, pp. 36–40.

[6] T. Wiegand, X. Zhang, and B. Girod, "Block-Based Hybrid Video Coding Using Motion-Compensated Long-Term Memory Prediction", in *Proceedings of the Picture Coding Symposium*, Berlin, Germany, Sept. 1997, pp. 153–158.

[7] T. Wiegand, X. Zhang, and B. Girod, "Motion-Compensating Long-Term Memory Prediction", in *Proceedings of the IEEE International Conference on Processing*, Santa Barbara, USA, Oct. 1997, vol. 2, pp. 53–56.

[8] T. Wiegand, E. Steinbach, A. Stensrud, and B. Girod, "Multiple Reference Picture Coding using Polynomial Motion Models", in *Proceedings of the SPIE Conference on Visual and Image Processing*, San Jose, USA, Feb. 1998, pp. 134–145.

[9] M. Flierl, T. Wiegand, and B. Girod, "A Local Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction", in *Proceedings of the IEEE Data Compression Conference*, Snowbird, USA, Mar. 1998, pp. 239–248.

[10] T. Wiegand, M. Flierl, and B. Girod, "Entropy-Constrained Linear Vector Prediction for Motion-Compensated Video Coding", in *Proceedings of the IEEE International Symposium on Information Theory*, Boston, USA, Aug. 1998, p. 409.

[11] B. Girod, T. Wiegand, E. Steinbach, M. Flierl, and X. Zhang, "High-Order Motion Compensation for Low Bit-Rate Video", in *Proceedings of the European Signal Processing Conference*, Island of Rhodes, Greece, Sept. 1998, pp. 253–256, invited paper.

[12] B. Girod, "Why B-Pictures Work: A Theory of Multihypothesis Motion-Compensated Prediction", in *Proceedings of the IEEE International Conference on Image Processing*, 1998.

[13] T. Wiegand, B. Lincoln, and B. Girod, "Fast Search for Long-Term Memory Motion-Compensated Prediction", in *Proceedings of the IEEE International Conference on Processing*, Chicago, USA, Oct. 1998, vol. 3, pp. 619–622.

[14] E. Steinbach, T. Wiegand, and B. Girod, "Using Multiple Global Motion Models for Improved Block-Based Video Coding", in *Proceedings of the IEEE International Conference on Processing*, Kobe, Japan, Oct. 1999, vol. 2, pp. 56–60.

[15] T. Wiegand, E. Steinbach, and B.Girod, "Long-Term Memory Prediction Using Affine Motion Compensation", in *Proceedings of the IEEE International Conference on Processing*, Kobe, Japan, Oct. 1999, vol. 1, pp. 51–54.

[16] M. Flierl, T. Wiegand, and B. Girod, "A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction", in *Proceedings of the SPIE Conference on Visual and Image Processing*, Perth, Australia, June 2000.

[17] M. Flierl, T. Wiegand, and B. Girod, "Rate-Constrained Multi-Hypothesis Motion-Compensated Prediction for Video Coding", in *Proceedings of the IEEE International Conference on Processing*, Vancouver, Canada, Oct. 2000.

[18] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Long-Term Memory Motion-Compensated Prediction for Robust Video Transmission", in *Proceedings of the IEEE International Conference on Processing*, Vancouver, Canada, Sept. 2000.

Standardisierungsbeiträge

[19] ITU-T/SG16/Q15-C-11, T. Wiegand, X. Zhang, and B.Andrews, "Long-Term Memory Motion-Compensated Prediction", Download via anonymous ftp to: standard.pictel.com/video-site/9712_Eib/q15c11.doc, Dec. 1997.

[20] ITU-T/SG16/Q15-D-13, T. Wiegand and B. Andrews, "An Improved H.263-Codec Using Rate-Distortion Optimization", Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d13.doc, Apr. 1998.

[21] ITU-T/SG16/Q15-D-54, T. Wiegand, B. Girod, and B. Andrews, "Long-Term Memory Motion-Compensated Prediction", Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d54.doc, Apr. 1998.

[22] ITU-T/SG16/Q15-D-55, T. Wiegand, B. Lincoln, B. Girod, and B. Andrews, "Fast Search for Long-Term Memory Motion-Compensated Prediction", Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d55.doc, Apr. 1998.

[23] ITU-T/SG16/Q15-E-25, T. Wiegand, B. Girod, and B. Andrews, "Results of the Core Experiment on Long-Term Memory Motion-Compensated Prediction", Download via anonymous ftp to: standard.pictel.com/video-site/9807_Whi/q15e25.doc, July 1998.

[24] ITU-T/SG16/Q15-E-44, T. Wiegand, N. Färber, B. Girod, and B. Andrews, "Long-Term Memory Motion-Compensated Prediction for Surveillance Applications", Download via anonymous ftp to: standard.pictel.com/ video-site/ 9807_Whi/ q15e44.doc, July 1998.

[25] ITU-T/SG16/Q15-E-52, T. Wiegand and B. Andrews, "Core Experiment Description for Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/ video-site/ 9807_Whi/ q15e52.doc, July 1998.

[26] ITU-T/SG16/Q15-F-32, T. Wiegand, N. Färber, B. Girod, and B. Andrews, "Proposed Draft for Annex on Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/video-site/9807_Seo/q15f32r1.doc, Nov. 1998.

[27] ITU-T/SG16/Q15-F-33, T. Wiegand, E. Steinbach, B. Girod, and B. Andrews, "Video Coding Using Long-Term Memory and Affine Motion Compensation", Download via anonymous ftp to: standard.pictel.com/video-site/9807_Seo/q15f33.doc, Nov. 1998.

[28] ITU-T/SG16/Q15-G-18, T. Wiegand, N. Färber, B. Girod, and B. Andrews, "Proposed Draft for Annex U on Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/video-site/9902_Mon/q15g18.doc, Feb. 1999.

[29] ITU-T/SG16/Q15-G-19, T. Wiegand, B. Girod, and B. Andrews, "Core Experiment Description for Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/video-site/9902_Mon/q15g19.doc, Feb. 1999.

[30] ITU-T/SG16/Q15-G-20, T. Wiegand, B. Girod, and B. Andrews, "Results for Core Experiment on Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/video-site/9902_Mon/q15g20.doc, Feb. 1999.

[31] ITU-T/SG16/Q15-G-21, T. Wiegand, E. Steinbach, B. Girod, and B. Andrews, "Video Coding Using Long-Term Memory and Affine Motion Compensation", Download via anonymous ftp to: standard.pictel.com/ video-site/9902_Mon/q15g21.doc, Feb. 1999.

[32] ITU-T/SG16/Q15-H-21, T. Wiegand, E. Steinbach, B. Girod, and B. Andrews, "Further Work on Affine Multiframe Motion Compensation", Download via anonymous ftp to: standard.pictel.com/ video-site/9908_Ber/q15h21.doc, Aug. 1999.

[33] ITU-T/SG16/Q15-H-24, T. Wiegand, N. Färber, and B. Girod, "Error Resilient Transmission Using Long-Term Memory Motion-Compensated Prediction", Download via anonymous ftp to: standard.pictel.com/ video-site/9908_Ber/q15h24.doc, Aug. 1999.

[34] ITU-T/SG16/Q15-H-30, T. Wiegand, "Proposed Draft for Annex U on Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/ video-site/9908_Ber/q15h30.doc, Aug. 1999.

[35] ITU-T/SG16/Q15-H-34, T. Wiegand and E. Steinbach, "Description of Affine Core Experiment", Download via anonymous ftp to: standard.pictel.com/ video-site/9908_Ber/q15h34.doc, Aug. 1999.

[36] ITU-T/SG16/Q15-I-42, T. Wiegand and E. Steinbach, "Core Experiment Description on Affine Multiframe Motion Compensation", Download via anonymous ftp to: standard.pictel.com/ video-site/9910_Red/q15i42.doc, Oct. 1999.

[37] ITU-T/SG16/Q15-I-43, T. Wiegand and E. Steinbach, "Results of Core Experiment on Affine Multiframe Motion Compensation", Download via anonymous ftp to: standard.pictel.com/ video-site/9910_Red/q15i43.doc, Oct. 1999.

[38] ITU-T/SG16/Q15-I-44, T. Wiegand, "Proposed Draft for Annex U on Enhanced Reference Picture Selection", Download via anonymous ftp to: standard.pictel.com/ video-site/9910_Red/q15i44.doc, Oct. 1999.

[39] ITU-T/SG16/Q15-I-45, T. Wiegand, "Multiframe Buffering Syntax for H.26L", Download via anonymous ftp to: standard.pictel.com/ video-site/9910_Red/q15i45.doc, Oct. 1999.

[40] ITU-T/SG16/Q15-J-33, T. Wiegand, "Modification of Annex U for Enhanced Error Resilience", Download via anonymous ftp to: standard.pictel.com/ video-site/0005_Osa/q15j33.doc, May 2000.

[41] ITU-T/SG16/Q15-J-49, T. Wiegand, "Proposed Editorial Changes to Annex U", Download via anonymous ftp to: standard.pictel.com/ video-site/0005_Osa/q15j49.doc, May 2000.

[42] ITU-T/SG16/Q15-J-50, T. Wiegand, "Interoperability of Annex U with other Annexes", Download via anonymous ftp to: standard.pictel.com/ video-site/0005_Osa/q15j50.doc, May 2000.

[43] ITU-T/SG16/Q15-J-51, T. Wiegand, "Test Model Description for an Annex U Encoder", Download via anonymous ftp to: standard.pictel.com/ video-site/0005_Osa/q15j51.doc, May 2000.

[44] ITU-T/SG16/Q15-J-66, T. Wiegand and M. Nilsson, "Annex U including Picture Numbers", Download via anonymous ftp to: standard.pictel.com/ video-site/0005_Osa/q15j66.doc, May 2000.

2 Arbeits- und Ergebnisbericht

2.1 Ausgangsfragen und Zielsetzung des Projektes

Die derzeit standardisierten Videocodierverfahren für niedrige Bitraten beruhen auf der Kombination aus bewegungskompensierter Prädiktion und Transformationscodierung des Prädiktionsfehlerbildes. Zur Bewegungskompensation wird nur das zuletzt decodierte Bild herangezogen, wodurch die Wirksamkeit der bewegungskompensierten Prädiktion eingeschränkt ist. Die Information aller früheren Bilder bleibt ungenutzt. In diesem Vorhaben sollte durch bewegungskompensierte Langzeitprädiktion die Information aus mehr als einem zuvor decodierten Bild benutzt werden.

Ziel des Vorhabens ist die Entwicklung eines verbesserten Codierverfahrens für die Übertragung von Bewegtbildsequenzen über Punkt-zu-Punkt-Verbindungen bei niedrigen Bitraten. Dazu sieht der Ansatz zur Langzeitprädiktion vor, die zwei Komponenten des Vektors der Bild-zu-Bild-Verschiebung um einen dritten Parameter zu ergänzen, der eine variable Zeitverzögerung repräsentiert. Es sollen vorangegangene Bilder über einen Zeitraum von mehreren Sekunden im Langzeitspeicher aufbewahrt werden und über eine Kombination von Verschiebungsvektor und Zeitverzögerung adressiert als Prädiktionssignal zur Verfügung stehen. Dazu muß die Zeitverzögerung zusätzlich als Seiteninformation übertragen werden.

Im Laufe des Forschungsvorhabens sind Erweiterungen des Langzeitprädiktionsverfahrens entwickelt worden, die als Ziel eine weitere Erhöhung der Bitrateneinsparungen hatten. Zum einen wurde ein Verfahren zur Bewegungskompensation mit Referenzbild-Vorverzerrung entwickelt, wobei zusätzlich zu den decodierten Bildern des Langzeitprädiktionscoders auch vorverzerrte Bilder benutzt werden. Affine Bewegungsmodelle werden zur Vorverzerrung benutzt, die als Seiteninformation übertragen werden müssen. Zum anderen wurde das Verfahren der Multihypothesen-Langzeitprädiktion entwickelt, wobei eine Linearkombination einer beliebigen Anzahl von Blöcken, die aus dem gesamten Langzeitprädiktionssuchraum stammen können, als Prädiktionssignal benutzt wird. Dabei werden die zu überlagernden Blöcke durch Verschiebungsvektoren und variable Zeitverzögerungsparameter adressiert.

Bei allen drei Verfahren wird eine verbesserte Bewegungskompensation durch eine Erhöhung der Bitrate für die Seiteninformation erzielt. Damit stellte sich das zentrale Problem, die Bitratenkontrolle der neuen Verfahren in einer Hybridcoderumgebung so zu entwerfen, daß die Bitrateneinsparungen bei der Prädiktionsfehlercodierung die erhöhte Bitrate für die Seiteninformation überwiegen. Weiterhin stellte sich das Problem einer drastischen Erhöhung des Aufwandes für die Bewegungsschätzung für die neuen Verfahren. Um Langzeitprädiktion auch bei gestörten Übertragungskanälen benutzen zu können, sollte auch untersucht werden, wie die Bitratenkontrolle des Coders für diesen Fall angepaßt werden kann, um die Fehlerfortpflanzung im Langzeitspeicher berücksichtigen zu können.

Die entwickelten Verfahren sollten bei der ITU-T/SG16/Q.15-Gruppe sowohl als Erweiterung des H.263-Standards als auch für das Langzeit-Projekt der ITU-T, H.26L, vorgeschlagen werden.

2.2 Durchgeführte Arbeiten

In dem Forschungsvorhaben zur bewegungskompensierten Langzeitprädiktion wurden neue Verfahren zur Videocodierung mit Rate-Distortion-Ansätzen entwickelt und durch das Einfügen in Hybridcoder auf ihre Leistungsfähigkeit hin untersucht. Die Ergebnisse des Vorhabens wurden in vier Zeitschriftenartikeln [4, 3, 2, 1] und in 14 Beiträgen auf internationalen Tagungen [18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5] publiziert.

Weiterhin sind aus dem Vorhaben 26 Beiträge zur internationalen Standardisierung entstanden [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44]. Die Mitwirkung der Arbeitsgruppe des Antragstellers bei der ITU-T/SG16/Q.15-Gruppe hat unter anderem dazu geführt, daß der im Vorhaben bearbeitete Ansatz zur bewegungskompensierten Langzeitprädiktion als Erweiterung des H.263-Videocodierstandards akzeptiert wurde. Das neue Projekt der ITU-T, H.26L, beinhaltet ebenfalls bewegungskompensierte Langzeitprädiktion.

2.2.1 Entwicklung codiereffzienter Verfahren zur Langzeitprädiktion

Im Forschungsvorhaben wurde das Verfahren zur bewegungskompensierten Langzeitprädiktion entwickelt [2, 11, 7, 6], bei dem die zwei Komponenten des Vektors der Bild-zu-Bild-Verschiebung („Verschiebungsvektor“) um einen dritten Parameter ergänzt werden, der eine variable Zeitverzögerung repräsentiert. Als Referenzsignale für die bewegungskompensierte Prädiktion wurden vorangegangene Bilder über einen Zeitraum von mehreren Sekunden im Langzeitspeicher aufbewahrt und über eine Kombination von Verschiebungsvektor und Zeitverzögerung adressiert.

Verschiebungsvektor und Zeitverzögerung müssen aber als „Seiteninformation“ übertragen werden und beanspruchen einen Teil der Bitrate. Langzeitprädiktion ist nur dann effektiv, wenn sich die bewegungskompensierte Prädiktion dadurch so verbessern läßt, daß die erhöhte Bitrate für die Seiteninformation gerechtfertigt ist. Daher wurde der im Abschnitt 2.1 des Erstantrags erläuterte iterative Entwurfsalgorithmus für Vektor-Quantisierer mit Entropie-Nebenbedingung (ECVQ) von Chou, Lookabaugh und Gray [51] auf den Entwurf von Verfahren zur Langzeitvorhersage angepaßt. In den entsprechenden Publikationen [5, 2] wurde die Anwendbarkeit des Verfahrens und die Wirkung des Trainings auf die Codiereffizienz nachgewiesen. Weiterhin konnte gezeigt werden, daß bei einer Bewegungsschätzung mit Raten-Nebenbedingung das Verfahren der Langzeitprädiktion bei gleicher oder erhöhter Bitrate für die Bewegungsparameter eine Verbesserung der Bewegungskompensation erreicht.

In einem nächsten Schritt wurde ein dem H.263-Standard entsprechender Coder um einen bewegungskompensierenden Langzeitprädiktor erweitert. Ein solcher Coder wird von uns in [2, 7, 6] vorgestellt. Die Bitratenkontrolle des Hybridcoders in [2, 7, 6] benutzt sowohl Bewegungsschätzung als auch Macroblock-Modeentscheidung mit Raten-Nebenbedingung. In [2, 7, 6] wird anhand von mehreren Testsequenzen gezeigt, daß bei der Verwendung von 50 Bildern im Langzeitspeicher bei erhöhter Bitrate für die Bewegungsparameter von etwa 5 % sich eine Reduktion der Bitrate für die Prädiktionsfehlercodierung von etwa 25 % ergibt, womit eine Bitratenreduktion für den Gesamtcodec von etwa 20 % im Vergleich zur herkömmlichen bewegungskompensierten Prädiktion erreicht

wird. Die Bitrateneinsparungen sind bei gleicher Qualität des rekonstruierten Videosignals gemessen worden. Bei einer bestimmten Testsequenz (*News*) wurde eine Bitrateneinsparung von 60 % gemessen [2]. Langzeitprädiktion funktioniert in diesem Fall so gut, weil sich bei dieser Testsequenz ein Teil des Bildinhaltes wiederholt.

Das Verfahren der bewegungskompensierten Langzeitprädiktion wurde als Erweiterung des H.263-Videocodierstandards vorgeschlagen [19, 21, 23, 24, 25, 26, 29, 30, 33, 34, 38, 39, 40, 41, 42, 43, 44]. Die gezeigte Leistungsfähigkeit des Verfahrens hat dazu geführt, daß der Vorschlag von der ITU-T/SG16/Q.15-Gruppe im Februar 1999 in Form eines Annexes angenommen wurde. Dieser neue Annex U wurde dann im Februar 2000 von der ITU-T/SG16 in einer Vorentscheidung angenommen. Die letzte Entscheidung über Annex U wird von der ITU-T im November 2000 getroffen werden, wenn die nationalen Standardierungsgreminen darüber entscheiden, ob das Verfahren der bewegungskompensierten Langzeitprädiktion als internationaler Videocodierstandard verabschiedet wird. Weiterhin beinhaltet der gegenwärtige Vorschlag für den H.26L-Standard der ITU-T das Langzeitprädiktionsverfahren als festen Bestandteil.

Die im Zusammenhang mit bewegungskompensierter Langzeitprädiktion besonders wichtige Bitratenkontrolle für die Bewegungsparameter ist durch die im Neuantrag beschriebenen Verfahren zur Bewegungsschätzung mit Raten-Nebenbedingung erzielt worden. Diese Verfahren haben aber nicht nur erhebliche Verbesserungen der Rate-Distortion-Effizienz des Langzeitprädiktionsverfahrens gebracht, sondern auch zu Gewinnen bei herkömmlichen H.263-Codern geführt. Durch Bewegungsschätzung und Macroblock-Modeentscheidung mit Raten-Nebenbedingung können Bitrateneinsparungen von 10 % erzielt werden im Vergleich zu einem H.263-Coder ohne Raten-Nebenbedingung bei Bewegungssuche und Macroblock-Modeentscheidung [2, 1]. Die Verbesserungen des H.263-Codieralgorithmus haben ihrerseits zu einem Vorschlag an die ITU-T/SG16/Q15-Gruppe geführt [20]. Daraufhin hat die ITU-T/SG16/Q15-Gruppe das in [20] vorgestellte Verfahren als Standard-Codiermethode für ihren Referenzcoder TMN-10 übernommen [56].

Die Gewinne durch Langzeitprädiktion von etwa 20 % addieren sich zu den 10 % Bitrateneinsparung die durch Rate-Distortion-Optimierung des H.263-Coders erzielt worden. Daher wurde eine Bitrateneinsparung von etwa 30 % durch Langzeitprädiktion mit Rate-Distortion-optimierter Bitratenkontrolle erreicht. Dieser Vergleich wurde gegenüber dem Referenzcodec ohne Rate-Distortion-Optimierung (TMN-9) [57] bei gleicher Qualität der rekonstruierten Videosignale angestellt. Der TMN-9-Coder war bis zur Festlegung des TMN-10-Coders als Referenzcoder die Standard-Codiermethode der ITU-T/SG16/Q15-Gruppe.

2.2.2 Entwicklung recheneffizienter Verfahren zur Langzeitprädiktion

Eine vollständige Suche zur Bewegungsschätzung ist für bewegungskompensierte Langzeitprädiktion besonders aufwendig, da nicht nur der Bewegungsvektor, sondern auch noch die Zeitverzögerung variiert werden muß. Deshalb sollten intelligente Suchverfahren entwickelt werden, die zu Reduktionen der zur Bewegungsschätzung notwendigen Rechenzeit führen.

Im Rahmen des Forschungsvorhabens konnte gezeigt werden, daß durch die Ausnutzung der Dreiecksungleichung zur Bestimmung von Schranken für den minimalen Abstand

zwischen zwei Vektoren erhebliche Verringerungen in der zur Bewegungssuche notwendigen Rechenzeit erzielt werden können [2, 13]. In [2, 13] wird gezeigt, wie eine Hierarchie von Dreiecksungleichungen zu einer Reduktion von bis zu 55 % der zur Bewegungsschätzung notwendigen Rechenzeit im Vergleich zu einer vollständigen Suche führen kann.

Weiterhin wurde ein Verfahren vorgestellt, das durch die Sortierung der Reihenfolge der Bewegungssuche anhand der zur Auswertung der Dreiecksungleichung vorberechneten Werte zusätzliche Geschwindigkeitsvorteile bietet [13]. Auch die Methode des "Half-Pixel-Refinement" wurde an die neue Suchreihenfolge angepaßt. Die in [13] vorgestellten Verfahren führen zu einer um den Faktor 5 verringerten Rechenzeit im Vergleich zu der für eine vollständige Suche notwendigen Rechenzeit.

Die neue Suchreihenfolge erlaubt weiterhin die Verwendung von verlustbehafteter Bewegungsschätzung. Hierbei lassen sich bei Verlusten des Prädiktionsgewinnes von weniger als 0.1 dB weitere Geschwindigkeitsvorteile erzielen, die zu einer Rechenzeit-Reduktion um den Faktor 10 führen [13]. Beispielsweise braucht die Bewegungssuche für sämtliche 16×16-Blöcke eines QCIF-Bildes der Sequenz *Foreman* bei einem Langzeitspeicher von 50 Bildern und einem örtlichen Suchbereich von ± 15 Bildpunkten in horizontaler und vertikaler Richtung im Mittel 2.1 Sekunden [13]. Dabei wurde ein Verlust des Prädiktionsgewinnes von $0,09dB$ im Vergleich zum Halb-Pixel-genauen Resultat der vollständigen Suche, die 23.7 Sekunden gedauert hat, gemessen. Die Messungen wurden mit einer Sun UltraSPARC durchgeführt, die mit einer mit 300 MHz getakteten CPU und 1 GByte Arbeitsspeicher ausgestattet ist. Das Verfahren in [13] ist auch bei der ITU-T/SG16/Q15-Gruppe im Zusammenhang mit dem Vorschlag von Langzeitprädiktion vorgestellt worden [22].

2.2.3 Affine Bewegungskompensation mit Referenzbild-Vorverzerrung

Im Forschungsvorhaben ist eine Erweiterung des translatorischen Bewegungsmodells auf ein affines Bewegungsmodell durchgeführt worden. Dazu wurde das Verfahren zur Bewegungskompensation mit Referenzbild-Vorverzerrung entwickelt [8, 14, 15]. Bei dem Verfahren zur Referenzbild-Vorverzerrung werden in einem ersten Schritt Bewegungscluster zwischen dem zu codierenden und dem zuletzt decodierten Bild geschätzt. Die geschätzten affinen Bewegungsmodelle werden jedoch auf das gesamte zuletzt decodierte Bild angewendet. Die so vorverzerrten Referenzbilder werden dann im nächsten Schritt zur blockbasierten, bewegungskompensierten Prädiktion herangezogen. Anschliessend wird mit Hilfe eines Lagrange-Ansatzes die Effizienz der affinen Parametersätze evaluiert.

Die Schätzung der affinen Bewegungsmodelle sieht zunächst eine translatorische Bewegungskompensation vor, die mit Hilfe eines Block-Matching-Verfahrens durchgeführt wird. Dann wird ein lineares Gleichungssystem für die affinen Bewegungsparameter gelöst, das mit Hilfe von geschätzten Bildgradienten aufgestellt wird. Die so erhaltenen affinen Bewegungsparameter werden mit dem translatorischen Bewegungvektoren kombiniert, um die Parameter für die Referenzbild-Vorverzerrung zu erhalten [14].

Weiterhin ist das Verfahren zur Bewegungskompensation mit Referenzbild-Vorverzerrung auf Langzeitprädiktion erweitert worden [15]. Die Erweiterung erfolgt dadurch, daß das gradientenbasierte Schätzverfahren auf jedes Referenzbild angewendet wird und dadurch ein optimaler Parametersatz pro Referenzbild gefunden wird. Dann wird unter

diesen affinen Parametersätzen der Optimale ausgesucht und zur Referenzbild-Vorverzerrung weiter verwendet. Somit ergibt sich für diesen Fall eine Erweiterung des Langzeitspeichers um die vorverzerrten Referenzbilder zur blockbarierten Bewegungskompensation. Dabei hat es sich sich gezeigt, daß die Prädiktionsgewinne aus beiden Verfahren, Langzeitprädiktion und Referenzbild-Vorverzerrung, sich fast vollständig ergänzen [15].

Zunächst wurde nur das zuletzt decodierte Bild zur Referenzbild-Vorverzerrung herangezogen [14]. Der Ansatz in [14] wurde in einen H.263-Coder eingebettet. Ein auf Lagrange-Optimierung basierender Algorithmus wurde zur Auswahl der affine Bewegungsparameter benutzt. Die dabei erzielten Gewinne waren in etwa der Größenordnung der Gewinne für Langzeitprädiktion. Für Sequenzen wie *Foreman* und *News* wurden etwa 20 % Bitrateneinsparungen gegenüber TMN-10 erzielt. Diese korrespondieren zu PSNR-Gewinnen zwischen 1-1.5 dB. Für Sequenzen mit komplizierter Bewegung und hochfrequenter Textur wie *Mobile & Calendar* und *Stefan* können sogar Bitrateneinsparungen bis zu 35 % erzielt werden, welche zu PSNR-Gewinnen von 2.2 dB korrespondieren [14].

Weiterhin konnten beide Verfahren, Referenzbild-Vorverzerrung und das Langzeitprädiktionsverfahren, in einem Hybridcoder gewinnbringend kombiniert werden [15]. Die dabei erzielten Ergebnisse stellen eine weitere Erhöhung der Bitrateneinsparungen dar. Diese Bitrateneinsparungen wurden für alle getesteten Bewegtbildsequenzen gemessen. Dabei wurden für die Sequenzen *Foreman* und *News* Bitrateneinsparungen zwischen 25 und 30 % erzielt. Für die Sequenz *Mobile & Calendar* konnte eine Bitrateneinsparung von mehr als 40 % gegenüber TMN-10 erzielt werden, welche zu einem PSNR-Gewinn von 3.5 dB korrespondiert.

Die Leistungsfähigkeit des Verfahrens zur Referenzbild-Vorverzerrung hat die Gruppe des Antragstellers dazu motiviert, das Konzept als Erweiterung von Langzeitprädiktion vorzuschlagen [27, 31, 32, 35]. Der Vorschlag wurde mit großem Interesse von der ITU-T/SG16-Q.15-Expertengruppe aufgenommen was sich in der Formulierung von Core-Experimenten unter Mitwirkung anderer Froschungsgruppen äußerte. Jedoch wurde von einer weiteren Verfolgung dieses Standardierungsvorschlages abgesehen, da der Ansatz eine zu große Abweichung vom existierenden H.263-Standard darstellt.

2.2.4 Multihypothesen-Langzeitprädiktion

Im Forschungsvorhaben wurde das Verfahren zur Multihypothesen-Langzeitprädiktion entwickelt [9, 10, 12, 3, 17, 16]. Multihypothesen-Langzeitprädiktion stellt eine Erweiterung der herkömmlichen Bewegungskompensation dar. Bei diesem Verfahren werden mehrere bewegungskompensierte Signale linear überlagert. Die überlagerten Signale werden als Hypothesen bezeichnet.

Eine ähnliche Methode wurde schon sehr erfolgreich im Rahmen der Standardisierung von MPEG-2 [53], MPEG-4 [54] und H.263+ [55] in der Form von bidirektionaler Prädiktion angewendet. Hierbei werden zwei Blöcke linear mit gleichem Gewicht überlagert, wobei der eine Block aus einem in der Bildsequenz zeitlich vorangegangenen Referenzbild und der andere aus einem zeitlich nachfolgenden Referenzbild stammen muß. Zur bidirektionalen Prädiktion müssen die beiden Referenzbilder rekonstruiert am Decoder vorliegen. Daher verursacht dieses Verfahren eine zusätzliche Verzögerung, die für interaktive Anwendungen der Bildkommunikation störend sein kann. Multihypothesen-

Langzeitprädiktion löst dieses Problem, da sämtliche Referenzbilder zu vorangegangenen Zeitpunkten korrespondieren und somit das aktuelle Bild immer aus der Vergangenheit vorhergesagt wird.

Multihypothesen-Langzeitprädiktion ist mit Hilfe des im Neuantrag beschriebenen iterativen Entwurfsverfahren entwickelt worden. Dabei wird, wie bei herkömmlicher Langzeitprädiktion, zur Bewegungsschätzung ein Lagrange-Funktional herangezogen, bei dem der Prädiktionsfehler zusammen mit der durch den Lagrange-Parameter gewichteten Bewegungsvektorrate minimiert wird. Dabei wurde das von Wu und Gersho [52] vorgeschlagene Suchverfahren für bidirektionale Prädiktion auf Multihypothesen-Langzeitprädiktion erweitert. Die Tabellen mit Codewörtern variabler Länge wurden mit Hilfe des Huffman-Algorithmus erzeugt und die skalaren Gewichte zur Überlagerung der Hypothesen wurden so trainiert, daß der mittlere quadratische Prädiktionsfehler mit Raten-Nebenbedingung minimiert wird. Dabei stellte sich heraus, daß ein Gewichtungsfaktor $1/N$ bei N Hypothesen eine gute Approximation der optimalen Gewichte darstellt [9].

Multihypothesen-Langzeitprädiktion wurde in einen Hybridcoder eingebettet [17, 16]. Dazu wurden die Ansätze zur Lagrange-Optimierung für Langzeitprädiktion auf den Multihypothesen-Ansatz erweitert. Neben dem iterativen Verfahren zur Bewegungsschätzung [52, 9] wurde die Macroblock-Modeentscheidung mit Lagrange-Ansatz [58, 2, 1] um einen Multihypothesen-Mode erweitert, wodurch eine effiziente Bitratenkontrolle erreicht wird. Experimente mit den Sequenzen *Foreman* und *Mobile & Calendar* zeigen, daß die Bitrateneinsparungen durch Multihypothesen-Langzeitprädiktion gegenüber Langzeitprädiktion weiter erhöht werden können. So läßt sich eine zusätzliche Einsparung von 10 bis 18 % für beide Sequenzen erreichen womit sich zum Beispiel für die Sequenz *Mobile & Calendar* eine Bitrateneinsparung von 32 % gegenüber TMN-10 ergibt. Diese Einsparung korrespondiert zu einem PSNR-Gewinn von 2.8 dB. Dabei ist es bewerkenswert, daß erst die Verwendung mehrerer Referenzbilder zu signifikanten Codiergewinnen führt [17].

2.2.5 Fehlerrobuste Hybridcodierung mit Langzeitprädiktion

Um Langzeitprädiktion auch bei gestörten Übertragungskanälen benutzen zu können, wurden Untersuchungen zur fehlerrobusten Langzeitprädiktion durchgeführt [4, 18]. Hierzu wurde ein Verfahren entwickelt, bei dem die Bitratenkontrolle einen Schätzwert der mittleren Distortion am Decoder verwendet. Die mittlere Distortion am Decoder ergibt sich einerseits durch den Codierfehler und andererseits durch den mittleren Übertragungsfehler.

Die Übertragungsfehler verursachen, daß Bildinhalt verlorengeht oder nicht decodiert werden kann. Die für diesen Bildinhalt vorgenommene Fehlerverschleierung führt dann zu unterschiedlichen Referenzbildern zwischen Coder und Decoder, welche widerum zur zeitlichen Fehlerfortpflanzung am Decoder führen. Der Umstand das die Übertragungsfehler zufällig auftreten bewirkt, daß das mittlere Übertragungsfehlersignal aus Fehlerverschleierung und zeitlichen Fehlerfortpflanzung verwendet werden muß. In den Publikationen wird eine Approximation dieses mittleren Übertragungsfehlersignals angegeben [4, 18].

Die Approximation des mittlere Übertragungsfehlersignals wurde in die auf Lagrange-Ansätzen basierende Bitratenkontrolle des fehlerrobusten Langzeitprädiktions-Coders einbezogen. Dazu wurde die Macroblock-Modeentscheidung und die Schätzung der Bewe-

gungsvektoren modifiziert. Die Veränderung der Macroblock-Modeentscheidung bewirkt, daß mehr Macroblöcke im INTRA-Mode codiert werden. Der INTRA-Mode stoppt die zeitliche Fehlerfortpflanzung da keine Bewegungskompensation durchgeführt wird. Der Einfluß der geschätzten mittleren Übertragungsfehlersignals auf die Schätzung der Bewegungsvektoren ist, daß die Auswahl der Referenzbilder modifiziert wird. Beispielsweise können dadurch verstärkt Macroblöcke, die im INTRA-Mode codiert sind, referenziert werden.

Für einen so erweiterten Langzeitprädiktionscoder konnte mit Hilfe von umfangreichen Experimenten gezeigt werden, daß das neue Verfahren auch erhebliche Gewinne im Fall von gestörten Kanälen aufweist. Die Experimente wurden mit und ohne Rückmeldung über fehlerhaft empfangenen Bildinhalt vom Decoder an den Coder durchgeführt [4, 18].

Literatur

[51] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 1, pp. 31–42, Jan. 1989.

[52] S.-W. Wu and A. Gersho, "Joint Estimation of Forward and Backward Motion Vectors for Interpolative Prediction of Video", *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 684–687, Sept. 1994.

[53] ISO/IEC 13818 (ITU-T H.262), "Generic Coding of Moving Pictures and Associated Audio Information, Part 2: Video", International Standard, Mar. 1994.

[54] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Video Verification Model", Draft, July 1997.

[55] ITU-T Recommendation H.263 Version 2 (H.263+), "Video Coding for Low Bitrate Communication", Jan. 1998.

[56] ITU-T/SG16/Q15-D-65, "Video Codec Test Model, Near Term, Version 10 (TMN-10), Draft 1", Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d65.doc, Apr. 1998.

[57] ITU-T/SG16/Q15-C-15, "Video Codec Test Model Number 9 (TMN-9)", Download via anonymous ftp to: standard.pictel.com/video-site/9712_Eib/q15c15.doc, Dec. 1997.

[58] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Video Coding and the Emerging H.263 Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 182–190, Apr. 1996.

3 Zusammenfassung

Die Ergebnisse des Forschungsvorhabens zur bewegungskompensierten Langzeitprädiktion haben gezeigt, daß mit den vorgeschlagenen Verfahren erhebliche Codiergewinne erzielt werden können. Es wurde anhand von mehreren Testsequenzen dokumentiert, daß bei der Verwendung von 50 Bildern im Langzeitspeicher eine Bitratenreduktion von etwa 20 % im Vergleich zur Hybridcodierung mit herkömmlicher bewegungskompensierter Prädiktion erreicht wird. Die Bitrateneinsparungen sind bei gleicher Qualität der rekonstruierten Videosignale gemessen worden. Das Verfahren zur bewegungskompensierten Langzeitprädiktion wurde erfolgreich zur Standardisierung bei der ITU-T/SG16/Q.15-Gruppe vorgeschlagen und wird voraussichtlich im November 2000 zum festen Bestandteil des internationalen Videocodier-Standards H.263. Das neue Projekt der ITU-T, H.26L, beinhaltet ebenfalls bewegungskompensierte Langzeitprädiktion.

Es wurde ein Verfahren zur affinen Bewegungskompensation entwickelt das auf einem neuen Ansatz zur Bewegungskompensation mit Referenzbild-Vorverzerrung basiert. Hierbei sind unter der Verwendung von 32 affinen Bewegungsclustern Bitrateneinsparungen zwischen 15 und 20 % im Vergleich zur Hybridcodierung mit herkömmlicher bewegungskompensierter Prädiktion erzielt worden. Weiterhin ist gezeigt worden, daß das Verfahren zur Referenzbild-Vorverzerrung mit Langzeitprädiktion effizient kombiniert werden kann. Hierbei wurden Bitrateneinsparungen zwischen 20 und 30 % gegenüber Hybridcodierung mit herkömmlicher bewegungskompensierter Prädiktion erzielt.

Im Forschungsvorhaben ist weiterhin das Verfahren zur Multihypothesen-Langzeitprädiktion entwickelt worden. Auch hier hat die Einbettung des neuen Verfahrens in einen Hybridcoder zu erheblichen Gewinnen in der Codiereffizienz geführt. Anhand mehrerer Testsequenzen wurde gezeigt, daß die Bitrateneinsparungen durch den Multihypothesenansatz unter Verwendung von Langzeitprädiktion mit 10 Bildern von 15-20% auf 20-30% gegenüber Hybridcodierung mit herkömmlicher bewegungskompensierter Prädiktion gesteigert werden können.

Durch die Ausnutzung von vorberechneten Werten und einer veränderten Suchreihenfolge wurden erhebliche Einsparungen in der zur Bewegungsschätzung für Langzeitprädiktion notwendigen Rechenzeit erzielt. Typischerweise läßt sich mit den vorgeschlagenen Verfahren bei Langzeitprädiktion mit 50 Bildern ein Faktor 10 an Rechenzeit einsparen.

Weiterhin wurde gezeigt, wie durch eine verbesserte Bitratenkontrolle das Verfahren der bewegungskompensierte Langzeitprädiktion auch zur Übertragung von Bewegtbildsequenzen über gestörte Kanäle benutzt werden kann. Dazu wird ein Schätzwert der mittleren Distortion am Decoder bei der Schätzung der Bewegungsvektoren und der Macroblock-Modeentscheidung berücksichtigt. Die mittlere Distortion am Decoder wird unter Berücksichtigung der Statistik des gestörten Übertragungskanals und der durchgeführten Bewegungskompensation geschätzt. Für einen so erweiterten Langzeitprädiktionscoder konnte experimentell gezeigt werden, daß das neue Verfahren auch erhebliche Gewinne im Fall von gestörten Übertragungskanälen aufweist. Die Experimente wurden mit und ohne Rückmeldung über fehlerhaft empfangenen Bildinhalt vom Decoder an den Coder durchgeführt.