**Zeitschrift:** Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafenbetriebe = Bulletin technique / Entreprise des postes, téléphones et télégraphes suisses = Bollettino tecnico / Azienda delle poste, dei telefoni e dei telegrafi svizzeri

**Herausgeber:** Schweizerische Post-, Telefon- und Telegrafenbetriebe

**Band:** 65 (1987)

**Heft:** 2

**Artikel:** Bildtelefon im digitalen Telekommunikationsnetz

**Autor:** May, Franz

**DOI:** https://doi.org/10.5169/seals-874796

**Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

**Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

**Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

**Download PDF:** 15.07.2025

**ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch**

# Bildtelefon im digitalen Telekommunikationsnetz

Franz MAY, Ulm[1]

Zusammenfassung. *Künftigen ISDN-Teilnehmern werden zwei Nutzkanäle mit jeweils 64 kbit/s Datenrate zur Verfügung stehen. Es eröffnen sich damit Möglichkeiten, parallel zur Sprachkommunikation Standbilder zu übertragen, aber auch echte Bewegtbildfolgen, wenn leistungsfähige Codierverfahren eingesetzt werden. Bei Verwendung von Bildcodieren mit Bewegungskompensation und bewegungsadaptiver Interpolation nicht codierter Teilbilder kann für typische Bildtelefonszenen eine ausreichend hohe Bildqualität erwartet werden.*

**Le visiophone dans le réseau téléphonique numérique**
Résumé. *Les futurs usagers du RNIS disposeront de deux canaux utiles véhiculant chacun un débit numérique de 64 kbit/s. Outre les communications vocales, ils peuvent servir à transmettre aussi bien des images inanimées que des séquences d'images animées par le recours à des procédés de codage performants. Si les codeurs d'image employés sont prévus pour la compensation et l'interpolation adaptative du mouvement d'images partielles non codées, on peut s'attendre à une qualité de présentation suffisante des scènes typiques en visiophonie.*

**Il videotelefono nella rete telefonica digitale**
Riassunto. *I futuri utenti ISDN avranno a disposizione due canali, ciascuno con una frequenza di cifra di 64 kbit/s. Sarà così possibile trasmettere, accanto alla voce, immagini fisse e, utilizzando un tipo di codifica ad alta capacità, anche sequenze di immagini in movimento. Per le tipiche scene videotelefoniche si potrà ottenere una qualità dell'immagine sufficientemente elevata impiegando la codificazione dell'immagine con compensazione del movimento e l'interpolazione adattativa del movimento delle immagini parziali non codificate.*

Die Deutsche Bundespost sieht ab 1988 die Einführung des diensteintegrierten digitalen Netzes ISDN (integrated services digital network) vor. Kennzeichen des ISDN ist die digitalisierte Übertragung von Teilnehmer zur Teilnehmer. Mit dem ISDN-Basisanschluss stehen dann jedem Teilnehmer jeweils zwei vermittelbare und transparente 64-kbit/s-Kanäle zur Verfügung. Zusätzlich zur Sprachkommunikation ermöglicht der zweite Kanal die simultane Übertragung von Daten, Grafiken und Dokumenten. Unter der Voraussetzung, dass eine leistungsfähige Codierung von Bewegtbildern mit 64 kbit/s eine akzeptable Qualität liefert, soll dann erstmals ein flächendeckender und kostengünstiger Individual-Bildtelefondienst möglich sein. Bedenkt man, dass digitales Fernsehen eine Datenrate von mehr als 200 Mbit/s benötigt, stellt ein Bildtelefon mit hoher Bildqualität bei bewegten Bildern und nur 64 kbit/s Übertragungsrate ein ehrgeiziges Ziel dar. Dabei können natürlich nicht alle Szeneninhalte eines Fernsehprogrammes originalgetreu wiedergegeben werden. Im Forschungsinstitut in Ulm hat AEG jedoch ein Verfahren entwickelt, das bei typischen Bildtelefonszenen, bei denen nur eine Person von einer festinstallierten Kamera aufgenommen wird, eine dem gewohnten Farbfernsehen nahekommende Qualität erzielt.

Bildtelefonieren bedeutet die Ergänzung der Sprachkommunikation durch gegenseitige Übertragung von Bewegtbildern. Das ISDN mit seinen beiden Kanälen erlaubt einen solchen Dienst zu verwirklichen, wobei das einfachstmögliche Endgerät gemäss *Figur 1* aufgebaut ist. Der zuschaltbare Bildkanal wird dabei über das Telefon gesteuert.

Die Deutsche Bundespost plant, bis 1990 den Bildtelefondienst zu standardisieren und die Kommunikationsprotokolle festzulegen. Es wird aber gerade in der Einführungsphase, die voraussichtlich im Bürobereich ihren Anfang nehmen wird, wichtig sein, nicht nur Bewegtbilder zu übertragen. Dokumentenübertragung mit hoher Auflösung, Datenverkehr, Integration von Telefax- und Teletexdienst werden die Attraktivität eines derartigen Dienstes wesentlich erhöhen. Da im Bürobereich zunehmend ISDN-fähige Personalcomputer und Arbeitsplatzrechner Verwendung finden, liegt es nahe, ein diensteintegrierendes Endgerät entsprechend *Figur 2* aufzubauen. Bei diesem Endgerät gruppieren sich die Komponenten, wie Dokumentenkamera und Papierausgabe, im wesentlichen um den kommunikationsfähigen Arbeitsplatzrechner. Die ISDN-Steckdose erlaubt, weitere Komponenten, z. B. Textfax, an die $S_0$-Schnittstelle anzuschliessen.

Die AEG arbeitet an der Entwicklung beider Typen von Endgeräten. Wesentliche Bedeutung kommt dabei der Bewegtbild- und der Standbildcodierung zu.

## Prinzipien der Bildcodierung

Fernsehsignale sind durch statistische Abhängigkeiten sowohl im Einzelbild (räumlich oder intraframe) als auch von Bild zu Bild (zeitlich oder interframe) gekennzeichnet. Die Abhängigkeiten innerhalb eines Bildes äussern sich beispielsweise dadurch, dass benachbarte Bildpunkte mit hoher Wahrscheinlichkeit gleiche Helligkeit und Farbe haben. Diese Redundanz lässt sich durch eine *Prädiktionscodierung*, z. B. *D*ifferentielle *P*uls-*C*ode-*M*odulation (DPCM), ausnutzen. Bei diesem Verfahren werden im wesentlichen nur die Abweichungen zwischen Bildpunkten codiert und damit die Datenrate gesenkt. Auch die Abhängigkeiten zwischen aufeinanderfolgen-

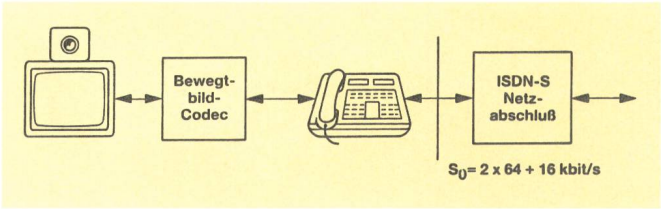


**Fig. 1**
**ISDN-Bildtelefon**

[1] Vortrag, gehalten am 21. Technischen Presse-Colloquium der AEG. Dr.-Ing. Franz May ist Wissenschaftler, Mitarbeiter in der Abteilung Informationstechnik im Forschungsinstitut, Ulm (BRD).

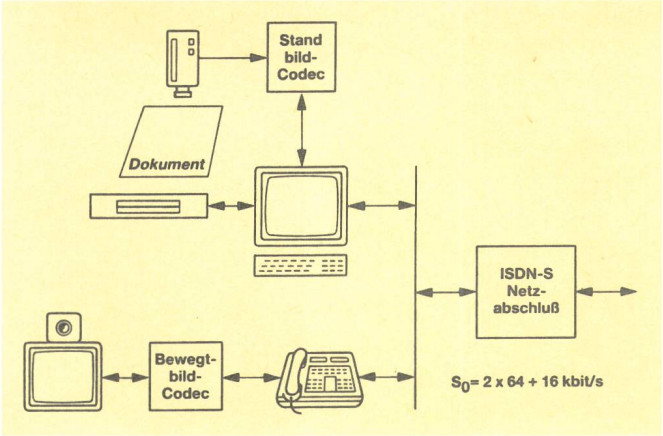


**Fig. 2**
**ISDN-Endgeräte im Büro mit Personalcomputer und Bildtelefon**

den Bildern in Bildfolgen lassen sich durch eine Bild-zu-Bild-Prädiktion verwenden. Dabei treten zu codierende Signale nur noch dort auf, wo sich der Bildinhalt geändert hat.

Statt der Prädiktion kann auch eine *Transformation* in einen Spektralbereich verwendet werden. Dazu wird ein Einzelbild in regelmässige Blöcke von beispielsweise 16 × 16 Bildpunkten zerlegt und deren zweidimensionales Spektrum berechnet. Die statistischen Abhängigkeiten äussern sich jetzt im Spektrum dadurch, dass Spektralkomponenten hoher Ordnung nur sehr geringe Amplituden erreichen, also nach Quantisierung mit wenigen Bits genügend genau darstellbar sind.

Beide Verfahren, die Prädiktions- und die Transformationscodierung, erlauben zusätzlich, Beobachtungsunschärfen des Auges zur weiteren Senkung der Datenrate auszunutzen (Irrelevanzreduktion). Sowohl die Prädiktionscodierung als auch die Transformationscodierung sind prinzipiell in der Lage, mit gleicher Effektivität zu arbeiten. Die Auswahl für das geeignetere Verfahren wird daher durch Kriterien wie erforderlicher Realisierungsaufwand, Anfälligkeit gegen Übertragungsfehler oder Art der typischen Fehler im empfangenen Bild bei stark beschränkter Kanalkapazität bestimmt.

Wird von den am Empfänger rekonstruierten Bildern verlangt, dass sie nicht vom Originalbild zu unterscheiden sind, können z. B. die 216 Mbit/s der CCIR-Empfehlung 601 auf 34 Mbit/s reduziert werden. Dies geschieht durch eine DPCM mit Intra- und Interframe-Prädiktion, adaptiver Quantisierung und Optimalcodierung. Ein derartiges Verfahren wird zurzeit in den zuständigen Gremien international genormt.

Für die typischen Szenen einer Videokonferenz oder eines Bildtelefondialoges lassen sich höhere Datenkompressionsfaktoren erreichen, wenn bei bestimmten, aber seltenen Bildfolgen Fehler zugelassen werden; das kann zum Beispiel ein nicht schlagartiger Bildaufbau bei einem Szenenwechsel sein oder eine Reduktion der Auflösung, wenn eine abgebildete Person sich sehr schnell bewegt.

In den derzeit installierten Videokonferenzstudios der Bundesrepublik Deutschland kommt ein Codec mit 2-Mbit/s-Übertragungsrate zum Einsatz. Dieser wurde innerhalb von COST (*Co*opération *S*cientifique et *T*echnique en Europe) entwickelt und vom CCITT (*C*omité *C*onsultatif *I*nternational *T*élégraphique et *T*éléphonique) genormt. Die Auflösung des Helligkeitssignals wurde gegenüber der Studionorm um etwa den Faktor 3, die Auflösung der Farbsignale um etwa den Faktor 6 reduziert. Mit einer Intra-/Interframe-DPCM, adaptiver Quantisierung und Optimalcodierung des Prädiktionsfehlersignals wird die Datenrate auf 2 Mbit/s reduziert. Diese Codierung hat zur Folge, dass die Bildqualität bei geringen Bewegungsvorgängen in der abgebildeten Szene recht gut ist, mit zunehmender Bewegung jedoch abnimmt und erst spürbar schlechter wird, wenn grosse Bereiche im Bild zu erneuern sind.

Zur Ausnutzung der Abhängigkeiten zwischen den Bildern (interframe) hat sich die Bild-zu-Bild-Prädiktion allgemein durchgesetzt. Sind hohe Kompressionsfaktoren gefordert, ist jedoch die Transformationscodierung zur Ausnutzung der Abhängigkeiten innerhalb eines Bildes (intraframe) deutlich besser geeignet. Der wesentliche Grund hierfür ist die Eigenschaft der Transformationscodierung, bei stark beschränkter Kanalkapazität «angenehmere» Verzerrungen zu verursachen als die Prädiktionscodierung. Es treten weniger hochfrequente Artefakte auf, und die rekonstruierten Bilder sind bei starker Bewegung im wesentlichen tiefpassgefilterte Bilder; diese stellen für den Gesichtssinn die angenehmere Verzerrung dar. Die Vorteile werden besonders deutlich bei einem Verbindungsaufbau oder einem Szenenwechsel, da hier keine Korrelation zu vorhergehenden Bildern bitsparend ausgenutzt werden kann. Hinzu kommt der sehr schnelle Schärfeaufbau der rekonstruierten Bilder im Laufe der Übertragung. Ein weiterer Vorteil der Transformationscodierung ist ihre einfache Adaptionsfähigkeit an unterschiedlichste Übertragungsraten.

## Bewegtbildcodierung mit Bewegungskompensation

Die Codierung von Bewegtbildfolgen für 64 kbit/s erfordert noch weitere Massnahmen, die über die zuvor genannten hinausgehen. Die wesentlichen sind die bewegungskompensierte Prädiktion und die bewegungsadaptive Interpolation ausgelassener Bilder. Auf dem Markt angebotene Codecs für 64 kbit/s der ersten Generation machen davon noch keinen Gebrauch. Das hat zur Folge, dass die Bilder zumeist unscharf sind und Bewegungen nur ruckhaft wiedergegeben werden. Die Einführung der angesprochenen Techniken ist aus der Sicht von AEG unbedingt erforderlich, um die Bildqualität des Bildtelefons auch bei Bewegung dem gewohnten Farbfernsehen anzunähern.

Bildfolgen enthalten auch bei Abbildung bewegter Objekte, bei Schwenk und Zoom der Kamera eine hohe Bild-zu-Bild-Redundanz, da der Inhalt des aktuellen Bildes dem Empfänger bereits weitgehend bekannt ist. Das Bild oder Teilbereiche davon sind jedoch im Vergleich zum vorhergehenden Bild verschoben, verkleinert, vergrössert oder verdreht. Bei Bewegung haben je zwei Bildpunkte aus aufeinanderfolgenden Bildern, welche die gleiche Stelle eines bewegten Objektes wiedergeben, nicht dieselben Koordinaten; sie sind entsprechend der Abbildung des bewegten Objektes gegeneinander

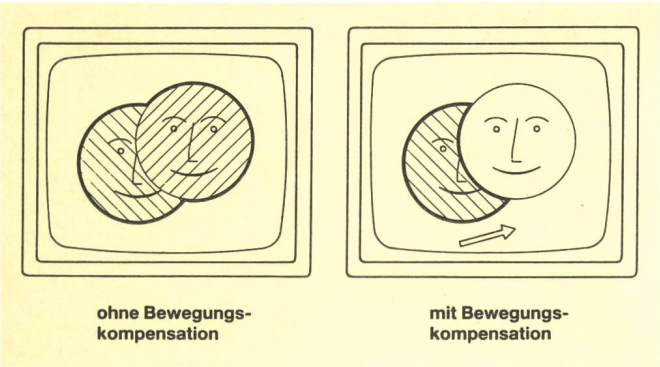


Fig. 3
**Bereiche mit Bild-zu-Bild-Prädiktionsfehlersignal**

versetzt. In diesen bewegten Bereichen tritt bei einfacher Bild-zu-Bild-Prädiktion ein hohes Fehlersignal auf, das bei beschränkter Kanalkapazität nicht vollständig übertragen werden kann.

Die Prädiktion kann nun verbessert werden, indem die Bewegungen der Objekte gemessen werden und zur Prädiktion ein Schätzbild berechnet wird. In diesem Bild werden die bewegten Objekte mit den zuvor gemessenen Verschiebungsgrössen an die «richtige» Stelle gerückt, so dass die Bewegung gerade kompensiert wird. In *Figur 3* hüllt die dicke Umrandung das jeweils zu übertragende Prädiktionsfehlersignal ein. Das Gesicht in seiner Lage im vorhergehenden Bild ist schraffiert dargestellt. Ohne Bewegungskompensation müssen der freigewordene Hintergrund und das ganze Gesicht neu übertragen werden, mit Bewegungskompensation nur noch der freigewordene Hintergrund. Die Wirklichkeit sieht natürlich etwas komplizierter aus als dargestellt; zu der zweidimensionalen Translation können auch Entfernungsänderungen, Rotationen und Gestaltänderungen kommen. Die Segmentierung eines Bildes in unbewegte und (gegebenenfalls unterschiedlich) bewegte Bildbereiche und die Schätzung der Verschiebegrössen dieser Bereiche sind die kennzeichnenden Probleme der bewegungskompensierten Prädiktion. Die Information, die die Bewegung im Bild beschreibt, wird dabei dem Empfänger mitgeteilt. Im wesentlichen ist dies eine angenäherte Beschreibung der Umrisslinie der bewegten Objekte und ihrer Bewegungsgrössen. Eine effektive bewegungskompensierte Prädiktion reicht noch nicht aus, bei einer Datenrate von 64 kbit/s gute Bildqualität zu garantieren. Bei 25 Vollbildern in der Sekunde stünden gerade 2560 Bit für ein codiertes Bild zur Verfügung. Das reicht kaum aus, die bei einem adaptiven Codierer mit vernünftig hoher Auflösung notwendige Steuerinformation zu übertragen. Diese besteht vor allem aus der Bewegungsinformation und der Einstellung der Quantisierer. Das Auslassen von Teil- und Vollbildern hilft aus diesem Dilemma, da dann die Bitmenge für jedes codierte Bild entsprechend vervielfacht wird, während die Steuerinformation nur unterproportional zunimmt. Um die Bewegungen der Szene nicht ruckhaft, sondern flüssig, d. h. «bewegungsrichtig» wiederzugeben, ist am Empfänger eine bewegungsadaptive Interpolation der ausgelassenen Teilbilder erforderlich.

Ein Bewegtbildcodierer der zweiten Generation, der alle angesprochenen Techniken benutzt, um auch bei 64 kbit/s hohe Bildqualität zu garantieren, wurde in Computersimulation entwickelt. Derzeit wird ein erstes in Echtzeit arbeitendes Funktionsmuster gebaut. Seine wesentlichen Schritte zur Senkung der Übertragungsrate bei möglichst kleinem Rekonstruktionsfehler sind:

- Zerlegung eines Bildes in unterschiedlich bewegte Bereiche und Berechnung der Bewegungsparameter dieser Bereiche unter Zugrundelegung eines Bewegungsmodells (object matching)
- Bewegungskompensierte Schätzbildberechnung aus dem letzten rekonstruierten Bild mit Hilfe der zu übertragenden Bewegungsparameter
- Codierung der verbleibenden Prädiktionsfehler in nicht prädizierbaren Bildbereichen durch adaptive Blockquantisierung (Cosinustransformation, Klassifizierung und Optimalcodierung)
- Bewegungsrichtige Interpolation ausgelassener Teil- und Vollbilder auf der Empfängerseite mit Hilfe der übertragenen Bewegungsparameter.

Die Struktur eines derartigen Codierers zeigt *Figur 4*.

Der Zählindex der Bildfolge ist m; n – 1 ist die Anzahl der ausgelassenen Bilder. Zunächst werden zwischen aufeinanderfolgenden zu codierenden Originalbildern $O_m$ und $O_{m-n}$ die bewegten Bereiche und deren Bewegungsparameter ermittelt (BA). Der Prädiktor P berechnet aus dem vorhergehenden rekonstruierten Bild $R_{m-n}$ das Schätzbild $S_m$ mit Hilfe der Bewegungsparameter. Das Prädiktionsfehlersignal $O_m$–$S_m$ wird transformiert (CT), quantisiert (Q) und für die im Zeitbereich arbeitende Rekonstruktionsschleife aus dem Spektralbereich rücktransformiert (ICT). Der Quantisierer Q führt auch die Auswahl der zu übertragenden Spektralkomponenten durch, die in C mit variabler Wortlänge optimal codiert werden. Die in Figur 4 als Doppellinien gezeichnete Steuerinformation, die dem Empfänger zusätzlich zu den ausgewählten quantisierten Spektralkomponenten zu übermitteln ist, besteht einerseits aus den Bewegungsdaten, die in codierter Form ein Verschiebungsfeld beschreiben, das für jeden Bildpunkt näherungsweise angibt, wo er im vorhergehenden Bild lag. Ander-

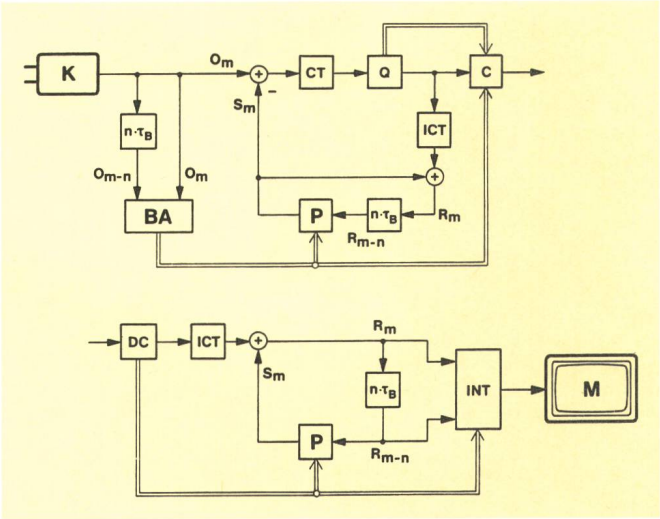


Fig. 4
**Struktur des Bildcodierers (oben) mit Bewegungskompensation, Blockquantisierung und Interpolation ausgelassener Bilder im Empfänger**

seits besteht die Steuerinformation aus der Einstellung der Quantisierer für die Spektralkomponenten und der durch eine Klassifizierung gewonnenen Adressinformation der ausgewählten Komponenten. Der Decoder trennt diese Informationsteile voneinander und berechnet dann genau wie der Coder das rekonstruierte Bild $R_m$. Zusätzlich führt der Interpolator (INT) die bewegungsadaptive Interpolation der nicht codierten Bilder mit den Indices $m-1...m-(n+1)$ durch. Dazu werden die für die Bewegungskompensation verwendeten Bewegungsgrössen benutzt. Dies ist nur möglich, wenn «richtige» Bewegungen gemessen werden. Die von AEG entwickelte Technik des «object matching» erreicht dies weitgehend. Nicht bildinhaltsbezogene und blockweise arbeitende Verfahren schätzen falsche Verschiebungen sehr leicht bei Blöcken, in denen die Gradienten der Helligkeitsfunktion nur schwach ausgeprägt sind oder ihre Richtung nicht ändern. Objektbildung und Ausnutzung von Nachbarschaftsbeziehungen zwischen bewegten Blöcken erlauben es, bei Mehrdeutigkeiten mit hoher Sicherheit die «richtige» Bewegung zu schätzen.

## Realisierung des Bildcodecs

Die AEG plant, 1988 mit Bildtelefon an den ISDN-Pilotversuchen der Deutschen Bundespost teilzunehmen. Die notwendigen Codecs werden zurzeit mit heute erhältlicher Technik aufgebaut. Die Realisierung basiert auf einem in einer höheren Sprache programmierbaren Mikroprozessor (MC 68020) und zwei sehr schnellen Signalprozessoren für die rechenaufwendigen Verarbeitungsschritte. Das sind vor allem die Bewegungsanalyse mit etwa 45 Millionen Operationen/s sowie die Hin- und Rücktransformation mit etwa 40 Millionen Operationen/s. In dieser Technik wird ein Codec zusammen mit den benötigten Bildspeichern (drei im Empfänger, vier im Sender) noch die Grösse eines Doppel-Europamagazins haben. AEG arbeitet ebenfalls am Entwurf von VLSI-Schaltkreisen für die rechenintensiven Verarbeitungsschritte. Aus heutiger Sicht kann davon ausgegangen werden, dass ein Codierer mit weniger als zehn VLSI-Schaltkreisen (einschliesslich Bildspeicher) aufgebaut werden kann, für den Decodierer genügt etwa die Hälfte. Die Bildcodierung wird dann kostengünstig und im Gehäuse des Endgerätes unterzubringen sein.

## Hauptmerkmale des Bildtelefon-Codecs für 64 kbit/s

Die Hauptmerkmale des bei der AEG entwickelten und derzeit in Realisierung befindlichen Bildtelefon-Codecs für 64 kbit/s sind:

- Reduktion der Bildwechselfrequenz auf 8⅓ Hz (es wird nur jedes sechste Teilbild in codierter Form übertragen)
- Objektbezogene Berechnung des Verschiebungsfeldes und der Schätzbildberechnung mit Bewegungskompensation
- Cosinustransformation und adaptive Quantisierung der verbleibenden Prädiktionsfehler
- Bewegungsrichtige Interpolation der ausgelassenen Bilder im Decoder mit Hilfe der übertragenen Bewegungsdaten.

Der Codec ermöglicht die Übertragung farbiger Bewegtbildfolgen mit Datenraten zwischen 64 kbit/s und 2 Mbit/s. Seine Hauptvorteile gegenüber konkurrierenden Codecs beruhen auf der objektbezogenen Arbeitsweise. Das ermöglicht eine sehr einfache «bewegungsrichtige» Interpolation ausgelassener Bilder, wodurch dem Benutzer ein nicht ruckhafter, sondern flüssiger Bewegungsablauf geboten wird.