Berichte aus der Kommunikations- und Informationstechnik

Band 20

Peter Eisert

Very Low Bit-Rate Video Coding Using 3-D Models

D 29 (Diss. Universität Erlangen-Nürnberg)

Shaker Verlag
Aachen 2001

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Eisert, Peter:
Very Low Bit-Rate Video Coding Using 3-D Models /
Peter Eisert. Aachen : Shaker, 2001
(Berichte aus der Kommunikations- und Informationstechnik ; Bd. 20)
Zugl.: Erlangen-Nürnberg, Univ., Diss., 2000
ISBN 3-8265-8308-6

Printed in Germany.

ISBN 3-8265-8308-6
ISSN 1432-489X

Shaker Verlag GmbH • P.O. BOX 1290 • D-52013 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • eMail: info@shaker.de

Video conferencing, tele-presence, and tele-teaching are video coding applications that have attracted considerable interest in recent years. The transmission of image sequences, however, imposes high demands on storage capacities and data rates of networks. For the uncompressed transmission of video in TV quality, e.g., bit-rates of more than 100 Mbps are necessary. Even when reducing the spatial and temporal resolution of the sequence significantly (e.g. QCIF, 10 Hz), the resulting bit-rates are still too large for transmission over today's low-cost channels. Common networks like the Public Switched Telephone Network (PSTN), the Integrated Services Digital Network (ISDN) and many computer networks provide only low bit-rates and cannot handle the amount of data arising when transmitting video uncompressed. Therefore, an efficient encoding of the image sequences is essential to enable the transmission of motion video.

In recent years, several video coding standards, such as H.261, MPEG-1, MPEG-2, and H.263, have been introduced to address the compression of digital video for storage and communication services. H.263 as well as the other standards describe a hybrid video coding scheme, which consists of block-based motion-compensated prediction (MCP) and DCT-based quantization of the prediction error. The recently determined MPEG-4 standard also follows the same video coding approach. These waveform-based schemes utilize the statistics of the video signal without knowledge of the semantic content of the frames and achieve compression ratios of about 100:1 at a reasonable quality.

If semantic information about a scene is suitably incorporated, higher coding efficiency can be achieved by employing more sophisticated source models. Model-based video codecs, for example, use 3-D models that describe the shape and texture of the objects in the scene. The 3-D object descriptions are typically encoded only once. During encoding of a video sequence, individual frames are characterized by a set of parameters describing 3-D motion and deformation of these objects. Since only a small number of parameters are transmitted, extremely low bit-rates are achievable. At the decoder, the video sequence is reconstructed by moving and deforming the 3-D models and rendering the 3-D scene using techniques known from computer graphics.

In this thesis, the analysis of image sequences is addressed by means of head-and-shoulder scenes frequently encountered in applications such as video-telephony or video-conferencing. Such sequences typically consist of a person moving in front of a static background and can therefore be described by a single 3-D head model that is animated to show facial expressions. A novel framework for model-based estimation of three-dimensional object motion and deformation from two-dimensional images is presented. Parameters describing facial expressions and head motion are estimated using optical flow information together with explicit motion constraints from the object model. The incorporation of photometric properties from the scene into the estimation framework leads to robust algorithms even under varying illumination. Experiments show that bit-rates of less than 1 kbps can be achieved when encoding head-and-shoulder scenes with the proposed algorithms. This enables the transmission of such video sequences over arbitrary networks and offers new opportunities for a wide variety of applications.