DECLARATION OF JUNE ANN MUNFORD

1. My name is June Ann Munford. I am over the age of 18, have personal knowledge of the facts set forth herein, and am competent to testify to the same.

2. I earned a Master of Library and Information Science (MLIS) from the University of Wisconsin-Milwaukee in 2009. I have over ten years of experience in the library/information science field. Beginning in 2004, I have served in various positions in the public library sector including Assistant Librarian, Youth Services Librarian and Library Director. I have attached my Curriculum Vitae as Appendix CV.

3. During my career in the library profession, I have been responsible for materials acquisition for multiple libraries. In that position, I have cataloged, purchased and processed incoming library works. That includes purchasing materials directly from vendors, recording publishing data from the material in question, creating detailed material records for library catalogs and physically preparing that material for circulation. In addition to my experience in acquisitions, I was also responsible for analyzing large collections of library materials, tailoring library records for optimal catalog

search performance and creating lending agreements between libraries during my time as a Library Director.

4. I am fully familiar with the catalog record creation process in the library sector. In preparing a material for public availability, a library catalog record describing that material would be created. These records are typically written in Machine Readable Catalog (herein referred to as "MARC") code and contain information such as a physical description of the material, metadata from the material's publisher, and date of library acquisition. In particular, the 008 field of the MARC record is reserved for denoting the date of creation of the library record itself. As this typically occurs during the process of preparing materials for public access, it is my experience that an item's MARC record indicates the date of an item's public availability.

5. Typically, in creating a MARC record, a librarian would gather various bits of metadata such as book title, publisher and subject headings among others and assign each value to a relevant numerical field. For example, a book's physical description is tracked in field 300 while title/attribution is tracked in field 245. The 008 field of the MARC record is reserved for denoting the creation of the library record itself. As this is the only date reflecting the inclusion of said materials within the library's collection, it is my experience

that an item's 008 field accurately indicates the date of an item's public availability.

6. I have reviewed Exhibit SAMSUNG-1007, "A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video" by Bo Martins and Soren Forchhammer as published in *IEEE Transactions on Circuits and Systems for Video Technology*, September 2002.

7. Attached hereto as Appendix MARTINS01 is a true and correct copy of the MARC record for *IEEE Transactions on Circuits and Systems for Video Technology* as held by the Carnegie Library of Pittsburgh. I secured this record myself from the library's public catalog. The MARC record contained within Appendix MARTINS01 accurately describes the title, author, publisher, and ISSN number of *IEEE Transactions on Circuits and Systems for Video Technology.* The 'Lib Has' field of this record visible on page 2 indicates this collection includes the September 2002 edition of *IEEE Transactions on Circuits and Systems for Video Technology* containing "A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video".

8. Attached hereto as Appendix MARTINS02 is a true and correct copy of "A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video" as published in *IEEE Transactions on Circuits and Systems for Video Technology* as held by the Carnegie Library of Pittsburgh. I secured these scans myself from the library's collection. In comparing Exhibit SAMSUNG-1007 to Appendix MARTINS02, it is my determination that Exhibit SAMSUNG-1007 is a true and correct copy of "A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video" by Bo Martins and Soren Forchhammer as published in *IEEE Transactions on Circuits and Systems for Video Technology* September 2002.

9. The 008 field of the MARC record in Appendix MARTINS01 indicates the date of record creation. The 008 field of Appendix MARTINS01 indicates the Carnegie Library of Pittsburgh first acquired this journal as of July 24, 1990 and held the journal in perpetuity until 2009. Considering this information, it is my determination that *IEEE Transactions on Circuits and Systems for Video Technology* September 2002 and therefore "A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video" was made available to the public shortly after its initial release in September 2002.

10. I have reviewed Exhibit SAMSUNG-1010 "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" by Heiko Schwarz, Detlev Marpe and Thomas Wiegand as published in *IEEE Transactions on Circuits and Systems for Video Technology*, September 2007.

11. Attached hereto as Appendix SCHWARZ01 is a true and correct copy of the MARC record for *IEEE Transactions on Circuits and Systems for Video Technology* as held by the Carnegie Library of Pittsburgh. I secured this record myself from the library's public catalog. The MARC record contained within Appendix SCHWARZ01 accurately describes the title, author, publisher, and ISSN number of *IEEE Transactions on Circuits and Systems for Video Technology.* The 'Lib Has' field of this record visible on page 2 indicates this collection includes the September 2007 edition of *IEEE Transactions on Circuits and Systems for Video Technology* containing "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard".

12. Attached hereto as Appendix SCHWARZ02 is a true and correct copy of "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" as published in *IEEE Transactions on Circuits and Systems for*

Video Technology as held by the Carnegie Library of Pittsburgh. I secured these scans myself from the library's collection. In comparing Exhibit SAMSUNG-1010 to Appendix SCHWARZ02, it is my determination that Exhibit SAMSUNG-1010 is a true and correct copy of "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" by Heiko Schwarz, Detlev Marpe and Thomas Wiegand as published in *IEEE Transactions on Circuits and Systems for Video Technology* September 2007.

13. The 008 field of the MARC record in Appendix SCHWARZ01 indicates the date of record creation. The 008 field of Appendix SCHWARZ01 indicates the Carnegie Library of Pittsburgh first acquired this journal as of July 24, 1990 and held the journal in perpetuity until 2009. Considering this information, it is my determination that *IEEE Transactions on Circuits and Systems for Video Technology* September 2007 and therefore "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" was made available to the public shortly after its initial release in September 2007.

14. I have reviewed Exhibit SAMSUNG-1006 "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension" by C. Andrew Segall and

Gary J. Sullivan as published in *IEEE Transactions on Circuits and Systems for Video Technology*, September 2007.

15. Attached hereto as Appendix SEGALL01 is a true and correct copy of the MARC record for *IEEE Transactions on Circuits and Systems for Video Technology* as held by the Carnegie Library of Pittsburgh. I secured this record myself from the library's public catalog. The MARC record contained within Appendix SEGALL01 accurately describes the title, author, publisher, and ISSN number of *IEEE Transactions on Circuits and Systems for Video Technology.* The 'Lib Has' field of this record visible on page 2 indicates this collection includes the September 2007 edition of *IEEE Transactions on Circuits and Systems for Video Technology* containing "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension".

16. Attached hereto as Appendix SEGALL02 is a true and correct copy of "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension" as published in *IEEE Transactions on Circuits and Systems for Video Technology* as held by the Carnegie Library of Pittsburgh. I secured these scans myself from the library's collection. In comparing Exhibit SAMSUNG-1006 to Appendix SEGALL02, it is my determination that

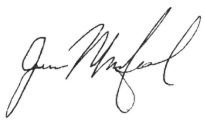
Exhibit SAMSUNG-1006 is a true and correct copy of "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension" by C. Andrew Segall and Gary J. Sullivan as published in *IEEE Transactions on Circuits and Systems for Video Technology* September 2007.

17. The 008 field of the MARC record in Appendix SEGALL01 indicates the date of record creation. The 008 field of Appendix SEGALL01 indicates the Carnegie Library of Pittsburgh first acquired this journal as of July 24, 1990 and held the journal in perpetuity until 2009. Considering this information, it is my determination that *IEEE Transactions on Circuits and Systems for Video Technology* September 2007 and therefore "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension" was made available to the public shortly after its initial release in September 2007.

18. I have been retained on behalf of the Petitioner to provide assistance in the above-illustrated matter in establishing the authenticity and public availability of the documents discussed in this declaration. I am being compensated for my services in this matter at the rate of $200.00 per hour plus reasonable expenses. My statements are objective, and my compensation does not depend on the outcome of this matter.

19. I declare under penalty of perjury that the foregoing is true and correct. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Dated: 1/7/2024



June Ann Munford

APPENDIX CV

June A. Munford
Curriculum Vitae

Education

University of Wisconsin-Milwaukee - MS, Library & Information Science, 2009
Milwaukee, WI

- Coursework included cataloging, metadata, data analysis, library systems, management strategies and collection development.
- Specialized in library advocacy, cataloging and public administration.

Grand Valley State University - BA, English Language & Literature, 2008
Allendale, MI

- Coursework included linguistics, documentation and literary analysis.
- Minor in political science with a focus in local-level economics and government.

Professional Experience

Researcher / Expert Witness, October 2017 – present
Freelance ● Pittsburgh, Pennsylvania & Grand Rapids, Michigan

- Material authentication and public accessibility determination. Declarations of authenticity and/or public accessibility provided upon research completion. Experienced with appeals and deposition process.

- Research provided on topics of public library operations, material publication history, digital database services and legacy web resources.

- Past clients include Alston & Bird, Arnold & Porter, Baker Botts, Fish & Richardson, Erise IP, Irell & Manella, O'Melveny & Myers, Perkins-Coie, Pillsbury Winthrop Shaw Pittman and Slayden Grubert Beard.

Library Director, February 2013 - March 2015
Dowagiac District Library ● Dowagiac, Michigan

- Executive administrator of the Dowagiac District Library. Located in

Southwest Michigan, this library has a service area of 13,000, an annual operating budget of over $400,000 and total assets of approximately $1,300,000.

● Developed careful budgeting guidelines to produce a 15% surplus during the 2013-2014 & 2014-2015 fiscal years while being audited.

● Using this budget surplus, oversaw significant library investments including the purchase of property for a future building site, demolition of existing buildings and building renovation projects on the current facility.

● Led the organization and digitization of the library's archival records.

● Served as the public representative for the library, developing business relationships with local school, museum and tribal government entities.

● Developed an objective-based analysis system for measuring library services - including a full collection analysis of the library's 50,000+ circulating items and their records.

November 2010 - January 2013
Librarian & Branch Manager, Anchorage Public Library ● Anchorage, Alaska

● Headed the 2013 Anchorage Reads community reading campaign including event planning, staging public performances and creating marketing materials for mass distribution.

● Co-led the social media department of the library's marketing team, drafting social media guidelines, creating original content and instituting long-term planning via content calendars.

● Developed business relationships with The Boys & Girls Club, Anchorage School District and the US Army to establish summer reading programs for children.

June 2004 - September 2005, September 2006 - October 2013
Library Assistant, Hart Area Public Library
Hart, MI

● Responsible for verifying imported MARC records and original MARC

cataloging for the local-level collection as well as the Michigan Electronic Library.

● Handled OCLC Worldcat interlibrary loan requests & fulfillment via ongoing communication with lending libraries.

Professional Involvement

Alaska Library Association - Anchorage Chapter
- Treasurer, 2012

Library Of Michigan
- Level VII Certification, 2008
- Level II Certification, 2013

Michigan Library Association Annual Conference 2014
- New Directors Conference Panel Member

Southwest Michigan Library Cooperative
- Represented the Dowagiac District Library, 2013-2015

Professional Development

Library Of Michigan Beginning Workshop, May 2008
Petoskey, MI
- Received training in cataloging, local history, collection management, children's literacy and reference service.

Public Library Association Intensive Library Management Training, October 2011
Nashville, TN
- Attended a five-day workshop focused on strategic planning, staff management, statistical analysis, collections and cataloging theory.

Alaska Library Association Annual Conference 2012 - Fairbanks, February 2012
Fairbanks, AK
- Attended seminars on EBSCO advanced search methods, budgeting, cataloging, database usage and marketing.

Depositions

2019 ● Fish & Richardson

Apple v. Qualcomm (IPR2018-001281, 39521-00421IP, IPR2018-01282 and 39521-00421IP2)

2019 ● Erise IP

Implicit, LLC v. Netscout Systems, Inc (Civil Action No. 2:18-cv-53-JRG)

2019 ● Perkins-Coie

Adobe Inc. v. RAH Color Technologies LLC (Cases IPR2019-00627, IPR2019-00628, IPR2019-00629 and IPR2019-00646)

2020 ● O'Melveny & Myers

Maxell, Ltd. v. Apple Inc. (Case No. 5:19-cv-00036-RWS)

2021 ● Pillsbury Winthrop Shaw Pittman LLP

Intel v. SRC (IPR2020-1449)

2022 ● Perkins-Coie

Realtek v. Future Link (IPR2021-01182)

2023 ● Fish & Richardson

Neuroderm Ltd. v. Abbvie, Inc (Case No. PGR2022-00040)

2023 ● Fish & Richardson

Nearmap US Inc. v. Pictometry International Corp. (IPR2022-00735)

Limited Case History & Potential Conflicts

Alston & Bird

- Ericsson
  - v. Collision Communications (IPR2022-01233)
- Nokia
  - v. Neptune Subsea, Xtera (Case No. 1:17-cv-01876)

Arnold & Porter

- Ivantis
  - v. Glaukos (Case No. 8:18-cv-00620)
- Samsung
  - v. Jawbone (Case No. 2:21-cv-00186)

Benesch Friedlander Coplan & Aronoff

- Voyis
  - v. Cathx (Case No. 5:21-cv-00077-RWS)

Erise I.P.

- Apple
  - v. Ericsson Inc. (IPR2022-00715)
  - v. Future Link Systems (IPRs 6317804, 6622108, 6807505, and 7917680)
  - v. INVT (Case No. 20-1881)
  - v. Navblazer LLC (IPR2020-01253)
  - v. Qualcomm (IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)
  - v. Quest Nettech Corp (Case No. 2:19-cv-00118-JRG)
  - v. Telefonaktiebolaget LM Ericsson (IPR2022-00275)

- Fanduel
    - v. CGT (Case No. 19-1393)
- Garmin
    - v. Phillips North America LLC (Case No. 2:19-cv-6301-AB-KS)
- Netscout
    - v. Longhorn HD LLC (Case No. 2:20-cv-00349)
    - v. Implicit, LLC (Civil Action No. 2:18-cv-53-JRG)
- Sony Interactive Entertainment LLC
    - v. Bot M8 LLC (IPR2020-01288)
    - v. Infernal Technology LLC (Case No. 2:19-CV-00248-JRG)
- Unified Patents
    - v. GE Video Compression (Civil Action No. 2:19-cv-248)

Fish & Richardson

- Apple
    - v. AliveCor (3:21-cv-03958)
    - v. LBS Innovations (Case No. 2:19-cv-00119-JRG-RSP)
    - v. Koss Corporation (IPR2021-00305)
    - v. Masimo (IPR 50095-0012IP1, 50095-0012IP2, 50095-0013IP1, 50095-0013IP2, 50095-0006IP1)
    - v. Neonode (Case No. 21-cv-08872-EMC)
    - v. Qualcomm (IPR2018-001281, 39521-00421IP, IPR2018-01282, 39521-00421IP2)
- Dell
    - v. Neo Wireless (IPR2022-00616)

- Dish Network
    - v. Realtime Adaptive Streaming (Case No. 1:17-CV-02097-RBJ)
    - v. TQ Delta LLC (Case No. 18-1798)
- Evapco Dry Cooling
    - v. SPG Dry Cooling (IPR2021-00688)
- Genetec
    - v. Sensormatic Electronics (Case No. 1:20-CV-00760)
- Huawei
    - v. Bell Northern Research LLC (IPR2019-01174)
- Kianxis
    - v. Blue Yonder (Case No. 3:20-cv-03636)
- LG Electronics
    - v. Bell Northern Research LLC (Case No. 3:18-cv-2864-CAB-BLM)
- Metaswitch
    - v. Sonus Networks (IPR2018-01719)
- Microsoft
    - v. Throughputer Inc (IPR2022-00757)
- MLC Intellectual Property
    - v. MicronTech (Case No. 3:14-cv-03657-SI)
- Nearmap Inc
    - v. EagleView Technologies (IPR2022-01009)
- Neuroderm Ltd.
    - v. Abbvie, Inc (Case No. PGR2022-00040)
- Realtek Semiconductor
    - v. Future Link (IPR2021-01182)

- Quectel
    - v. Koninklijke Philips (Case No. 1:20-cv-01710)
- Samsung
    - v. Aire Technology (IPR2022-00877)
    - v. Bell Northern Research (Case No. 2:19-cv-00286-JRG)
    - v. Communication Technologies Inc (IPR2022-01221)
    - v. Jawbone Innovations (IPR2022-00865)
    - v. MemoryWeb LLC (IPR2022-00885)
    - v. Telefonaktiebolaget LM Ericsson (IPR2021-00615)
- Texas Instruments
    - v. Vantage Micro LLC (IPR2020-01261)
- Xilinx
    - v. Sentient Sensors LCC (Case No. 1:22-cv-00173)

Irell & Manella

- Curium

O'Melveny & Myers

- Apple
    - v. Maxell (Case No. 5:19-cv-00036-RWS)

Perkins-Coie

- Heru Industries
    - v. The UAB Research Foundation (IPR2022-01148)
- Realtek Semiconductor
    - v. Future Link (IPR2021-01182)

- Twitter Inc

  v. VOIP-Pal.com (Case No. 3:20-cv-02397-JD)

- TCL Industries

  v. Koninklijke Philips NV (IPR2021-00495, IPR2021-00496 and IPR2021-00497)

Pillsbury Winthrop Shaw Pittman

- Intel

  v. FG SRC LLC (Case No. 6:20-cv-00315)

- Gravel Rating Systems

  v. Costco (Case No. 4:21-cv-149)
  v. Lowe's Home Centers (Case No. 4:21-cv-150)
  v. T-Mobile USA (Case No. 4:21-cv-152)
  v. Kohl's Inc. (Case No. 4:21-cv-258)
  v. Under Armor (Case No. 4:21-cv-356)

APPENDIX MARTINS01

screenshot-iiisy1.einetwork.net-2024.01.04-18_59_55
https://iiisy1.einetwork.net/search~S1?/tIEEE+Transactions+on+Circuits+and+Systems+for+Video+Technology+/tieee+transactions+on+circuits+and+systems+for+video+technology/1%2C1%2C1%2CB/marc&FF=tieee+transactions+on+circuits+and+systems+for+video+technology&1%2C1%2C/indexsort=-
04.01.2024


```
LEADER 00000cas  22000005a 4500
001    22134460
003    OCoLC
005    20150414135429.0
008    900724c19919999nyuqr p       0   a0eng d
010    91642188 |zsn 90001281
022 0  1051-8215
030    ITCTEM
032    006329|bUSPS
037    |bIEEE Service Center, 445 Hoes Lane, P.O. Box 1331,
       Piscataway, NJ 08855-1331
040    NSD|cNSD|dNYG|dMYG|dHUL|dNSD|dDLC|dNSD|dDLC|dCAS|dWAU|dOCL
       |dNSD|dELM|dNST|dNSD|dDLC|dWaOLN|dUtOrBLW
042    nsdp|alc
050 00 IN PROCESS
050 14 TK6630.A1|bI4
082 10 621|212
210 0  IEEE trans. circuits syst. video technol.
222  0 IEEE transactions on circuits and systems for video
       technology
245 00 IEEE transactions on circuits and systems for video
       technology :|ba publication of the Circuits and Systems
       Society.
246 10 Circuits and systems for video technology
246 23 Institute of Electrical and Electronics Engineers
       transactions on circuits and systems for video technology
260    New York, NY :|bInstitute of Electrical and Electronics
       Engineers,|c1991-
300    volumes :|billustrations ;|c28 cm
310    Quarterly
336    text|btxt|2rdacontent
337    unmediated|bn|2rdamedia
338    volume|bnc|2rdacarrier
362 0  Vol. 1, no. 1 (Mar. 1991)-
500    Title from cover.
510 1  Computer & control abstracts|x0036-8113|bMarch 1991-
510 1  Electrical & electronics abstracts|x0036-8105|bMarch 1991-
510 1  Physics abstracts|x0036-8091|bMarch 1991-
650  0 Television|xEquipment and supplies|vPeriodicals.
650  0 Television circuits|vPeriodicals.
650  0 Image processing|vPeriodicals.
650  0 Signal processing|vPeriodicals.
710 2  IEEE Circuits and Systems Society.
850    AAP|aAU|aArU|aAzTeS|aCCC|aCPT|aCSt|aCU-A|aCU-RivP|aCU-SB
       |aCoFS|aCoU|aDGW|aDLC|aGAT|aICarbS|aIaU|aInLP|aKyU|aMBU
       |aMCM|aMH-GM|aMdBJ|aMMet|aMiEM|aMiU|aMoKL|aMoU|aMsSM|aN
       |aNN|aNNC|aNRU|aNSyU|aNcD|aNcRS|aNjP|aNjR|aNmU|aOCU|aOU
       |aOrCS|aPPD|aPPiC|aPU|aRPB-S|aTU|aTxCM|aTxHR|aViBlbV|aWM
       |aWaU
890    IEEE transactions on circuits and systems for video
       technology.
901    |cSer
935    ADO-0676
936    Vol. 1, no. 2 (June 1991) LIC
```

Location	CLP Main (Oakland) Second Floor - Magazines
Lib Has	Volume 1 (1991) to Volume 19, no. 2 (2009).

screenshot-iiisy1.einetwork.net-2024.01.04-18_59_55
https://iiisy1.einetwork.net/search~S1?/tIEEE+Transactions+on+Circuits+and+Systems+for+Video+Technology+/tieee+transactions+on+circuits+and+systems+for+video+technology/1%2C1%2C1%2CB/marc&FF=tieee+transactions+on+circuits+and+systems+for+video+technology&1%2C1%2C/indexsort=-
04.01.2024

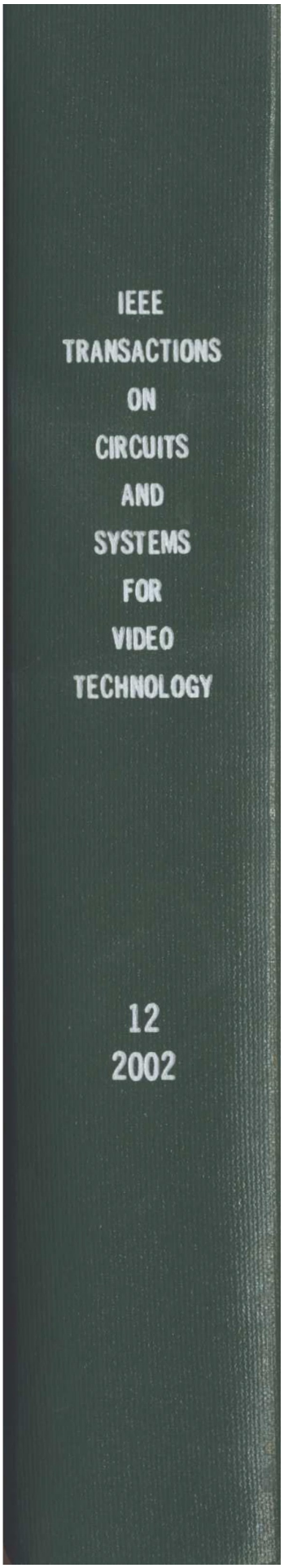
Title	**IEEE transactions on circuits and systems for video technology : a publication of the Circuits and Systems Society.**
Publisher	New York, NY : Institute of Electrical and Electronics Engineers, 1991-
Description	volumes : illustrations ; 28 cm
Frequency	Quarterly
Pub Date	Vol. 1, no. 1 (Mar. 1991)-

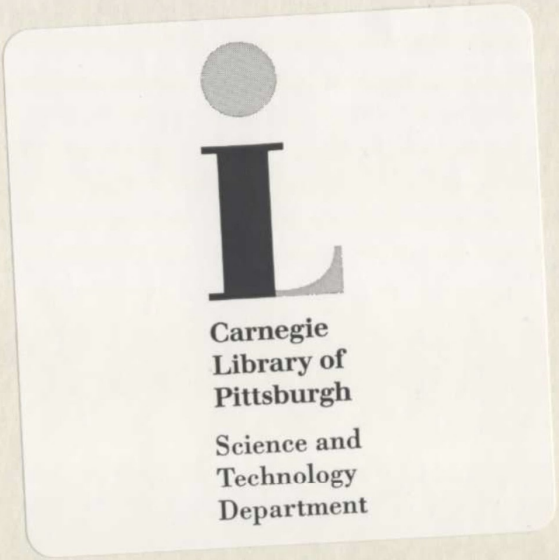
Location	CLP Main (Oakland) Second Floor - Magazines
Lib Has	Volume 1 (1991) to Volume 19, no. 2 (2009).

Details

Note	Title from cover.
Indexed In:	Computer & control abstracts 0036-8113 March 1991- Electrical & electronics abstracts 0036-8105 March 1991- Physics abstracts 0036-8091 March 1991-
Subject	Television -- Equipment and supplies -- Periodicals. *Show similar items* Television circuits -- Periodicals. Image processing -- Periodicals. Signal processing -- Periodicals.
Addt'l Author	IEEE Circuits and Systems Society.
Key Title	IEEE trans. circuits syst. video technol. IEEE transactions on circuits and systems for video technology
Addt'l Title	Institute of Electrical and Electronics Engineers transactions on circuits and systems for video technology
ISSN	1051-8215

APPENDIX MARTINS02


IEEE
TRANSACTIONS
ON
CIRCUITS
AND
SYSTEMS
FOR
VIDEO
TECHNOLOGY
12
2002


Carnegie
Library of
Pittsburgh

Science and
Technology
Department

IEEE TRANSACTIONS ON

CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY

A PUBLICATION OF THE IEEE CIRCUITS AND SYSTEMS SOCIETY

WWW.IEEE-CAS.ORG

SEPTEMBER 2002 VOLUME 12 NUMBER 9 ITCTEM (ISSN 1051-8215)

TRANSACTIONS PAPERS

VLSI Architecture Design of MPEG-4 Shape Coding. . . . . H.-C. Chang, Y.-C. Chang, Y.-C. Wang, W.-M. Chao, and L.-G. Chen 741
Lapped Orthogonal Transforms Designed for Error-Resilient Image Coding . . . . . D.-M. Chung and Y. Wang 752
Unsupervised Video-Shot Segmentation and Model-Free Anchorperson Detection for News Video Story Parsing . . . . . X. Gao and X. Tang 765
Supporting Image and Video Applications in a Multihop Radio Environment Using Path Diversity and Multiple Description Coding . . . . . N. Gogate, D.-M. Chung, S. S. Panwar , and Y. Wang 777
New Techniques for Multi-Resolution Motion Estimation. . . . . J. Zan, M. O. Ahmad, and M. N. S. Swamy 793

TRANSACTIONS LETTERS

A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video . . . . . B. Martins and S. Forchhammer 803
Memory-Constrained 3-D Wavelet Transform for Video Coding Without Boundary Effects . . . . . J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang 812
Morphological Representation of DCT Coefficients for Image Compression . . . . . D. Zhao, W. Gao, and Y. K. Chan 819

CALLS FOR PAPERS

Special Issue on Authentication, Copyright Protection and Information Hiding. 824
Special Issue on Audio and Video Analysis for Multimedia Interactive Services. 825
ISCAS 2003 826
WIAMIS 2003. 827

Transactions Letters

A Unified Approach to Restoration, Deinterlacing and Resolution Enhancement in Decoding MPEG-2 Video

Bo Martins and Søren Forchhammer

Abstract—**The quality and spatial resolution of video can be improved by combining multiple pictures to form a single superresolution picture. We address the special problems associated with pictures of variable but somehow parameterized quality such as MPEG-decoded video. Our algorithm provides a unified approach to restoration, chrominance upsampling, deinterlacing, and resolution enhancement. A decoded MPEG-2 sequence for interlaced standard definition television (SDTV) in 4:2:0 is converted to: 1) improved quality interlaced SDTV in 4:2:0; 2) interlaced SDTV in 4:4:4; 3) progressive SDTV in 4:4:4; 4) interlaced high-definition TV (HDTV) in 4:2:0; and 5) progressive HDTV in 4:2:0. These conversions also provide features as freeze frame and zoom. The algorithm is mainly targeted at bit rates of 4–8 Mb/s. The algorithm is based on motion-compensated spatial upsampling from multiple images and decimation to the desired format. The processing involves an estimated quality of individual pixels based on MPEG image type and local quantization value. The mean-squared error (MSE) is reduced, compared to the directly decoded sequence, and annoying ringing artifacts including mosquito noise are effectively suppressed. The superresolution pictures obtained by the algorithm are of much higher visual quality and have lower MSE than superresolution pictures obtained by simple spatial interpolation.**

Index Terms—**Deinterlacing, enhanced decoding, motion-compensated processing, MPEG-2, SDTV to HDTV conversion, video decoding.**

I. Introduction

MPEG-2 [1] is currently the most popular method for compressing digital video. It is used for storing video on digital versatile disks (DVDs) and it is used in the contribution and distribution of video for TV. We base this paper on the MPEG reference software encoder [2] for which a bit rate of 5–7 Mb/s yields a quality which is equivalent to (analog) distribution phase alternating line (PAL) TV quality. Lower bit rates are also used in TV distribution to save bandwidth and because professional encoders may provide better quality than the reference software.

Manuscript received December 1, 1999; revised May 2, 2002. This work was supported in part by The Danish National Centre for IT Research. This paper was recommended by Associate Editor A. Tabatabai.

B. Martins was with the Department of Telecommunication, Technical University of Denmark, DK-2800 Lyngby, Denmark. He is now with Scientific-Atlanta Denmark A/S, DK-2860 Søborg, Denmark (e-mail: bo.martins@sciatl.com).

S. Forchhammer is with Research Center COM, 371, Technical University of Denmark, DK-2800 Lyngby, Denmark (e-mail: sf@com.dtu.dk).

Publisher Item Identifier 10.1109/TCSVT.2002.803227.

At these bit rates, a sequence decoded from an MPEG-2 bitstream is of lower quality than the original digital sequence in terms of sharpness and color resolution but still acceptable (except for very demanding material). This overall reduction of quality is less annoying to a human observer than the artifacts typically found in compressed video. The most annoying artifacts are *ringing artifacts*[1] and in particular *mosquito noise,* which occurs when the appearance of the ringing changes from picture to picture.

The primary goal of this paper is to improve MPEG-2 decoding, or rather to postprocess the decoded sequence re-using information in the MPEG-2 bitstream to obtain a sequence of higher fidelity, especially with regard to the artifacts. The resulting output is a sequence in the same format as the directly decoded one, which in our case is interlaced standard TV in 4:2:0. In addition, we demonstrate how the approach can be used to obtain progressive (deinterlaced) or high-definition TV (HDTV) from the same bitstream. This also facilitates features such as frame freeze and zoom.

Previous work on postprocessing includes projections onto convex sets (POCS) [3] and regularization [4]. For low-bit-rate (high compression) JPEG-compressed still images and MPEG-1-coded moving pictures, the main artifact is *blocking,* i.e., visible discontinuities at coding block boundaries. This artifact can be dealt with efficiently using the POCS framework [5], as well as by other methods [6]. By regularization, POCS constraints can be combined with "soft" assumptions about the sequence. Thus, Choi *et al.* [4] restored very-low bit-rate video encoded by H.261 and H.263 according to the following desired (soft) properties: 1) smoothness across block boundaries; 2) small distance between the directly decoded sequence and the reconstructed sequence; and 3) smoothness along motion trajectories. Elad and Feuer [7] presented a unified methodology for superresolution restoration requiring explicit knowledge of parameters as warping and blurring. As this knowledge is not available in our case, we do not take the risk of processing based on estimating such parameters. Patti *et al.* [8] also addressed the superresolution problem in a general setting modeling the system components. They applied POCS performing projections for each pixel of each reference image in each iteration. Recently [9] this approach was modified to obtain superresolution from images of an MPEG-1 sequence captured by a specific video camera. Projections were carried

[1]Ringing artifacts are caused by the quantization error of high-frequency content, e.g., at edges. They appear as ringing adjacent to the edge.

1051-8215/02$17.00 © 2002 IEEE

out in the transform domain. Our goal is to develop simpler techniques (which could be combined with POCS).

The starting point of our work is the sequence decoded by an ordinary MPEG-2 decoder [2]. The material to be processed in this paper is of higher quality than MPEG-1 material or the low-bit-rate material of [4]. Consequently, there is a higher risk of degrading the material. Enforcing assumptions of smoothness of the material will almost surely lead to a decrease of sharpness. The basic idea of our restoration scheme is to apply a conservative form of filtering along motion trajectories utilizing the assumed quality of the pixels on each trajectory. The assumed quality of each pixel in the decoded sequence is given by the MPEG picture structure (i.e., what type of motion compensation is applied) and the quantization step size for the corresponding macroblock.

The algorithm has two steps. In the first step, a superresolution version (default is quadruple resolution) of each directly decoded picture[2] is constructed. In the second step, the superresolution picture is decimated to the desired format. Depending on the degree of decimation of the chrominance and luminance in the second step, the problem addressed is one of restoration, chrominance upsampling, deinterlacing, or resolution enhancement, e.g., conversion to HDTV. The aim in restoration is to enhance the decoding quality. For the other applications, the resolution is also enhanced.

In the first part of the upsampling, directly decoded pixels are placed very accurately in a superresolution picture before further processing. This approach is motivated by the fact that the individual pictures of the original sequence are undersampled [9], [10]. We do not want to trade resolution for improved peak signal-to-noise ratio (PSNR) by spatial filtering at this stage so the noise reducing filtering is deferred to the decimation step.

The paper is organized as follows. In Section II, a quality value is assigned to each pixel in the decoded sequence. Part one (upsampling) of our enhancement algorithm is described in Section III. The second part (decimation) is described in Section IV. Results on a number of test sequences are presented in Section V.

II. Processing Based on MPEG-Quality

MPEG-2 [1] partitions a picture into 16×16 blocks of picture material (*macroblocks*). A macroblock is usually predicted from one or more reference pictures. The different types of pictures are referred to as I, P, and B pictures. I pictures are intracoded, i.e., no temporal prediction. Macroblocks in P pictures may be unidirectionally predicted and macroblocks in B pictures may be uni- or bidirectionally predicted. (Macroblocks in B and P pictures may also be intracoded as macroblocks in I-pictures.)

The error block, resulting from the prediction, is partitioned into four luminance and two, four, or eight chrominance blocks of 8 × 8 pixels, depending on the format. For the 4 : 2 : 0 format, each macroblock has two chrominance blocks. The discrete cosine transform (DCT) is applied to each 8 × 8 block. The DCT coefficients are subjected to scalar quantization before being coded to form the bitstream.

[2]In this paper, all pictures are field pictures.

A. Quality Measure for Pixels in an MPEG Sequence

From the MPEG code stream, the type (I, P, or B) and the quantization step size are extracted for each macroblock. Based on this information, we shall estimate a quality parameter for each pixel which is used in a motion-compensated (MC) filtering. MPEG specifies the code-stream syntax but not the encoder itself. Our work is based on the reference MPEG-2 software encoder [2], for which the quantizers may be characterized as follows. The nonintra quantizer used for DCT coefficient $C(i, j)$ is (very close to) a uniform quantizer with quantization step q_s and a deadzone of $2q_s$ around zero. The intra quantizer used for DCT coefficient $C(i, j)$ has a deadzone of $5/4\ q_s(i, j)$ around zero. For larger values, it is a uniform quantizer with quantization step $q_s(i, j)$, and the dequantizer reconstruction point has a bias of $1/8\ q_s(i, j)$ toward zero. In [2], as is usually the case, all DCT coefficients in all blocks are being quantized independently as scalars.

The mean-squared error (MSE) caused by the quantization depends on the distribution of $C(i, j)$. This distribution varies with the image content and is hard to estimate accurately. We may approximate the expected error by the expression for a uniform distribution of errors, within each quantization interval, resulting from a uniform quantizer with quantization step q_s applied to $C(i, j)$

$$E\left\{\left(C(i,j) - C\left(\hat{i},\hat{j}\right)\right)^2\right\} = \frac{1}{12}q_s^2. \tag{1}$$

This expression may underestimate the error as it neglects the influence of the dead zone, and it may overestimate the error as the distribution of $C(i, j)$ is usually quite peaked around zero, especially for the high frequencies.

The DCT transform is unitary (when appropriate scaling is applied). Thus, the sum of squares over a block is the same in the DCT and spatial domains. Applying this to the quantization errors and introducing the expected values gives the following relationship for each DCT block:

$$\sum_{m=0,\,n=0}^{m=7,\,n=7} E\left\{\left(X(m,n) - \hat{X}(m,n)\right)^2\right\} = \sum_{i=0,\,j=0}^{i=7,\,j=7} E\left\{\left(C(i,j) - \hat{C}(i,j)\right)^2\right\} \tag{2}$$

where the DCT coefficients $C(i, j)$ are scaled as specified in [1] (Annex A) and $X(m, n)$ denotes the pixel value variables. As an approximation, we assume that the expected squared quantization errors are the same for all the pixel positions (m, n) within the DCT block. Based on this assumption, the expected value of the squared error for pixel $x(m, n)$ is given by

$$E\left\{\left(X(m,n) - \hat{X}(m,n)\right)^2\right\} = \frac{1}{64}\sum_{i=0,\,j=0}^{i=7,\,j=7} E\left\{\left(C(i,j) - \hat{C}(i,j)\right)^2\right\} \tag{3}$$

for all n, m within the DCT block having coefficients $C(i, j)$.

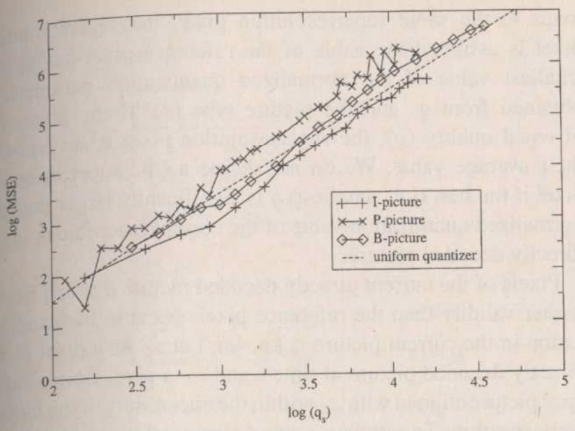


Fig. 1. MSE measured for sequence *table* as a function of the quantization step size q_s (depicted using natural logarithms). For intra pictures, q_s is defined as the quantization step size for the DCT coefficient at (1, 1).

Fig. 1 depicts the logarithm of the MSE as a function of $\log(q_s)$ for the luminance component of I, B, and P pictures. The figure reflects the fact that bidirectional prediction is better than unidirectional prediction, and that intra pictures and nonintra pictures are different. It is noted that we can use the expression (1) as a general approximation for the MSE of picture type y as long as we replace q_s with $q_s r_y$, where r_y is a constant which depends on the picture type, i.e.

$$E\left\{\left(X(m,\,n)-\hat{X}(m,\,n)\right)^2\right\}=\frac{1}{12}\,q^2,\qquad q\triangleq q_s r_y. \quad (4)$$

From the data in Fig. 1, we measure $r_B = 1$, $r_P = 1.2$, and $r_I = 0.9$. These values are used in all the experiments reported. The intra and nonintra quantization matrices used [2] are different. This is, in part, addressed by the values of r_y. [The value of r_I was measured with q_s defined as the quantization step size for $C(1,\,1)$.] The normalized quantization parameters, q in (4) are used as the quality value we assign to each pixel within the block. This measure is only used for relative comparisons and not as an absolute measure. It could be improved by taking the specific frequency content into account, as well as the precise quantization for each coefficient.

In general, pixels in the interior of an 8×8 DCT block have a smaller MSE than pixels on the border. We could assign a different value of r_y for interior pixels and pixels on the border. Experiments lead to our decision of ignoring the small difference at our (high) bit rates and as an approximation use the same quality value (4) for all pixels in a block.

III. Upsampling to Superresolution Using Motion Compensation

To process a given (directly decoded) picture we combine the information from the current frame and the N_f previous frames and the N_f subsequent frames, where N_f is a parameter and each frame consists of two field pictures. We first describe how to align pixels of the current picture at time t with pixels of one of the reference pictures using motion estimation. Section III-A then describes how to combine the information from all the reference pictures to form a single superresolution picture at t. The term *superresolution picture* is used to refer to the initial MC upsampled high-resolution image. An overview of the algorithm is given in Fig. 2.

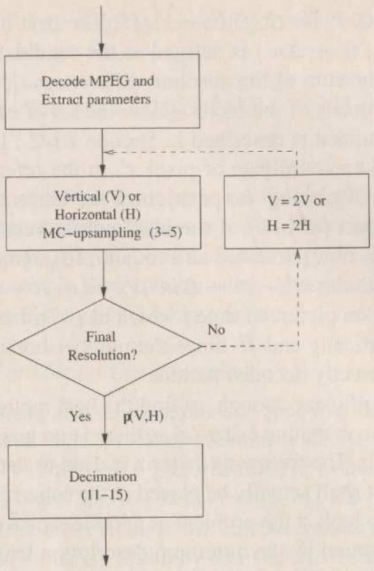


Fig. 2. Overview block diagram. MC upsampling alternates between doubling the resolution vertically and horizontally. Then final step is decimation to the desired format. Equation numbers are given in (). Dashed line marks control flow.

The motion field, relative to one of the reference pictures, is determined on the directly decoded sequence by block-based motion estimation using blocks of size 8×8. This block size is our compromise between larger blocks for robustness and smaller blocks for accuracy, e.g., at object boundaries. A motion vector is calculated at subpixel accuracy for each pixel x_0 of the current picture relative to the reference field picture considered. Based on the position of x_0 and the associated motion vector, one pixel x_r' shall be chosen in the reference picture.

The motion vector is found by searching the reference picture for the best match of the 8×8 block, which has x_0 positioned as the lower-right of the four center pixels. The displacements are denoted by $(m+\Delta m,\, n+\Delta n)$, where $(m,\, n)$ is the integer and $(\Delta m,\, \Delta n)$ the (positive) fractional part of the displacement relative to the position in the current picture. $m+\Delta m$ is the vertical displacement. For a given candidate vector $(m+\Delta m,\, n+\Delta n)$, each pixel x of the 8×8 block is matched against an estimated value $\hat{x}$ which is formed by bilinear interpolation of four neighboring pixels $x_r(m,\, n)$, $x_r(m+1,\, n)$, $x_r(m,\, n+1)$, and $x_r(m+1,\, n+1)$ in the reference picture

$$\hat{x}=\lfloor(1-\Delta m)(1-\Delta n)\,x_r(m,n)+(1-\Delta m)\Delta n\,x_r(m,n+1)$$
$$+\Delta m(1-\Delta n)\,x_r(m+1,n)+\Delta m\Delta n\,x_r(m+1,n+1)\rfloor \quad (5)$$

where $x_r(m,\, n)$ is the pixel in the reference picture displaced $(m,\, n)$ relative to the pixel x in the current picture. (The coordinate systems of the two pictures are aligned such that the positions of the pixels coincide with the lattice given by the integer coordinates.) The subpixel resolution of the motion field, specified vertically by V and horizontally by H, determines the allowed values of Δm and Δn: $\Delta m = 0,\, 1/V,\, \ldots,\, (V-1)/V$.

and $\Delta n = 0, 1/H, \ldots, (H-1)/H$. The best motion vector $(m' + \Delta m', n' + \Delta n')$ is defined as the candidate vector that minimizes the sum of the absolute differences ($|x - \hat{x}|$) taken over the 64 pixels of the block. (How the set of candidate vectors is determined is described in Section III-C.) Let (m'_r, n'_r) be the absolute coordinate of pixel x'_r in the reference picture obtained by displacing the position of the current pixel x_0 by the integer part (m', n') of the best motion vector. The pixel value of x'_r is now perceived as a (quantized) sample value of a pixel at position $((m'_r - m' - \Delta m')V, (n'_r - n' - \Delta n')H)$ in a superresolution picture at time t which has V times the number of pixels vertically and H times the number horizontally relative to the directly decoded picture.

It is not sufficient, though, to find the best motion vector according to the matching criterion as there is no guarantee this is a *good* match. The following criteria is used to decide for each x'_r whether it shall actually be placed in the superresolution picture. We may look at the problem as a lossless data compression problem (inspired by the minimum description length principle [11]). Let there be two alternative predictive descriptions of the pixels of the current 8×8 block, one utilizing a block of the reference picture and one which does not. If the best compression method that utilizes the reference block is better than the best method which does not, then we rely on the match. In practice, we do not know the best data compression scheme, but instead some of the best compression schemes in the literature may be used. For lossless still-image coding, we use JPEG-LS [12]. For lossless compression utilizing motion compensation, we chose the technique in [13], which may be characterized as JPEG-LS with motion compensation. For simplicity, the comparison is based on the sum of absolute differences. The JPEG-LS predictor [12] is given by

$$\hat{x} = \begin{cases} \min(a, b), & \text{if } c \geq \max(a, b) \\ \max(a, b), & \text{if } c \leq \min(a, b) \\ a + b - c, & \text{otherwise} \end{cases} \tag{6}$$

where a denotes the pixel to the left of x, b denotes the pixel on top of x, and c the top-left pixel.

We compare the (intra picture) JPEG-LS predictor and the best MC bilinear predictor (5). If the former yields a better prediction of the pixels of the 8×8 surrounding block, we leave the superresolution pixel undefined (or unchanged) by not inserting (or modifying) a MC pixel at the position $((m'_r - m' - \Delta m')V, (n'_r - n' - \Delta n')H)$.

Checking the match reduces the risk of errors in the motion compensation process, e.g., at occlusions. Occlusions are also handled by performing the motion compensation in both directions time wise, and by performing motion compensation at pixel level. This leads to a fairly robust handling of occlusions to within 3–4 pixels of the edge.

A. Forming the Superresolution Picture

The superresolution picture is initially formed by mapping pixels from each of the reference pictures as described above. The implemented block-based motion-compensation scheme is described in Section III-C. If more than one reference pixel maps to the same superresolution pixel, the superresolution pixel is assigned the value of the reference pixel having the smallest value of the normalized quantization parameter q obtained from q_s and the picture type (4). If the pixels are of equal quality (q), the superresolution pixel is set equal to their average value. We do not define a MC superresolution pixel if the best (i.e., smallest) q is significantly larger than the normalized quantization value of the current macroblock in the directly decoded picture.

Pixels of the current directly decoded picture *a priori* have a higher validity than the reference pixels because the exact location in the current picture is known. Let x_2 be a pixel of the directly decoded picture at time t and x_1 a pixel from a reference picture aligned with x_2 within the uncertainty of the motion compensation. To estimate a new (superresolution) pixel value x at the original sample position of x_2, we calculate a weighted value $\hat{x}$ of x_1 and x_2 by

$$\hat{x} = h_1 x_1 + h_2 x_2. \tag{7}$$

The filter coefficients in (7) may be estimated in a training session using original data. The (MSE) optimal linear filter is given by solving the Wiener–Hopf equations

$$\begin{pmatrix} E\{X_1X_1\} & E\{X_1X_2\} \\ E\{X_1X_2\} & E\{X_2X_2\} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} E\{XX_1\} \\ E\{XX_2\} \end{pmatrix} \tag{8}$$

where X, X_1, and X_2 are the stochastic variables of the pixels in (7). The variables X_1 and X_2 represent quantized pixel values, whereas X represents a (superresolution) pixel at a sample position in the picture with the original resolution. The Wiener filter coefficients could, alternatively, be computed under the constraint that $h_1 + h_2 = 1$ in order to preserve the mean value. In our experiments on actual data applying (8), $h_1 + h_2$ was fairly close to 1, so we just proceeded with these estimates. Given enough training data, the second-order mean values in (8) could be conditioned on the quality of (x_1, x_2), i.e., (q_1, q_2), and on the types of the pictures of x_1 and x_2 as well as other MPEG parameters. In this paper, the picture type is reflected by (4) and the number of free parameters is reduced by fitting a smooth function to the samples $h_1(q_2/q_1)$. We choose the function below as it is monotonically increasing in q_2/q_1 from 0 to 1 and as its behavior can be adjusted by just two parameters as follows:

$$h_1 = 1 - (1 - \alpha)^{(q_2/q_1)^\beta} \tag{9}$$

$$h_2 = 1 - h_1. \tag{10}$$

The parameter α specifies the *a priori* weight that x_1 should carry. The parameter β specifies how much the difference in the qualities of x_1 and x_2 should influence x. The filter (9) has the property that for $0 \leq \alpha \leq 1$, $\beta \geq 0$ and $q_1, q_2 \geq 0$, we have $0 \leq h_1 \leq 1$.

The MC superresolution pixels, which do not coincide with the sample positions in the current image, maintain the quality value they were assigned in the reference picture. Pixels in the original sample positions ($x_2 = x(m, n)$), determined by (7), are assigned the quality value

$$q(m, n) = h_1 q_1 + h_2 q_2. \tag{11}$$

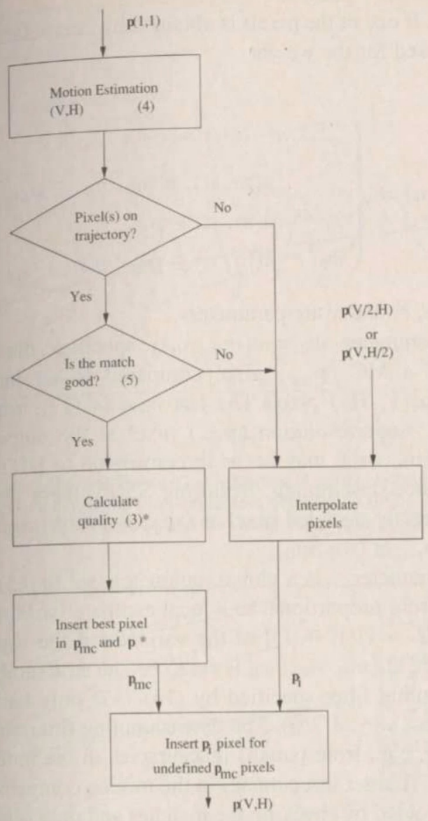


Fig. 3. Block diagram of MC upsampling doubling the vertical or horizontal resolution. Equation numbers are given in (). ∗Averages as expressed by (8)–(10) may also be used.

B. Completing the Superresolution Picture by Interpolation

A block diagram of the MC upsampling is given in Fig. 3. Let $\mathbf{p}_{\mathrm{mc}}(V, H)$ denote a superresolution picture created by MC upsampling as described previously. V and H specify the resolution of the motion compensation (5). Usually, some of the pixels of $\mathbf{p}_{\mathrm{mc}}(V, H)$ are undefined because there was no accurate match (of adequate quality) in any of the reference pictures. These pixels are assigned values from an interpolated superresolution picture $\mathbf{p}_i(V, H)$ having the same resolution as $\mathbf{p}_{\mathrm{mc}}(V, H)$. The resulting image is denoted $\mathbf{p}(V, H)$. The picture $\mathbf{p}_i(V, H)$ is created by a 2:1 spatial interpolation of the high-resolution picture $\mathbf{p}(V/2, H)$ (if $V = 2H$) or the high-resolution picture $\mathbf{p}(V, H/2)$ (if $V = H$). This upsampling alternates between horizontal and vertical 2:1 upsampling.

The upsampling process is first initialized by setting $\mathbf{p}(1, 1)$ equal to the directly decoded picture which has the original resolution. Thereafter, the initialization is completed by defining $\mathbf{p}(2, 1) = \mathbf{p}_i(2, 1)$, where $\mathbf{p}_i(2, 1)$ is created by spatial interpolation of $\mathbf{p}(1, 1)$. Hereafter, $\mathbf{p}(2, 2)$, $\mathbf{p}(4, 2)$, and $\mathbf{p}(4, 4)$[3] may be created in turn building up the resolution, alternating between horizontal and vertical 2:1 upsampling.

The odd samples being interpolated in the upsampled picture are obtained with a symmetric finite-impulse response (FIR) filter h_c used in the software coder [2] for 4:2:2–4:4:4 conversion

$$h_c = (\ldots, 0, 0.082, -0.203, 0.621, 0.621, -0.203, 0.082, 0, \ldots). \tag{12}$$

Each pixel of the resulting superresolution picture $\mathbf{p}(V, H)$ is assigned the attribute of whether it was determined by motion compensation $\mathbf{p}_{\mathrm{mc}}(V, H)$ or interpolation $\mathbf{p}_i(V, H)$. The MC pixels also maintain their quality value determined by (4) [and possibly modified by (11)] as an attribute.

C. Speedup of Motion Compensation

The following scheme is applied to speed up the estimation of the $4N_f + 1$ high-resolution motion fields that are required for the $4N_f + 1$ reference pictures relative to the current picture. The very first motion field (estimating the displacement of pixels of the other field of the current frame relative to the current field picture) is found by an exhaustive search within a small rectangular window (± 3 vertically and ± 7 horizontally). For each of the remaining $4N_f$ reference pictures, we initially *predict* the motion field before actually estimating the field by a search over a reduced set of candidate motion vectors. The motion field is initially predicted from the previously estimated motion fields using linear prediction, simply extrapolating the motion based on two motion vectors taken from two previous fields. (The offset in relative pixel positions between fields of different parity is taken into account in the extrapolation. After this the motion vectors implicitly takes care of the parity issue.) Having the predicted motion field (truncated to integer precision), we collect a list over the J most common motion vectors appearing in the predicted motion field. Thereafter, the search is restricted to the small set of this list for the integer part (m, n) of the motion vector in (5). All $H \cdot V$ fractional values of a motion vector are combined with the J integer vectors on the list. Consequently, the final motion vector search consists of trying out $J \cdot H \cdot V$ vectors. This way, we hope to track the motion vectors at picture level without requiring the tracking locally. Thus, even with a small initial search area, between the two fields of a frame, the magnitude of the motion vectors on the list may increase considerably with no explicit limit to the magnitude. Very fast motion, exceeding the initial search area between two fields of the same frame, is not captured though. In the experiments, we use a fixed-size ($J = 35$) candidate list. The size of the list can be adjusted according to different criterias. As an example, including all motion vectors on the list with an occurrence count greater than some threshold T in the predicted motion field reduces the risk of overlooking the motion vector of an object composed of more than T pixels, as a motion vector is estimated for each pixel. An additional increase in speed for higher-resolution motion fields ($H > 2 \vee V > 2$) is obtained by letting them be simple subpixel refinements of the motion field found for $H = V = 2$. The processing time for creating the high-resolution motion field is proportional to $J \cdot 4 + H \cdot V$ instead of $J \cdot H \cdot V$, i.e., approximately a reduction by a factor of four for the usual resolution ($H = V = 4$). As the size of the list with the updated vectors is fixed, the complexity is also proportional to the number of pictures specified by N_f.

[3]The block-based motion-estimation method applied does not warrant higher precision of the motion field.

In order to keep the algorithmic complexity down, we base the decisions in the enhancement algorithm on analysis of the luminance component only, always performing the same operations on a chrominance pixel as the corresponding luminance pixel. Additionally, no special action is taken at the picture boundaries apart from zero padding. The original motion vectors coming with the bit stream were disregarded as a higher resolution is desired. They could be used though, e.g., by including them on the list of predicted motion vectors.

IV. Decimation

The upsampling procedure only performed quality-based filtering for pixels located on the same motion trajectory (within our accuracy). In this section, we state a downsampling scheme applying quality based spatial filtering of the superresolution pictures. The filter coefficient for each pixel should reflect the quality and the spatial distance of the pixel. The quality attributes are dependent on the MPEG quantization (4) and whether the pixel is MC or interpolated. For all possible combinations of quality attributes within the filter window, the optimal filter could be determined given enough training data. Instead, we take the simpler approach of first assigning individual weights to each pixel depending on its attributes relative to the current pixel and then normalizing the filter coefficients.

A two-dimensional linear filter g is applied to the samples of the superresolution picture $\mathbf{p}$ $(=\mathbf{p}(V, H))$ in the vicinity of each sample position (m_0, n_0) in the resulting output image of lower resolution. The filter is a product of a symmetric vertical filter, a symmetric horizontal filter and a function reflecting the quality. The weight of the pixel $p(m, n)$ at (m, n) in $\mathbf{p}$ is

$$g(m, n; m_0, n_0, V, H) = K\, g_V(|m - m_0|)\, g_H(|n - n_0|)\, w(m, n). \quad (13)$$

In this expression, the weight $w(m, n)$ is a function of the quality attributes of the pixel $p(m, n)$ and K is a normalizing factor $(\sum g = 1)$. The 1-D filters g_V and g_H, reflecting the spatial distance, are defined as follows:

$$g_1 = (\ldots, 0, 0, 1, 0, 0 \cdots) \quad (14)$$

$$g_2 = (\ldots, 0, a, 1, a, 0 \cdots) \quad (15)$$

$$g_4 = g_2 * g_2 = (\ldots, a^2, 2a, 1 + 2a^2, 2a, a^2 \cdots). \quad (16)$$

It is noticed that the support of the low-pass filter is $(H + 1) \cdot (V + 1)$ superresolution pixels or approximately the area of one low-resolution pixel. This very small region of support is chosen to reduce the risk of blurring across edges in the decimation process. Furthermore, the value of a should be quite small because very often the individual pictures are undersampled. In the experiments, we use the parameter value $a = 0.1$.

The function $w(m, n)$ (13), reflecting the quality, depends on whether $p(m, n)$ and $p(m_0, n_0)$ are MC superresolution pixels or whether they were found through interpolation. When both pixels are MC [i.e., defined by $\mathbf{p}_{\mathrm{mc}} = \mathbf{p}_{\mathrm{mc}}(V, H)$], their relative quality parameters are used to determine the weight of $p(m, n)$. If one of the pixels is obtained by interpolation, a constant is used for the weight

$$w(m, n) = \begin{cases} \dfrac{w_0}{\gamma}\, \gamma^{(q(m,n)/q(m_0,n_0))^{\delta}} & \\ \qquad p(m, n),\ p(m_0, n_0) \in \mathbf{p}_{\mathrm{mc}} & \\ 1, \quad p(m, n) \notin \mathbf{p}_{\mathrm{mc}}, & \\ w_0, \quad p(m, n) \in \mathbf{p}_{\mathrm{mc}},\ p(m_0, n_0) \notin \mathbf{p}_{\mathrm{mc}} & \end{cases} \quad (17)$$

where w_0, δ, and γ are parameters.

The parameter $w_0 (= w(m_0, n_0))$ specifies the *a priori* worth of a MC $(\mathbf{p}_{\mathrm{mc}})$ pixel compared to an interpolated $(\mathbf{p}_i = \mathbf{p}_i(V, H))$ pixel. The last case in (17), where there is no MC superresolution $(\mathbf{p}_{\mathrm{mc}})$ pixel at the output sample position (m_0, n_0), may occur in conversion to HDTV and in chrominance upsampling. Restoring SDTV there will always be the directly decoded pixel at (m_0, n_0) ensuring a defined pixel in $\mathbf{p}_{\mathrm{mc}}$ at (m_0, n_0).

The parameter γ is a global parameter (set to 0.5) whereas δ is inversely proportional to a local estimate [within a region of size $(H + 1)(V + 1)$] of the variance of the superresolution picture at (m_0, n_0). w_0 is set to 6. The structurally simple downsampling filter specified by (13)–(17) only has the four parameters (a, γ, δ, w_0). The downsampling filter also attenuates noise, e.g., from (small) inaccuracies in the motion compensation. (Larger inaccuracies in the motion compensation are largely avoided by checking the matches and only operating on a reduced list of candidate motion vectors.)

V. Results

Four sequences were encoded: *table, mobcal, tambour-sdtv,* and *tambour-hdtv*. The extremely complex *tambour* sequence has been used both as interlaced SDTV and in HDTV format. For SDTV, the format is 4:2:0 PAL TV, i.e., the luminance frame size is 720 × 576 and the frame rate is 25 frames/sec. For HDTV, the resolution is doubled horizontally and vertically.

The parameters (α, β) of the filter expression (9) are estimated using a small number of frames of the sequence *mobcal*. Calculating the Wiener filter (8), we assume implicitly that the "original" pixels of the superresolution picture taken at the sample positions are equal to the original (low-resolution) pixels of the SDTV test sequence. This yields the curve $h(q_2/q_1)$ depicted in Fig. 4. Fitting the filter parameters of (9) to this curve yields $\alpha = 0.25$ and $\beta = 0.5$. These parameters are used in the processing of all the test sequences. Besides the curve $h_1(q_2/q_1)$ based on average values over all q_1, curves of $h_1(q_2/q_1)$ were recorded for different fixed values of q_1. These curves differ from the average in shape, as well as in level, e.g., expressed by the value for $q_1 = q_2$, i.e., $h_1(1)$. For most of the occurrences, q_2/q_1 was close to 1. The irregular shape of the curves for larger values of q_2/q_1 reflects the sparse statistics and due to this the dependency on the specific data that was used for estimating the Wiener filter. The overall level of h_1 was observed to increase with increasing q_1, reflecting the fact that the motion estimation inaccuracy becomes less important when the quantization error is large.

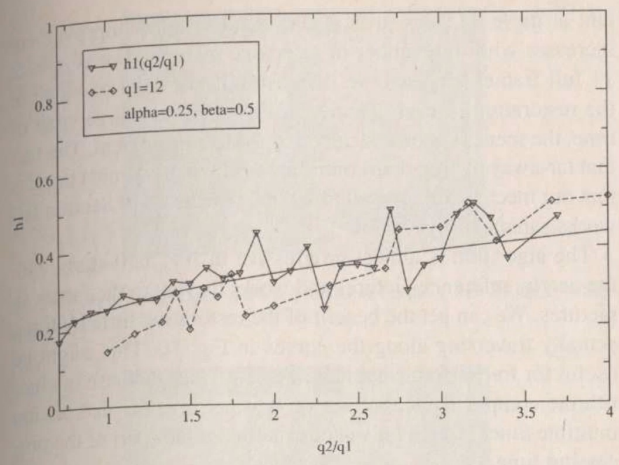


Fig. 4. Wiener filter coefficient h_1 as a function of q_2/q_1 for a piece of *mobcal*. The smooth function is the filter expression obtained by fitting α and β. The curves are $h_1(q_2/q_1)$ for all q_1 and for a small fixed value of $q_1(=12)$.

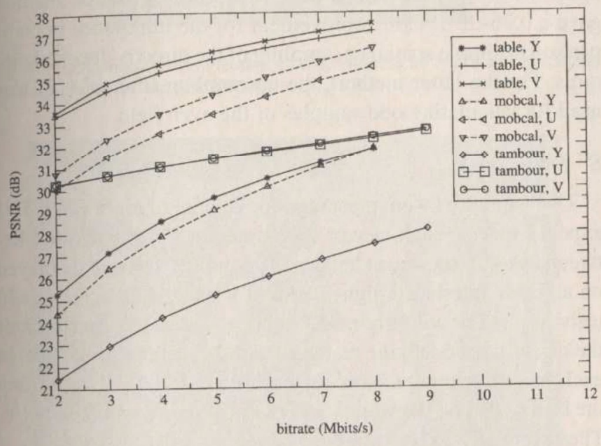


Fig. 5. PSNR of directly decoded sequences as a function of the bit rate.

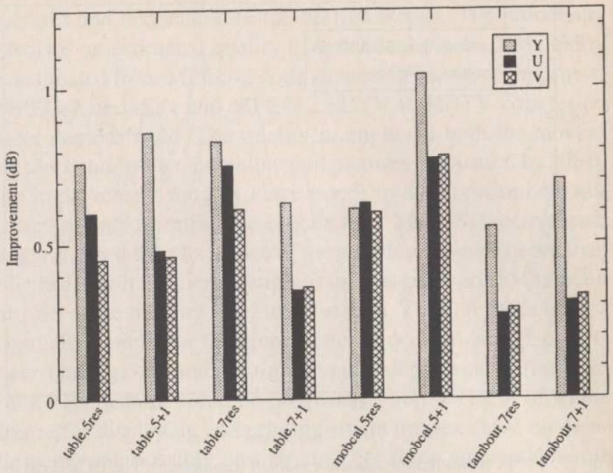


Fig. 6. Average improvement in PSNR for luminance and chrominance for all sequences (res) using parameters $H = V = 4$ and $N_f = 5$. The result of increasing the bit rate by 1 Mb/s (+1) is given for comparison.

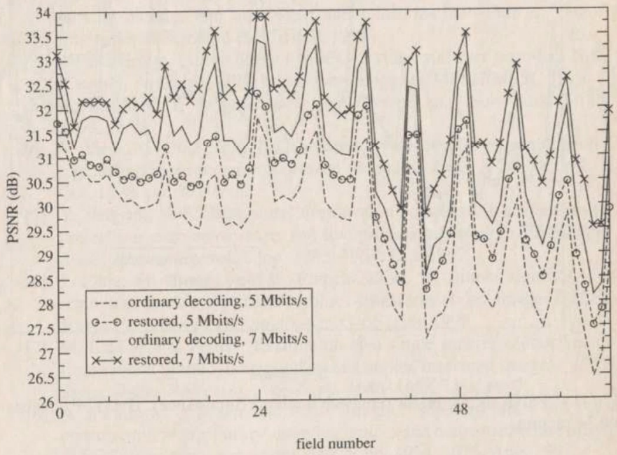


Fig. 7. PSNR measured for sequence *table* (luminance). The GOP consists of 24 pictures.

A. MSE Results

Fig. 5 shows the PSNR of the directly decoded sequences (for the first 33 frames in each sequence, which is the part being used in the tests). The average PSNR improvement for the sequences using our algorithm is depicted in Fig. 6. For comparison, the improvement obtained by increasing the coded bit rate by 1 Mb/s is also shown. Over these sequences, the average improvement achieved by our algorithm is roughly the same as the improvement obtained by increasing the bit rate by 1 Mb/s.

Figs. 7–9 show the PSNR for the individual pictures in the sequence. (The group of picture (GOP) structure consists of 12 frames and thereby 24 pictures: I/P, B/B, B/B, P/P, B/B, B/B, P/P $\cdots$). It is remarkable that the directly decoded sequences display such different characteristics: for *table* and *tambour*, the P pictures have much better PSNR than B pictures, while for *mobcal* this is not so. The restoration algorithm improves all pictures, regardless of their directly decoded quality. The magnitude of the improvement depends on two factors: 1) the relative quality of the directly decoded picture compared to the surrounding pictures and 2) to which degree the temporal redundancy was exploited during MPEG-2 coding. Consequently, the largest improvement (up to 1.7 dB) is recorded for the I pictures of *mobcal*. The P-pictures of *mobcal* being relatively poor and only unidirectionally predicted also display high improvement (about 1 dB). The B pictures of *table* being much worse than the corresponding P pictures display the highest improvement (about 1 dB) for this sequence. Whereas the algorithm generally improves poor pictures the most, some areas may be so poor (e.g., due to occlusions), that the algorithm fails to improve them. This is a consequence of the conservative strategy of requiring a good block match in the reference picture in order for it to influence the current picture. This is also the reason why *tambour* displays a relatively modest improvement and why *table* at 7 Mb/s has a larger improvement than *table* at 5 Mb/s.

In Fig. 10, the influence of the upsampling factors H and V, as well as N_f, is depicted. The superresolution picture is constructed using four $N_f + 2$ field pictures, namely the current field and the four $N_f + 1$ reference field pictures. The results are evaluated by the average improvement in PSNR reconstructing

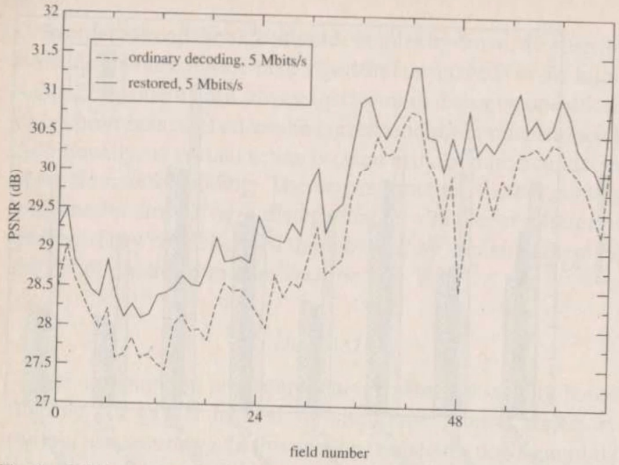


Fig. 8. PSNR measured for sequence *mobcal* (luminance). The GOP consists of 24 pictures.

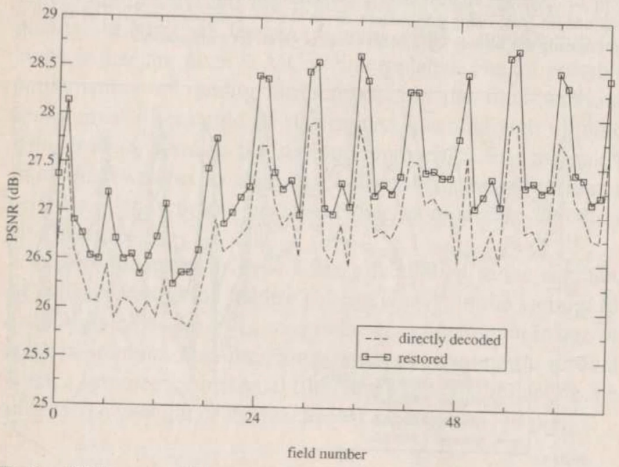


Fig. 9. PSNR measured for sequence *tambour* (luminance). The GOP consists of 24 pictures.

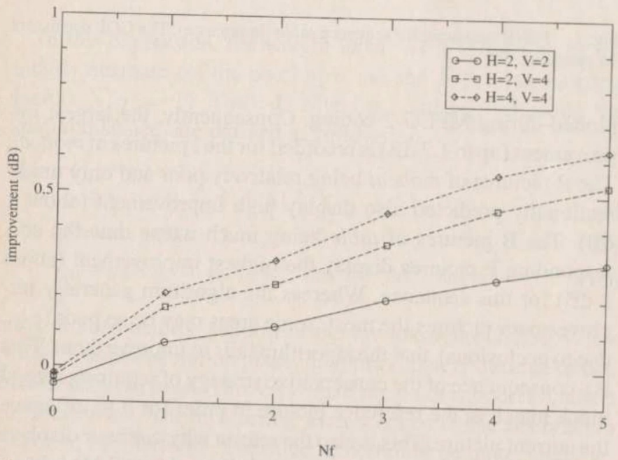


Fig. 10. Average PSNR improvement of the luminance for a GOP of *mobcal* (pictures 24–47) encoded at 5 Mb/s as a function of N_f and the upsampling factors (V, H).

SDTV measured over one GOP of *mobcal*. It is noted that, in this test, the improvement increases with the upsampling factor, implying that the accuracy of the motion field is very important at these bit rates. It is also noticed that the improvement increases with the number of reference pictures. For $N_f = 5$, 11 full frames are used, i.e., almost half a second of video in the restoration of each picture. In this relatively large span of time, the scene is geometrically warped to some extent. The fact that far-away pictures can contribute to the improvement implies that our mechanism for excluding bad matches (see Section III) works satisfactorily.

The algorithm is almost progressive in N_f, as it starts with the nearby reference pictures and works its way to the far-away pictures. We can get the benefit of the restoration little by little, actually traversing along the curves in Fig. 10. This might be useful for freeze-frame applications. The only increase in algorithmic complexity is that we have to perform the decimation multiple times, which only accounts for a minor part of the processing time.

For *tambour*, we can measure the performance of restoration to HDTV. The PSNR is only measured for the even fields. (This is in order to exclude the effect of a resampling in the measurements.) For *tambour* coded at 7 Mb/s, the restoration method gave a 0.76-dB PSNR improvement for the luminance in comparison to simple spatial upsampling of the directly decoded pictures. For the latter method, the upsampling filter of (12) was used for calculating odd samples of the even field.

B. Panel Tests

The sequences were presented for a panel of eight (PAL TV) expert viewers. Each viewer was seated at a fixed distance between two and six screen heights. The sequences were displayed on a 50-Hz interlaced high-fidelity TV using split screen in 20 tests in all. The viewers made blind pair-wise comparisons of the directly decoded, the restored, and the original sequence. In each pair-wise comparison, they scored $(-1, 0, +1)$ indicating the best $(+1)$ and the worst (-1) of the two or equal quality (0). They were also asked to judge sharpness, artifacts, etc.

The reconstructed sequences were overall rated as equally good or better than the corresponding directly decoded sequence (with an average overall score of 0.5 on the -1 to $+1$ scale). The overall evaluation was highly correlated with the degree the artifacts were evaluated to be reduced in the restored sequences. The sharpness was also evaluated to be improved by the restoration but less noticeable.

In a comparison between a directly decoded (*table*) sequence coded at 7 Mb/s and a restored sequence coded at 5 Mb/s, the panel judged the sequences to be of equal quality overall. Some viewers observed that the 7 Mb/s sequence was sharper.

Using our method for upsampling a decoded sequence to HDTV produced acceptable results for *mobcal* and *table*. The restored HDTV sequence of the very complex *tambour* was too bleak and lacked details though. For all sequences, our results were visually significantly better than simple spatial upsampling.

Deinterlacing was tested in a frame-freeze setting viewing single images of a progressive sequence. The images obtained by our enhancement algorithm were also evaluated as being of acceptable quality. Figs. 11–13 show part of an image of mobcal resulting from deinterlacing to progressive format. Fig. 11 depicts the result of using simple upsampling of a directly decoded



Fig. 11. Direct progressive SDTV. Mobcal at 5 Mbs/s. Part of the I-picture (frame 24, top field). PSNR = 28.5 dB.



Fig. 12. Enhanced to progressive SDTV. Mobcal at 5 Mb/s. Part of the I-picture (frame 24, top field). PSNR = 30.1 dB.



Fig. 13. Enhanced to progressive HDTV in 4 : 2 : 0. Mobcal at 5 Mb/s. Part of the I picture (frame 24, top field).

sequence. Figs. 12 and 13 depict the results of our enhancement to progressive SDTV and HDTV images, respectively.

VI. Conclusion

We have achieved a significant improvement of the decoding quality of MPEG-2 encoded sequences coded at bit rates that are usually considered to provide good quality for distribution. The algorithm is based on MC spatial upsampling from multiple pictures and decimation to the desired format. The processing involves an estimated quality of individual pixels. The quality is estimated from MPEG-2 code streams in our work. Improved MPEG-2 decoding and MPEG-2 SDTV to HDTV conversion were demonstrated. The quality is improved both for moving pictures and for the individual still pictures. Measured by MSE, the improvement roughly corresponds to the improvement obtained by incrementing the bit rate by 1 Mb/s. Subjective tests suggest that the performance of the algorithm is even better than this because it efficiently suppresses mosquito noise, the main artifact at the bit rates used in these tests. The algorithm is conceptually simple but the computational demand is high as it is based on high-accuracy estimation of a dense motion field. An initial application could be progressive improvement of frame freeze for displaying and printing single images based on deinterlacing and possibly upsampling. For these applications, the technique could be combined with POCS.

References

[1] ISO/IEC 13818-2, "Information technology—Generic coding of moving pictures and associated audio information—Part 2: Video," International Standard (MPEG-2), 1996.

[2] MPEG Group. (1996) MSSG MPEG-2 video software encoder, TM5. [Online]. Available: URL http://www.mpeg.org/MPEG/MSSG

[3] H. Stark and Y. Yang, *Vector Space Projections*. New York: Wiley, 1998.

[4] Y. Yang, M. Choi, and N. Galatsanos, "New results on multichannel regularized recovery of compressed video," in *Proc. ICIP '98*, vol. 1, Oct. 1998, pp. 391–395.

[5] Y. Yang and N. P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *IEEE Trans. Image Processing*, vol. 6, pp. 1345–1357, Oct. 1997.

[6] J. Chou, M. Crouse, and K. Ramchandran, "A simple algorithm for removing blocking artifacts in block-transform coded images," *IEEE Signal Processing Lett.*, vol. 5, pp. 33–35, Feb. 1998.

[7] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.

[8] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064–1076, Aug. 1997.

[9] A. J. Patti and Y. Altunbasak, "Super-resolution image estimation for transform coded video with application to MPEG," in *Proc. ICIP*, 1999, pp. 179–183.

[10] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[11] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Singapore: World Scientific, 1989.

[12] JPEG-LS, "IS 14495-1, lossless and near-lossless coding of continuous tone still images (JPEG-LS)," ISO/IEC International Standard, 1998.

[13] B. Martins and S. Forchhammer, "Lossless compression of video using motion compensation," in *Proc. 7th Danish Conf. Pattern Recognition and Image Analysis*, 1998, pp. 59–67.

INFORMATION FOR AUTHORS

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY covers all aspects of visual information technology, including but not limited to video A/D and D/A, display technology, image analysis and processing, video signal characterization and representation, image and video compression techniques, signal processing, multidimensional filters and transforms, analog video signal processing, neural networks for video applications, nonlinear video signal processing, video archiving and retrieval, computer vision, video transport, network video, streaming video, high-speed real-time circuits, synthetic video, object-based representation, MPEG technologies, VLSI architecture and implementation for video technology, multiprocessor systems—hardware and software, video systems architecture, video quality assessment, and error resilient video coding. The TRANSACTIONS consists of two sections.

1) *Transactions Papers*—This section is intended for self-contained articles that provide new and significant contributions to the field covered by the TRANSACTIONS, with coherent coverage of theoretical background, technical contributions, practical applications, and experimental results.
2) *Transactions Letters*—This section is intended for short articles that describe new and significant contributions to the field covered by the TRANSACTIONS, to be published in the most expeditious manner. This category includes: i) disclosure of potentially important ideas with convincing and encouraging results; ii) a condensed version of a submitted Transactions Paper; and iii) updates and correspondence on previously published papers.

Submission of a contribution is taken to manifest the fact that the submission has not been submitted, accepted, published, or copyrighted elsewhere. However, manuscripts submitted to or accepted by this Transactions may also be submitted for presentation at the IEEE International Symposium on Circuits and Systems or other conferences sponsored by the IEEE Circuits and Systems Society.

A. Submission of a Transactions Paper

1) Only electronic submissions of Transactions Papers via e-mail will be accepted. To submit a Transactions Paper, send an e-mail with subject "Transactions Paper" to the address: submission_tcsvt@hhi.de with the manuscript contained as an attachment strictly in PDF format, readable using Adobe Acrobat Reader. The body of the e-mail as well as the manuscript itself should contain the following information: Manuscript Type, Transactions Paper, Color illustrations: Yes/No, Author Name(s), Author Institution(s), Author E-mail(s), Paper Title, Abstract, Key Words, and Any Comments to the Editor-in-Chief.
2) Manuscript should state the significance of the problem in the Introduction.
3) The total length of a manuscript should not exceed ten (10) pages when printed using IEEE two-column format including figures (approximately 30 double-spaced typewritten pages of 8 1/2 × 11). The style files for LaTeX and Word can be downloaded from http://www.ieee.org/organizations/pubs/authors.html. The final submission format must be PDF.
4) Originals for the illustrations should be ready to be submitted immediately upon acceptance of the Transactions Paper.
5) See instructions on the web site for completing the IEEE copyright form for the manuscript.
6) If the manuscript has been presented, published, or submitted for publication elsewhere, please so inform the Editor-in-Chief in the comment form. Our primary objective is to publish technical material not available elsewhere, but on occasion we publish papers of unusual merit which have appeared or will appear before other audiences.

B. Submission of a Transactions Letter

1) Only electronic submissions of Transactions Letters via e-mail will be accepted. To submit a Transactions Letter, send an e-mail with subject "Transactions Letter" to the address: submission_tcsvt@hhi.de with the manuscript contained as an attachment strictly in PDF format, readable using Adobe Acrobat Reader. The body of the e-mail as well as the manuscript itself should contain the following information: Manuscript Type, Transactions Letter, Color illustrations: Yes/No, Author Name(s), Author Institution(s),Author E-mail(s), Paper Title, Abstract, Key Words, and Any Comments to the Editor-in-Chief.
2) Manuscript should state the significance of the problem in the Introduction.
3) The total length of a manuscript should not exceed five (5) pages when printed using IEEE two-column format including figures (approximately 15 double-spaced typewritten pages of 8 1/2 x 11). The style files for LaTeX and Words can be downloaded from http://www.ieee.org/organizations/pubs/authors.html. The final submission format must be PDF.
4) Originals for the illustrations should be ready to be submitted immediately upon acceptance of the Transactions Letter.
5) See instructions on the web site for completing the IEEE copyright form for the manuscript.
6) If the manuscript has been presented, published, or submitted for publication elsewhere, please so inform the Editor-in-Chief in the comment form. Our primary objective is to publish technical material not available elsewhere, but on occasion we publish papers of unusual merit which have appeared or will appear before other audiences.

RECEIVED
SEP 16 2002

C. Style for Manuscript

1) The manuscript should be typewritten using double space.
2) Provide an informative abstract of 100 to 250 words at the beginning of the manuscript.
3) Provide a separate double-spaced page listing all footnotes. Acknowledgment of financial support is often placed at the end of the footnote.
4) References should appear in a separate reference section at the end of the paper, with items referred to by numerals in square brackets, e.g., [12]. References should be complete and in the IEEE style. Style for papers: author, first initial followed by last name, title, volume, inclusive page numbers, month, and year. Style for books: author, title, location and publisher, year, chapter or page numbers (if desired). See any recent issues for examples.
5) Provide a separate page listing all figure captions, in proper style for the typesetter, e.g., "Fig. 6. The final wide-band amplifier design."
6) For further information, see the booklet Information for IEEE Transactions and Journal Authors, available from the IEEE Periodicals Department, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

D. Style for Illustrations

1) Originals for illustrations should be sharp, noise-free, and of good contrast. We regret that we cannot provide drafting or art service.
2) Line drawings should be in India ink on drafting cloth, paper, or board. Use only 8 1/2 × 11 size sheets to simplify handling of the manuscript.
3) On graphs, show only the coordinate axes, or at most the major grid lines, to avoid a dense hard-thread result.
4) All lettering should be large enough to permit legible reduction of the figure to column width, perhaps as much as 4:1. Typing on figures is not acceptable.
5) Photographs should be glossy prints, of good contrast and gradation, and any reasonable size.
6) Number each original on the back, or at the bottom of the front, and also indicate the author's name.
7) Note C-5 above. Captions lettered on figures will be blocked out in production, in favor of typeset captions.

E. Page and Color Photo Charges

After a manuscript has been accepted for publication, the author's institution will be approached with a request to pay a voluntary charge of $110 per page to cover part of the cost of publication. If the voluntary page charge is not honored, IEEE will charge a mandatory page charge of $175 per page for each page in excess of ten (10) for a TRANSACTIONS PAPER or in excess of five (5) for a TRANSACTIONS LETTER. The author will receive 100 free reprints (without cover) only if the voluntary page charge is honored. Detailed instructions will accompany the page proofs. Color photos judged to be essential to show the results of a paper may be printed in color. The cost of printing color photos will be shared by the author(s) and the Society.

F. Copyright

It is the policy of the IEEE to own the copyright to the technical contributions it publishes on behalf of the interests of the IEEE, its authors, and their employers and to facilitate the appropriate reuse of this material by others. To comply with the U.S. Copyright Law, authors are required to sign an IEEE Copyright Form before publication. This form returns to authors and their employers full rights to reuse their material for their own purposes. Authors must submit a signed copy of this form with their manuscript.

G. Special Issues

Special issues of this TRANSACTIONS may be proposed to the Editor-in-Chief by any individual when a new, unique, exciting, and/or timely subject area is identified.


Wert
Bookbinding
Grantville, PA
JULY-DEC 2003
"We're Quality Bound"

APPENDIX SCHWARZ01

screenshot-iiisy1.einetwork.net-2024.01.04-18_59_55
https://iiisy1.einetwork.net/search~S1?/tIEEE+Transactions+on+Circuits+and+Systems+for+Video+Technology+/tieee+transactions+on+circuits+and+systems+for+video+technology/1%2C1%2C1%2CB/marc&FF=tieee+transactions+on+circuits+and+systems+for+video+technology&1%2C1%2C/indexsort=-
04.01.2024


```
LEADER 00000cas  22000005a 4500
001    22134460
003    OCoLC
005    20150414135429.0
008    900724c19919999nyuqr p       0   a0eng d
010    91642188 |zsn 90001281
022 0  1051-8215
030    ITCTEM
032    006329|bUSPS
037    |bIEEE Service Center, 445 Hoes Lane, P.O. Box 1331,
       Piscataway, NJ 08855-1331
040    NSD|cNSD|dNYG|dMYG|dHUL|dNSD|dDLC|dNSD|dDLC|dCAS|dWAU|dOCL
       |dNSD|dELM|dNST|dNSD|dDLC|dWaOLN|dUtOrBLW
042    nsdp|alc
050 00 IN PROCESS
050 14 TK6630.A1|bI4
082 10 621|212
210 0  IEEE trans. circuits syst. video technol.
222  0 IEEE transactions on circuits and systems for video
       technology
245 00 IEEE transactions on circuits and systems for video
       technology :|ba publication of the Circuits and Systems
       Society.
246 10 Circuits and systems for video technology
246 23 Institute of Electrical and Electronics Engineers
       transactions on circuits and systems for video technology
260    New York, NY :|bInstitute of Electrical and Electronics
       Engineers,|c1991-
300    volumes :|billustrations ;|c28 cm
310    Quarterly
336    text|btxt|2rdacontent
337    unmediated|bn|2rdamedia
338    volume|bnc|2rdacarrier
362 0  Vol. 1, no. 1 (Mar. 1991)-
500    Title from cover.
510 1  Computer & control abstracts|x0036-8113|bMarch 1991-
510 1  Electrical & electronics abstracts|x0036-8105|bMarch 1991-
510 1  Physics abstracts|x0036-8091|bMarch 1991-
650  0 Television|xEquipment and supplies|vPeriodicals.
650  0 Television circuits|vPeriodicals.
650  0 Image processing|vPeriodicals.
650  0 Signal processing|vPeriodicals.
710 2  IEEE Circuits and Systems Society.
850    AAP|aAU|aArU|aAzTeS|aCCC|aCPT|aCSt|aCU-A|aCU-RivP|aCU-SB
       |aCoFS|aCoU|aDGW|aDLC|aGAT|aICarbS|aIaU|aInLP|aKyU|aMBU
       |aMCM|aMH-GM|aMdBJ|aMMet|aMiEM|aMiU|aMoKL|aMoU|aMsSM|aN
       |aNN|aNNC|aNRU|aNSyU|aNcD|aNcRS|aNjP|aNjR|aNmU|aOCU|aOU
       |aOrCS|aPPD|aPPiC|aPU|aRPB-S|aTU|aTxCM|aTxHR|aViBlbV|aWM
       |aWaU
890    IEEE transactions on circuits and systems for video
       technology.
901    |cSer
935    ADO-0676
936    Vol. 1, no. 2 (June 1991) LIC
```

Location	CLP Main (Oakland) Second Floor - Magazines
Lib Has	Volume 1 (1991) to Volume 19, no. 2 (2009).

screenshot-iiisy1.einetwork.net-2024.01.04-18_59_55
https://iiisy1.einetwork.net/search~S1?/tIEEE+Transactions+on+Circuits+and+Systems+for+Video+Technology+/tieee+transactions+on+circuits+and+systems+for+video+technology/1%2C1%2C1%2CB/marc&FF=tieee+transactions+on+circuits+and+systems+for+video+technology&1%2C1%2C/indexsort=-
04.01.2024

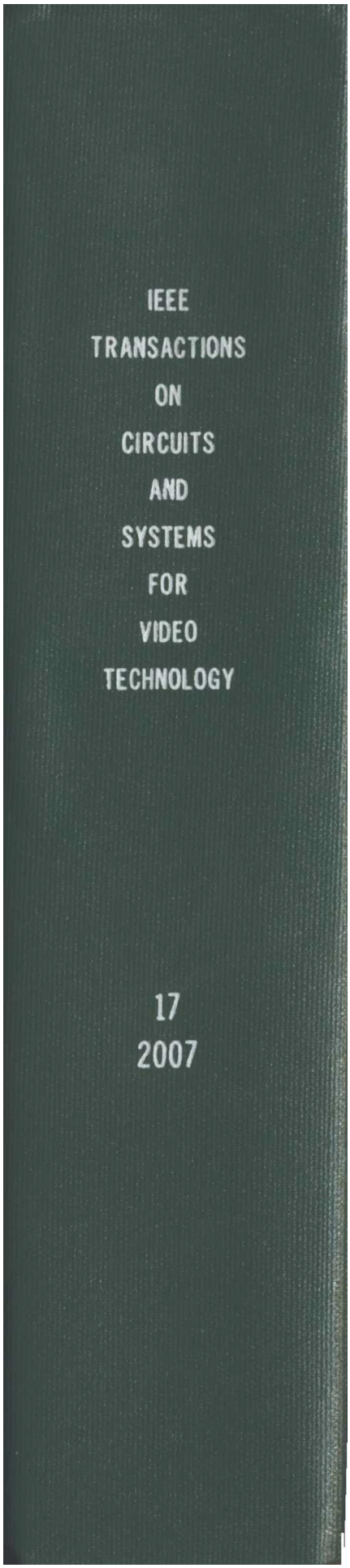
Title	**IEEE transactions on circuits and systems for video technology : a publication of the Circuits and Systems Society.**
Publisher	New York, NY : Institute of Electrical and Electronics Engineers, 1991-
Description	volumes : illustrations ; 28 cm
Frequency	Quarterly
Pub Date	Vol. 1, no. 1 (Mar. 1991)-

Location	CLP Main (Oakland) Second Floor - Magazines
Lib Has	Volume 1 (1991) to Volume 19, no. 2 (2009).

Details

Note	Title from cover.
Indexed In:	Computer & control abstracts 0036-8113 March 1991- Electrical & electronics abstracts 0036-8105 March 1991- Physics abstracts 0036-8091 March 1991-
Subject	Television -- Equipment and supplies -- Periodicals. *Show similar items* Television circuits -- Periodicals. Image processing -- Periodicals. Signal processing -- Periodicals.
Addt'l Author	IEEE Circuits and Systems Society.
Key Title	IEEE trans. circuits syst. video technol. IEEE transactions on circuits and systems for video technology
Addt'l Title	Institute of Electrical and Electronics Engineers transactions on circuits and systems for video technology
ISSN	1051-8215

APPENDIX SCHWARZ02


IEEE
TRANSACTIONS
ON
CIRCUITS
AND
SYSTEMS
FOR
VIDEO
TECHNOLOGY
17
2007

Carnegie
Library of
Pittsburgh
Main

IEEE TRANSACTIONS ON

CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY

A PUBLICATION OF THE IEEE CIRCUITS AND SYSTEMS SOCIETY

WWW.IEEE-CAS.ORG

SEPTEMBER 2007 VOLUME 17 NUMBER 9 ITCTEM (ISSN 1051-8215)

SPECIAL ISSUE ON SCALABLE VIDEO CODING—STANDARDIZATION AND BEYOND

EDITORIALS

Message from the Editor-in-Chief ..... *C.-W. Chen* 1097
Introduction to the Special Issue on Scalable Video Coding—Standardization and Beyond ..... *T. Wiegand, G. J. Sullivan, J.-R. Ohm, A. K. Luthra* 1099

INVITED PAPERS

Overview of the Scalable Video Coding Extension of the H.264/AVC Standard ..... *H. Schwarz, D. Marpe, and T. Wiegand* 1103
Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension ..... *C. A. Segall and G. J. Sullivan* 1121
Interlaced Coding in SVC ..... *E. François, J. Viéron, and V. Bottreau* 1136
System and Transport Interface of SVC ..... *Y.-K. Wang, M. M. Hannuksela, S. Pateux, A. Eleftheriadis, and S. Wenger* 1149
Transport and Signaling of SVC in IP Networks ..... *S. Wenger, Y.-K. Wang, and T. Schierl* 1164
File Format for Scalable Video Coding ..... *P. Amon, T. Rathgen, and D. Singer* 1174
Optimized Rate-Distortion Extraction With Quality Layers in the Scalable Extension of H.264/AVC ..... *I. Amonou, N. Cammas, S. Kervadec, and S. Pateux* 1186
Performance Analysis of SVC ..... *M. Wien, H. Schwarz, and T. Oelbaum* 1194
Mobile Video Transmission Using Scalable Video Coding ..... *T. Schierl, T. Stockhammer, and T. Wiegand* 1204
Scalable H.264/AVC Video Transmission Over MIMO Wireless Systems With Adaptive Channel Selection Based on Partial Channel Information ..... *D. Song and C. W. Chen* 1218
Real-Time System for Adaptive Video Streaming Based on SVC ..... *M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon* 1227
State-of-the-Art and Trends in Scalable Video Compression With Wavelet Based Approaches ..... *N. Adami, A. Signoroni, and R. Leonardi* 1238
Barbell-Lifting Based 3-D Wavelet Coding Scheme ..... *R. Xiong, J. Xu, F. Wu, and S. Li* 1256

CALLS FOR PAPERS

ISCAS 2008 1270
Special Issue on Video Surveillance 1271
Special Issue on Event Analysis in Videos 1272

**********ORIGIN MIXED ADC 180
153 092007 22010702 00/005
CARNEGIE LIB-PITTSBURGH 0006
ACQUISITIONS/PERIODICALS
4400 FORBES AVE
PITTSBURGH PA 15213-4007 000231

Overview of the Scalable Video Coding Extension of the H.264/AVC Standard

Heiko Schwarz, Detlev Marpe, *Member, IEEE*, and Thomas Wiegand, *Member, IEEE*

(Invited Paper)

***Abstract*—With the introduction of the H.264/AVC video coding standard, significant improvements have recently been demonstrated in video compression capability. The Joint Video Team of the ITU-T VCEG and the ISO/IEC MPEG has now also standardized a Scalable Video Coding (SVC) extension of the H.264/AVC standard. SVC enables the transmission and decoding of partial bit streams to provide video services with lower temporal or spatial resolutions or reduced fidelity while retaining a reconstruction quality that is high relative to the rate of the partial bit streams. Hence, SVC provides functionalities such as graceful degradation in lossy transmission environments as well as bit rate, format, and power adaptation. These functionalities provide enhancements to transmission and storage applications. SVC has achieved significant improvements in coding efficiency with an increased degree of supported scalability relative to the scalable profiles of prior video coding standards. This paper provides an overview of the basic concepts for extending H.264/AVC towards SVC. Moreover, the basic tools for providing temporal, spatial, and quality scalability are described in detail and experimentally analyzed regarding their efficiency and complexity.**

***Index Terms*—H.264/AVC, MPEG-4, Scalable Video Coding (SVC), standards, video.**

I. Introduction

ADVANCES in video coding technology and standardization [1]–[6] along with the rapid developments and improvements of network infrastructures, storage capacity, and computing power are enabling an increasing number of video applications. Application areas today range from multimedia messaging, video telephony, and video conferencing over mobile TV, wireless and wired Internet video streaming, standard- and high-definition TV broadcasting to DVD, Blu-ray Disc, and HD DVD optical storage media. For these applications, a variety of video transmission and storage systems may be employed.

Traditional digital video transmission and storage systems are based on H.222.0 | MPEG-2 systems [7] for broadcasting services over satellite, cable, and terrestrial transmission channels, and for DVD storage, or on H.320 [8] for conversational video conferencing services. These channels are typically characterized by a fixed spatio–temporal format of the video signal (SDTV or HDTV or CIF for H.320 video telephone). Their application behavior in such systems typically falls into one of the two categories: it works or it does not work.

Modern video transmission and storage systems using the Internet and mobile networks are typically based on RTP/IP [9] for real-time services (conversational and streaming) and on computer file formats like mp4 or 3gp. Most RTP/IP access networks are typically characterized by a wide range of connection qualities and receiving devices. The varying connection quality is resulting from adaptive resource sharing mechanisms of these networks addressing the time varying data throughput requirements of a varying number of users. The variety of devices with different capabilities ranging from cell phones with small screens and restricted processing power to high-end PCs with high-definition displays results from the continuous evolution of these endpoints.

Scalable Video Coding (SVC) is a highly attractive solution to the problems posed by the characteristics of modern video transmission systems. The term "scalability" in this paper refers to the removal of parts of the video bit stream in order to adapt it to the various needs or preferences of end users as well as to varying terminal capabilities or network conditions. The term SVC is used interchangeably in this paper for both the concept of SVC in general and for the particular new design that has been standardized as an extension of the H.264/AVC standard. The objective of the SVC standardization has been to enable the encoding of a high-quality video bit stream that contains one or more subset bit streams that can themselves be decoded with a complexity and reconstruction quality similar to that achieved using the existing H.264/AVC design with the same quantity of data as in the subset bit stream.

SVC has been an active research and standardization area for at least 20 years. The prior international video coding standards H.262 | MPEG-2 Video [3], H.263 [4], and MPEG-4 Visual [5] already include several tools by which the most important scalability modes can be supported. However, the scalable profiles of those standards have rarely been used. Reasons for that include the characteristics of traditional video transmission systems as well as the fact that the spatial and quality scalability features came along with a significant loss in coding efficiency as well as a large increase in decoder complexity as compared to the corresponding nonscalable profiles. It should be noted that two or more single-layer streams, i.e., nonscalable streams, can always be transmitted by the method of *simulcast*, which in principle provides similar functionalities as a scalable bit stream,

Manuscript received October 6, 2006; revised July 15, 2007. This paper was recommended by Guest Editor T. Wiegand.

The authors are with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, 10587 Berlin, Germany (e-mail: hschwarz@hhi.hg.de; marpe@hhi.fhg.de; wiegand@hhi.fhg.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.905532

1051-8215/$25.00 © 2007 IEEE

although typically at the cost of a significant increase in bit rate. Moreover, the adaptation of a single stream can be achieved through transcoding, which is currently used in multipoint control units in video conferencing systems or for streaming services in 3G systems. Hence, a scalable video codec has to compete against these alternatives.

This paper describes the SVC extension of H.264/AVC and is organized as follows. Section II explains the fundamental scalability types and discusses some representative applications of SVC as well as their implications in terms of essential requirements. Section III gives the history of SVC. Section IV briefly reviews basic design concepts of H.264/AVC. In Section V, the concepts for extending H.264/AVC toward na SVC standard are described in detail and analyzed regarding effectiveness and complexity. The SVC high-level design is summarized in Section VI. For more detailed information about SVC, the reader is referred to the draft standard [10].

II. Types of Scalability, Applications, and Requirements

In general, a video bit stream is called scalable when parts of the stream can be removed in a way that the resulting substream forms another valid bit stream for some target decoder, and the substream represents the source content with a reconstruction quality that is less than that of the complete original bit stream but is high when considering the lower quantity of remaining data. Bit streams that do not provide this property are referred to as single-layer bit streams. The usual modes of scalability are temporal, spatial, and quality scalability. Spatial scalability and temporal scalability describe cases in which subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) or frame rate (temporal resolution), respectively. With quality scalability, the substream provides the same spatio–temporal resolution as the complete bit stream, but with a lower fidelity—where fidelity is often informally referred to as signal-to-noise ratio (SNR). Quality scalability is also commonly referred to as fidelity or SNR scalability. More rarely required scalability modes are region-of-interest (ROI) and object-based scalability, in which the substreams typically represent spatially contiguous regions of the original picture area. The different types of scalability can also be combined, so that a multitude of representations with different spatio–temporal resolutions and bit rates can be supported within a single scalable bit stream.

Efficient SVC provides a number of benefits in terms of applications [11]–[13]—a few of which will be briefly discussed in the following. Consider, for instance, the scenario of a video transmission service with heterogeneous clients, where multiple bit streams of the same source content differing in coded picture size, frame rate, and bit rate should be provided simultaneously. With the application of a properly configured SVC scheme, the source content has to be encoded only once—for the highest required resolution and bit rate, resulting in a scalable bit stream from which representations with lower resolution and/or quality can be obtained by discarding selected data. For instance, a client with restricted resources (display resolution, processing power, or battery power) needs to decode only a part of the delivered bit stream. Similarly, in a multicast scenario, terminals with different capabilities can be served by a single scalable bit stream. In an alternative scenario, an existing video format (like QVGA) can be extended in a backward compatible way by an enhancement video format (like VGA).

Another benefit of SVC is that a scalable bit stream usually contains parts with different importance in terms of decoded video quality. This property in conjunction with unequal error protection is especially useful in any transmission scenario with unpredictable throughput variations and/or relatively high packet loss rates. By using a stronger protection of the more important information, error resilience with graceful degradation can be achieved up to a certain degree of transmission errors. Media-Aware Network Elements (MANEs), which receive feedback messages about the terminal capabilities and/or channel conditions, can remove the nonrequired parts from a scalable bit stream, before forwarding it. Thus, the loss of important transmission units due to congestion can be avoided and the overall error robustness of the video transmission service can be substantially improved.

SVC is also highly desirable for surveillance applications, in which video sources not only need to be viewed on multiple devices ranging from high-definition monitors to videophones or PDAs, but also need to be stored and archived. With SVC, for instance, high-resolution/high-quality parts of a bit stream can ordinarily be deleted after some expiration time, so that only low-quality copies of the video are kept for long-term archival. The latter approach may also become an interesting feature in personal video recorders and home networking.

Even though SVC schemes offer such a variety of valuable functionalities, the scalable profiles of existing standards have rarely been used in the past, mainly because spatial and quality scalability have historically come at the price of increased decoder complexity and significantly decreased coding efficiency. In contrast to that, temporal scalability is often supported, e.g., in H.264/AVC-based applications, but mainly because it comes along with a substantial coding efficiency improvement (cf. Section V-A.2).

H.264/AVC is the most recent international video coding standard. It provides significantly improved coding efficiency in comparison to all prior standards [14]. H.264/AVC has attracted a lot of attention from industry and has been adopted by various application standards and is increasingly used in a broad variety of applications. It is expected that in the near-term future H.264/AVC will be commonly used in most video applications. Given this high degree of adoption and deployment of the new standard and taking into account the large investments that have already been taken place for preparing and developing H.264/AVC-based products, it is quite natural to now build a SVC scheme as an extension of H.264/AVC and to reuse its key features.

Considering the needs of today's and future video applications as well as the experiences with scalable profiles in the past, the success of any future SVC standard critically depends on the following essential requirements.

- Similar coding efficiency compared to single-layer coding—for each subset of the scalable bit stream.

- Little increase in decoding complexity compared to single-layer decoding that scales with the decoded spatio–temporal resolution and bit rate.
- Support of temporal, spatial, and quality scalability.
- Support of a backward compatible base layer (H.264/AVC in this case).
- Support of simple bit stream adaptations after encoding.

In any case, the coding efficiency of scalable coding should be clearly superior to that of "simulcasting" the supported spatio–temporal resolutions and bit rates in separate bit streams. In comparison to single-layer coding, bit rate increases of 10% to 50% for the same fidelity might be tolerable depending on the specific needs of an application and the supported degree of scalability.

This paper provides an overview how these requirements have been addressed in the design of the SVC extension of H.264/AVC.

III. History of SVC

Hybrid video coding, as found in H.264/AVC [6] and all past video coding designs that are in widespread application use, is based on motion-compensated temporal differential pulse code modulation (DPCM) together with spatial decorrelating transformations [15]. DPCM is characterized by the use of synchronous prediction loops at the encoder and decoder. Differences between these prediction loops lead to a "drift" that can accumulate over time and produce annoying artifacts. However, the scalability bit stream adaptation operation, i.e., the removal of parts of the video bit stream can produce such differences.

Subband or transform coding does not have the drift property of DPCM. Therefore, video coding techniques based on motion-compensated 3-D wavelet transforms have been studied extensively for use in SVC [16]–[19]. The progress in wavelet-based video coding caused MPEG to start an activity on exploring this technology. As a result, MPEG issued a call for proposals for efficient SVC technology in October 2003 with the intention to develop a new SVC standard. Twelve of the 14 submitted proposals in response to this call [20] represented scalable video codecs based on 3-D wavelet transforms, while the remaining two proposals were extensions of H.264/AVC [6]. After a six-month evaluation phase, in which several subjective tests for a variety of conditions were carried out and the proposals were carefully analyzed regarding their potential for a successful future standard, the scalable extension of H.264/AVC as proposed in [21] was chosen as the starting point [22] of MPEG's SVC project in October 2004. In January 2005, MPEG and VCEG agreed to jointly finalize the SVC project as an Amendment of H.264/AVC within the Joint Video Team.

Although the initial design [21] included a wavelet-like decomposition structure in temporal direction, it was later removed from the SVC specification [10]. Reasons for that removal included drastically reduced encoder and decoder complexity and improvements in coding efficiency. It was shown that an adjustment of the DPCM prediction structure can lead to a significantly improved drift control as will be shown in the paper. Despite this change, most components of the proposal in [21] remained unchanged from the first model [22] to the latest draft [10] being augmented by methods for nondyadic scalability and interlaced processing which were not included in the initial design.

IV. H.264/AVC Basics

SVC was standardized as an extension of H.264/AVC. In order to keep the paper self-contained, the following brief description of H.264/AVC is limited to those key features that are relevant for understanding the concepts of extending H.264/AVC towards SVC. For more detailed information about H.264/AVC, the reader is referred to the standard [6] or corresponding overview papers [23]–[26].

Conceptually, the design of H.264/AVC covers a *Video Coding Layer* (VCL) and a *Network Abstraction Layer* (NAL). While the VCL creates a coded representation of the source content, the NAL formats these data and provides header information in a way that enables simple and effective customization of the use of VCL data for a broad variety of systems.

A. Network Abstraction Layer (NAL)

The coded video data are organized into NAL units, which are packets that each contains an integer number of bytes. A NAL unit starts with a one-byte header, which signals the type of the contained data. The remaining bytes represent payload data. NAL units are classified into VCL NAL units, which contain coded slices or coded slice data partitions, and non-VCL NAL units, which contain associated additional information. The most important non-VCL NAL units are parameter sets and Supplemental Enhancement Information (SEI). The sequence and picture parameter sets contain infrequently changing information for a video sequence. SEI messages are not required for decoding the samples of a video sequence. They provide additional information which can assist the decoding process or related processes like bit stream manipulation or display. A set of consecutive NAL units with specific properties is referred to as an access unit. The decoding of an access unit results in exactly one decoded picture. A set of consecutive access units with certain properties is referred to as a coded video sequence. A coded video sequence represents an independently decodable part of a NAL unit bit stream. It always starts with an instantaneous decoding refresh (IDR) access unit, which signals that the IDR access unit and all following access units can be decoded without decoding any previous pictures of the bit stream.

B. Video Coding Layer (VCL)

The VCL of H.264/AVC follows the so-called block-based hybrid video coding approach. Although its basic design is very similar to that of prior video coding standards such as H.261, MPEG-1 Video, H.262 | MPEG-2 Video, H.263, or MPEG-4 Visual, H.264/AVC includes new features that enable it to achieve a significant improvement in compression efficiency relative to any prior video coding standard [14]. The main difference to previous standards is the largely increased flexibility and adaptability of H.264/AVC.

The way pictures are partitioned into smaller coding units in H.264/AVC, however, follows the rather traditional concept of subdivision into *macroblocks* and *slices*. Each picture is partitioned into macroblocks that each covers a rectangular pic-

ture area of 16 × 16 luma samples and, in the case of video in 4:2:0 chroma sampling format, 8 × 8 samples of each of the two chroma components. The samples of a macroblock are either spatially or temporally predicted, and the resulting prediction residual signal is represented using transform coding. The macroblocks of a picture are organized in slices, each of which can be parsed independently of other slices in a picture. Depending on the degree of freedom for generating the prediction signal, H.264/AVC supports three basic slice coding types.

1) I-slice: *intra-picture* predictive coding using spatial prediction from neighboring regions,
2) P-slice: intra-picture predictive coding and inter-picture *predictive* coding with one prediction signal for each predicted region,
3) B-slice: intra-picture predictive coding, inter-picture predictive coding, and inter-picture *bipredictive* coding with two prediction signals that are combined with a weighted average to form the region prediction.

For I-slices, H.264/AVC provides several directional spatial intra-prediction modes, in which the prediction signal is generated by using neighboring samples of blocks that precede the block to be predicted in coding order. For the luma component, the intra-prediction is either applied to 4 × 4, 8 × 8, or 16 × 16 blocks, whereas for the chroma components, it is always applied on a macroblock basis.[1]

For P- and B-slices, H.264/AVC additionally permits variable block size motion-compensated prediction with multiple reference pictures [27]. The macroblock type signals the partitioning of a macroblock into blocks of 16 × 16, 16 × 8, 8 × 16, or 8 × 8 luma samples. When a macroblock type specifies partitioning into four 8 × 8 blocks, each of these so-called *submacroblocks* can be further split into 8 × 4, 4 × 8, or 4 × 4 blocks, which is indicated through the submacroblock type. For P-slices, one motion vector is transmitted for each block. In addition, the used reference picture can be independently chosen for each 16 × 16, 16 × 8, or 8 × 16 macroblock partition or 8 × 8 submacroblock. It is signaled via a reference index parameter, which is an index into a list of reference pictures that is replicated at the decoder.

In B-slices, two distinct reference picture lists are utilized, and for each 16 × 16, 16 × 8, or 8 × 16 macroblock partition or 8 × 8 submacroblock, the prediction method can be selected between *list 0*, *list 1*, or *biprediction*. While list 0 and list 1 prediction refer to unidirectional prediction using a reference picture of reference picture list 0 or 1, respectively, in the bipredictive mode, the prediction signal is formed by a weighted sum of a list 0 and list 1 prediction signal. In addition, special modes as so-called *direct modes* in B-slices and *skip modes* in P- and B-slices are provided, in which such data as motion vectors and reference indexes are derived from previously transmitted information.

For transform coding, H.264/AVC specifies a set of *integer transforms* of different block sizes. While for intra-macroblocks the transform size is directly coupled to the intra-prediction block size, the luma signal of motion-compensated macroblocks that do not contain blocks smaller than 8 × 8 can be coded by using either a 4 × 4 or 8 × 8 transform. For the chroma components a two-stage transform, consisting of 4 × 4 transforms and a Hadamard transform of the resulting DC coefficients is employed.[1] A similar hierarchical transform is also used for the luma component of macroblocks coded in intra 16 × 16 mode. All inverse transforms are specified by exact integer operations, so that inverse-transform mismatches are avoided. H.264/AVC uses *uniform reconstruction quantizers*. One of 52 quantization step sizes[1] can be selected for each macroblock by the quantization parameter QP. The scaling operations for the quantization step sizes are arranged with logarithmic step size increments, such that an increment of the QP by 6 corresponds to a doubling of quantization step size.

For reducing blocking artifacts, which are typically the most disturbing artifacts in block-based coding, H.264/AVC specifies an *adaptive deblocking filter*, which operates within the motion-compensated prediction loop.

H.264/AVC supports two methods of entropy coding, which both use context-based adaptivity to improve performance relative to prior standards. While context-based adaptive variable-length coding (CAVLC) uses variable-length codes and its adaptivity is restricted to the coding of transform coefficient levels, context-based adaptive binary arithmetic coding (CABAC) utilizes arithmetic coding and a more sophisticated mechanism for employing statistical dependencies, which leads to typical bit rate savings of 10%–15% relative to CAVLC.

In addition to the increased flexibility on the macroblock level, H.264/AVC also allows much more flexibility on a picture and sequence level compared to prior video coding standards. Here we mainly refer to *reference picture memory control*. In H.264/AVC, the coding and display order of pictures is completely decoupled. Furthermore, any picture can be marked as reference picture for use in motion-compensated prediction of following pictures, independent of the slice coding types. The behavior of the *decoded picture buffer* (DPB), which can hold up to 16 frames (depending on the used conformance point and picture size), can be adaptively controlled by *memory management control operation* (MMCO) commands, and the reference picture lists that are used for coding of P- or B-slices can be arbitrarily constructed from the pictures available in the DPB via *reference picture list reordering* (RPLR) commands.

In order to enable a flexible partitioning of a picture into slices, the concept of *slice groups* was introduced in H.264/AVC. The macroblocks of a picture can be arbitrarily partitioned into slice groups via a *slice group map*. The slice group map, which is specified by the content of the picture parameter set and some slice header information, assigns a unique slice group identifier to each macroblock of a picture. And each slice is obtained by scanning the macroblocks of a picture that have the same slice group identifier as the first macroblock of the slice in raster-scan order. Similar to prior video coding standards, a *picture* comprises the set of slices representing a complete frame or one field of a frame (such that, e.g., an interlaced-scan picture can be either coded as a single frame picture or two separate field pictures). Additionally, H.264/AVC supports a macroblock-adaptive switching between frame and field coding. For that, a pair of vertically adjacent macroblocks is considered as a single coding unit,

[1]Some details of the profiles of H.264/AVC that were designed primarily to serve the needs of professional application environments are neglected in this description, particularly in relation to chroma processing and range of step sizes.

which can be either transmitted as two spatially neighboring frame macroblocks, or as interleaved top and a bottom field macroblocks.

V. Basic Concepts for Extending H.264/AVC Towards an SVC Standard

Apart from the required support of all common types of scalability, the most important design criteria for a successful SVC standard are coding efficiency and complexity, as was noted in Section II. Since SVC was developed as an extension of H.264/AVC with all of its well-designed core coding tools being inherited, one of the design principles of SVC was that new tools should only be added if necessary for efficiently supporting the required types of scalability.

A. Temporal Scalability

A bit stream provides temporal scalability when the set of corresponding access units can be partitioned into a temporal base layer and one or more temporal enhancement layers with the following property. Let the temporal layers be identified by a temporal layer identifier T, which starts from 0 for the base layer and is increased by 1 from one temporal layer to the next. Then for each natural number k, the bit stream that is obtained by removing all access units of all temporal layers with a temporal layer identifier T greater than k forms another valid bit stream for the given decoder.

For hybrid video codecs, temporal scalability can generally be enabled by restricting motion-compensated prediction to reference pictures with a temporal layer identifier that is less than or equal to the temporal layer identifier of the picture to be predicted. The prior video coding standards MPEG-1 [2], H.262 | MPEG-2 Video [3], H.263 [4], and MPEG-4 Visual [5] all support temporal scalability to some degree. H.264/AVC [6] provides a significantly increased flexibility for temporal scalability because of its reference picture memory control. It allows the coding of picture sequences with arbitrary temporal dependencies, which are only restricted by the maximum usable DPB size. Hence, for supporting temporal scalability with a reasonable number of temporal layers, no changes to the design of H.264/AVC were required. The only related change in SVC refers to the signaling of temporal layers, which is described in Section VI.

1) Hierarchical Prediction Structures: Temporal scalability with dyadic temporal enhancement layers can be very efficiently provided with the concept of hierarchical B-pictures [28], [29] as illustrated in Fig. 1(a).[2] The enhancement layer pictures are typically coded as B-pictures, where the reference picture lists 0 and 1 are restricted to the temporally preceding and succeeding picture, respectively, with a temporal layer identifier less than the temporal layer identifier of the predicted picture. Each set of temporal layers $\{T_0, \ldots, T_k\}$ can be decoded independently of all layers with a temporal layer identifier $T > k$. In the following, the set of pictures between two successive pictures of the temporal base layer together with the succeeding base layer picture is referred to as a *group of pictures* (GOP).

[2]As described above, neither P- or B-slices are directly coupled with the management of reference pictures in H.264/AVC. Hence, backward prediction is not necessarily coupled with the use of B-slices and the temporal coding structure of Fig. 1(a) can also be realized using P-slices resulting in a structure that is often called hierarchical P-pictures.

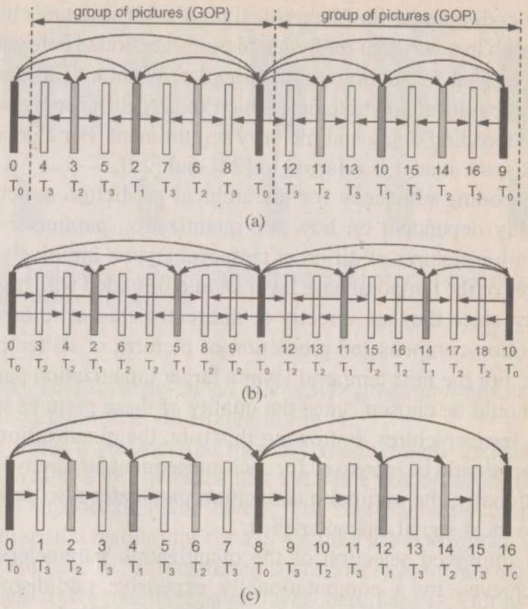


Fig. 1. Hierarchical prediction structures for enabling temporal scalability. (a) Coding with hierarchical B-pictures. (b) Nondyadic hierarchical prediction structure. (c) Hierarchical prediction structure with a structural encoding/decoding delay of zero. The numbers directly below the pictures specify the coding order, the symbols T_k specify the temporal layers with k representing the corresponding temporal layer identifier.

Although the described prediction structure with hierarchical B-pictures provides temporal scalability and also shows excellent coding efficiency as will be demonstrated later, it represents a special case. In general, hierarchical prediction structures for enabling temporal scalability can always be combined with the multiple reference picture concept of H.264/AVC. This means that the reference picture lists can be constructed by using more than one reference picture, and they can also include pictures with the same temporal level as the picture to be predicted. Furthermore, hierarchical prediction structures are not restricted to the dyadic case. As an example, Fig. 1(b) illustrates a nondyadic hierarchical prediction structure, which provides 2 independently decodable subsequences with 1/9th and 1/3rd of the full frame rate. It should further be noted that it is possible to arbitrarily modify the prediction structure of the temporal base layer, e.g., in order to increase the coding efficiency. The chosen temporal prediction structure does not need to be constant over time.

Note that it is possible to arbitrarily adjust the structural delay between encoding and decoding a picture by restricting motion-compensated prediction from pictures that follow the picture to be predicted in display order. As an example, Fig. 1(c) shows a hierarchical prediction structure, which does not employ motion-compensated prediction from pictures in the future. Although this structure provides the same degree of temporal scalability as the prediction structure of Fig. 1(a), its structural delay is equal to zero compared to 7 pictures for the prediction structure in Fig. 1(a). However, such low-delay structures typically decrease coding efficiency.