Implementation and evaluation of packet loss concealment schemes with the JM reference software
Implementation and evaluation of packet loss concealment schemes with the JM reference software

Examensarbete utfört i Kommunikationssystem vid Tekniska högskolan i Linköping av

Henrik Cooke

LiTH-ISY-EX–10/4339–SE

Handledare: Daniel Persson
ISY, Linköpings universitet
Renat Vafin
Skype

Examinator: Jonas Eriksson
ISY, Linköpings universitet

Linköping, 15 April, 2010
Implementation och utvärdering av metoder för att dölja paketförluster med JM-referensmjukvaran

Implementation and evaluation of packet loss concealment schemes with the JM reference software

Henrik Cooke

Communication over today’s IP-based networks are to some extent subject to packet loss. Most real-time applications, such as video streaming, need methods to hide this effect, since resending lost packets may introduce unacceptable delays. For IP-based video streaming applications such a method is referred to as a packet loss concealment scheme.

In this thesis a recently proposed mixture model and least squares-based packet loss concealment scheme is implemented and evaluated together with three more well known concealment methods. The JM reference software is used as basis for the implementation, which is a public available software codec for the H.264 video coding standard. The evaluation is carried out by comparing the schemes in terms of objective measurements, subjective observations and a study with human observers.

The recently proposed packet loss concealment scheme shows good performance with respect to the objective measures, and careful observations indicate better concealment of scenes with fast motion and rapidly changing video content. The study with human observers verifies the results for the case when a more sophisticated packetization technique is used.

A new packet loss concealment scheme, based on joint modeling of motion vectors and pixels, is also investigated in the last chapter as an additional contribution of the thesis.

Keywords
packet loss concealment, mixture model, H.264
Abstract

Communication over today’s IP-based networks are to some extent subject to packet loss. Most real-time applications, such as video streaming, need methods to hide this effect, since resending lost packets may introduce unacceptable delays. For IP-based video streaming applications such a method is referred to as a packet loss concealment scheme.

In this thesis a recently proposed mixture model and least squares-based packet loss concealment scheme is implemented and evaluated together with three more well known concealment methods. The JM reference software is used as basis for the implementation, which is a public available software codec for the H.264 video coding standard. The evaluation is carried out by comparing the schemes in terms of objective measurements, subjective observations and a study with human observers.

The recently proposed packet loss concealment scheme shows good performance with respect to the objective measures, and careful observations indicate better concealment of scenes with fast motion and rapidly changing video content. The study with human observers verifies the results for the case when a more sophisticated packetization technique is used.

A new packet loss concealment scheme, based on joint modeling of motion vectors and pixels, is also investigated in the last chapter as an additional contribution of the thesis.
I would like to thank my supervisor at the University, Daniel Persson, for all help, inspiring ideas and the research which this thesis is based upon, and my supervisor at Skype, Renat Vafin, for his guidance and for taking the initiative to the thesis.

During my time at the University I have had help from various persons at the Division of Communication Systems who I would like to thank for their input. Many thanks to my room mate and parallel master thesis worker, Kristoffer Eriksson, who has helped me with a little bit of everything, from math to LaTeX.

Last but not least I would like to thank my lovely wife, Hannah Cooke, and my family for supporting me through the years of studies and exams.
Notation

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>error concealment</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian mixture model</td>
</tr>
<tr>
<td>JM</td>
<td>joint model</td>
</tr>
<tr>
<td>JVT</td>
<td>joint video team</td>
</tr>
<tr>
<td>MB</td>
<td>macroblock</td>
</tr>
<tr>
<td>MV</td>
<td>motion vector</td>
</tr>
<tr>
<td>MMLS</td>
<td>mixture model and least square</td>
</tr>
<tr>
<td>NAL</td>
<td>network adaptation layer</td>
</tr>
<tr>
<td>PDF</td>
<td>probability density function</td>
</tr>
<tr>
<td>PLC</td>
<td>packet loss concealment</td>
</tr>
<tr>
<td>PSNR</td>
<td>peak signal-to-noise ration</td>
</tr>
<tr>
<td>VCL</td>
<td>video coding layer</td>
</tr>
</tbody>
</table>
# Contents

1 Introduction
   1.1 Problem definition ........................................... 1
   1.2 Methodology ..................................................... 1
   1.3 About the report .............................................. 2
      1.3.1 Disposition .................................................. 2
      1.3.2 Prerequisites ............................................... 2
      1.3.3 A new packet loss concealment scheme .................. 2

2 H.264
   2.1 Colour space and sampling formats .......................... 3
   2.2 Macroblocks, motion-compensation and inter-prediction .... 4
   2.3 Intra-prediction ............................................... 4
   2.4 Slices and slice groups ...................................... 4

3 Packet loss concealment schemes
   3.1 Previous efforts ............................................... 7
   3.2 Mixture Model- and Least Squares-Based Packet Video Error Concealment ......................... 8
      3.2.1 Background ................................................ 8
      3.2.2 Theory ..................................................... 9
      3.2.3 Parameters and estimators ................................ 11
   3.3 Packet loss concealment in the H.264 JM reference software .... 12
      3.3.1 Temporal error concealment ............................... 12
      3.3.2 Parameters ............................................... 12
   3.4 Q.-F. Zhu’s spatio-temporal packet loss concealment scheme .... 13
      3.4.1 Theory ..................................................... 13
      3.4.2 Parameters ............................................... 14
   3.5 W. Zhu’s spatial packet loss concealment scheme ............. 14
      3.5.1 Theory ..................................................... 14
      3.5.2 Parameters ............................................... 15
   3.6 Common properties and strategies ................................ 15
      3.6.1 Concealment block order ................................ 16
      3.6.2 Concealment of entire frame losses ...................... 16
      3.6.3 Motion vector estimation ................................ 16
Chapter 1

Introduction

Communication over IP-based networks are subject to various transmission errors, resulting in lost IP packets. Real-time applications, such as video streaming over the Internet, where resending lost packets will result in significant delays, are in need of methods to hide this effect. Methods implemented at the decoder side of an IP-based video streaming application, with the purpose of concealing the lost packets and which do not require adding any extra redundancy to the video stream, are referred to as packet loss concealment (PLC) schemes.

Recently D. Persson and T. Eriksson proposed a mixture model- and least squares (MMLS)-based PLC scheme [16] which has shown good performance with respect to peak signal-to-noise ratio (PSNR) and complexity. The method had at the time of publication yet to be implemented within the context of a real decoder for further evaluation.

1.1 Problem definition

This thesis aims to implement and evaluate the proposed MMLS-based PLC scheme in [16], with the H.264 video codec in the form of the JM (Joint Model) reference software. The evaluation includes testing the performance, objectively and subjectively, for a set of relevant codec settings and loss scenarios, and comparing against other PLC schemes. Also some minor adjustments of the algorithm can come into play, if experiments indicate a gain in performance or clearly motivated simplifications or approximations can be applied without significant loss in performance.

1.2 Methodology

The JM reference decoder (a part of the reference software) provides the basis for the implementation of the proposed MMLS-based PLC scheme and three other PLC schemes, which are subject to comparisons as a part of the evaluation. Further the parameter estimation, for the model used in the proposed method, is
implemented with help of Matlab. Two different video databases are built by randomly choosing videos from an open source video archive [1]. The videos in [1] are a mixture of different types of movies, collected over many years. One of the video databases, referred to as the video training database, is used for parameter estimation and settings investigation. The other database is used in the objective evaluation of the PLC methods, and is referred to as the video evaluation database. Two different packet loss models are used in the evaluation.

For the objective evaluation, the notion of PSNR is introduced as means of measure. For the subjective evaluation, a study with human observers is performed in conjunction with observations of video sequences restored with the different PLC schemes.

1.3 About the report

1.3.1 Disposition

The report starts with a short theoretical background of the H.264 standard, and the tools it provides to build an error resilient video stream. As an introduction to the theory of the PLC schemes, a section is dedicated to give a historical background to the packet loss concealment problem, and what has been achieved in this area of research. Further the theory of the MMLS-based concealment scheme will be presented together with three other concealment schemes. In the end of the report, the simulation configurations, measurements and results are presented followed by conclusions.

1.3.2 Prerequisites

It is assumed that the reader is familiar with basic statistical mathematics and signal processing. A short introduction to the concepts of modern video compression techniques is given, but for a reader who are new to the subject further studies may be required to fully grasp the details of this report. Recommended reading is Introduction to Data Compression by Khalid Sayood [19] and H.264 and MPEG-4 Video Compression - Video Coding for Next-generation Multimedia [18].

1.3.3 A new packet loss concealment scheme

As an additional contribution to the thesis, a new PLC scheme is proposed and investigated in chapter 6. It is a separate and preliminary investigation of a new PLC method, and it is not a part of the problem definition.
Chapter 2

H.264

This chapter describes some of the components in the H.264 standard which are important for the understanding of the PLC problem. Some descriptions are somewhat simplified, to not introduce aspects which are not covered by the thesis. For a more detailed overview of the standard see [26].

H.264, also known as ISO/IEC14496-10 or MPEG-4 Part 10 Advanced Video Codec (AVC), is a standard by the joint video team (JVT), consisting of members from MPEG and the ITU’s Video Coding Experts Group [25]. The standard is described from decoder point of view, constraining the format of the bitstream and syntax rather than describing how to encode a raw video stream. H.264 divides the codec into two layers: the video coding layer (VCL) which is responsible for the compression of the raw video data, and the network adaption layer (NAL) which adapts the bit stream generated by the VCL to various network environments.

2.1 Colour space and sampling formats

The standard is defined for the YCbCr colour space (sometimes referred to as YUV), which is a linear transform from the RGB colour space (which is assumed to be familiar to the reader). Y is the luminance component, and CbCr (UV) are the colour components, called chrominance or chroma. The transformation from RGB to YCbCr, with the coefficients used in the ITU-R recommendation BT.601 [11], is described by equations (2.1).

\[
\begin{align*}
Y &= 0.299R + 0.587G + 0.114B \\
Cb &= 0.564(B - Y) \\
Cr &= 0.713(R - Y)
\end{align*}
\] (2.1)

The human visual system is less sensitive to colour than to luminance [18], and because of this the resolution can be lower for the CbCr components than for the Y component, without significant loss in perceived image quality. Several sampling formats for the components exists. In this thesis we will work with 4:2:0
H.264 sampling format (YUV420), where the CbCr components are represented by a quarter as many samples each as the Y component. This is achieved with half sample rate for the CbCr components, compared to the sample rate for the Y component, both horizontally and vertically. The notion of pixel and sample will be used interchangeably in this report, and does only refer to a single scalar value of one of the three components at a time. The three components are dealt with separately in the decoding process, if not otherwise noted.

2.2 Macroblocks, motion-compensation and inter-prediction

A video sequence is divided into frames, where one frame is a rectangular area of pixels associated with a timestamp. The VCL divides every frame in a video stream into macroblocks (MBs), consisting of 16 × 16 luminance samples each and its corresponding two blocks of 8 × 8 chroma samples. For inter-predictively encoded frames (P-frames or temporally predicted frames), the decoder receives motion vectors (MVs) and pixel differences, called residuals, for every MB. Such a MV represents the movement of a group of pixels (both luminance and chroma) from some previously decoded frame, known as reference frame, and the current frame to be decoded. The residual is the difference between the pixels in the reference frame, pointed to by the MV, and the actual pixels in the MB to be decoded. The decoder uses the MVs to motion-compensate the blocks in the current frame, and then applies the residuals by addition to obtain the decoded frame. In this thesis we only consider the case when the MVs are transmitted in the same packet as their corresponding residuals. Further we assume that the motion-compensation is done using the last encoded frame as reference frame.

2.3 Intra-prediction

For various reasons one may want to decode a frame independently of previously decoded frames. For example when one wants to stop errors introduced in previously decoded frames from propagating into subsequent frames. For this scenario a frame may be intra-coded (spatially coded), meaning that the pixels are predicted from their spatial neighbors. This spatial prediction is causal and works on per-pixel level, when indexing pixels in raster scan order, from left to right and downwards. The details of the intra-prediction is not covered by this overview, and intra-prediction is only used for the very first frame for the sequences in this thesis.

2.4 Slices and slice groups

To increase robustness, which means to make the video stream less sensitive to packet loss, and to deal with packet size limitations, the MBs of a frame are grouped into slices which may be sent in separate packets. The MBs within a slice
must be encoded such that they do not depend on any other MBs in any other slice of the same frame. This constraint allows the receiver to decode the slices independently of each other. The MBs are put into the slices in raster scan order, and when the first slice is full the following MBs are put into the next slice and so forth.

To further increase robustness an optional feature called slice groups may be enabled. It allows the MBs of a frame to be put into slices in a predefined order, defined by a macroblock to slice group map. The frame is divided into slice groups by the macroblock to slice group map and the MBs within such a group are put into slices in raster scan order, as described above.

The different options for dividing the frames into slices and slice groups are in this thesis referred to as packetization techniques. Three packetization techniques, illustrated in figure 2.1, are of special interest: the simple slice setup with each frame divided into two slices (figure 2.1(a)), two interleaved slice groups (figure 2.1(b)) and two dispersed slice groups (figure 2.1(c)). In the presented interleaved mapping scheme all MBs on the same MB-row is mapped to the same slice group. In the shown dispersed mapping scheme every even numbered MB is put into one slice group and every odd numbered MB is put into another slice group, where the MBs are numbered in raster scan order.

A slice can be of one of five different slice types, containing MBs coded with different methods. The two slice types used in this thesis are the I-slice and the P-slice, where the MBs of an I-slice are intra-predicted and the MBs of a P-slice are either inter-predicted or intra-predicted.
Figure 2.1. Three packetization techniques.
Chapter 3

Packet loss concealment schemes

With the proposed MMLS-based scheme included, a total of four different PLC schemes are compared. Three of them are implemented in the thesis, and one is already included in the JM reference software. The three methods which the MMLS-based scheme is to be compared against are chosen to include various techniques and has shown good performance [27] [28]. This chapter will present all four techniques in theory, with practical concerns regarding special cases, improvements implemented in this thesis and the parameter values used in our simulations. Moreover the first section will give a short summary of previous efforts in the concealment area, and the last section will describe algorithms common to all four concealment schemes. In the descriptions of the PLC schemes it is assumed that the lost MVs are estimated prior to the concealment, as explained in section 3.6.3.

3.1 Previous efforts

PLC methods are usually categorized into spatial methods and temporal methods. Spatial methods only use spatially surrounding pixels for concealment of lost blocks and temporal methods try to estimate the motion and use pixels from previous frames for concealment. Spatial methods tend to yield better performance than temporal methods in sequences with high motion, or after a scene change. See [28], [23], [14] and [21] for some contributions.

Spatial methods can not recreate details inside lost blocks, in this case temporal information may improve the result. Instead of using the block in the previous frame at the same spatial position as the lost block for temporal PLC, the motion-compensated block should be used. When the MV is lost together with the lost block, it has to be estimated. One estimation technique is to use the median of the MVs of surrounding blocks, or the MV of the corresponding block in the previous frame [10]. Another technique, employed in the JM reference software, builds on smoothing constraints in the spatial domain as described in [13].
From an information-theoretic view, spatio-temporal approaches should improve the pure spatial and temporal efforts. The method in [27] extended the method in [23] to also include pixels from a previous frame in the concealment. In [20], MVs are estimated, the prediction error is modeled as a Gaussian Markov random field, and a maximum a posteriori estimate of the prediction error for the lost block is formed. Other recent contributions are mode selection in [5], combined and iterative spatio-temporal error concealment in [22], and error concealment by spatio-temporal boundary matching and a partial differential equation-based algorithm in [8].

3.2 Mixture Model- and Least Squares-Based Packet Video Error Concealment

The proposed mixture model- and least squares (MMLS)-based PLC scheme is a spatio-temporal method. The scheme was preceded by a spatio-temporal Gaussian mixture model (GMM)-based PLC method, presented in [17]. This preceding method is briefly described in the following section to give a background to the MMLS-based method.

3.2.1 Background

The preceding Gaussian mixture model PLC scheme is a spatio-temporal method based on Gaussian mixture modeling (GMM). The method described here was first presented in [17], but GMMs are commonly used for a number of tasks in image processing, and it is known that a GMM may describe probability distributions arbitrary well with an increasing amount of mixture components [17]. The different Gaussian distributions, called component distributions, define the components in the GMM, and combined together with a priori weights (component weights) they form the distribution that defines the GMM.

Let a group of lost pixels in a frame be denoted by a stochastic vector \( x \), and its temporally and spatially surrounding pixels be denoted by a stochastic vector \( y \). The lost pixels \( x \) may than be estimated from \( y \) using a GMM with \( M \) components which is described by the distribution

\[
f(x, y) = \sum_{m=1}^{M} \rho^{(m)} f^{(m)}(x, y),
\]

where \( \rho^{(m)} \) are the a priori weights. The functions \( f^{(m)}(x, y) \) are Gaussian distributions with covariance matrices \( C^{(m)}_{xy} \) and \( C^{(m)}_{yy} \), and mean vectors \( \mu^{(m)}_{x} \) and \( \mu^{(m)}_{y} \). An estimator \( \hat{x} \) for the lost pixels \( x \) is then given in (3.2) by considering a minimum mean squared error (MMSE) problem

\[
\hat{x}(y) = E_{X}(x|y) = \int x f(x|y)dx.
\]
3.2 Mixture Model- and Least Squares- Based Packet Video Error Concealment

With help of (3.1) the estimator can be written [17]

$$\hat{x}(y) = \sum_{m=1}^{M} \phi^{(m)}(y)(C_{xy}^{(m)}(C_{yy}^{(m)})^{-1}(y - \mu_y^{(m)}) + \mu_x^{(m)}) \quad (3.3)$$

where

$$\phi^{(m)}(y) = \frac{\rho^{(m)} f^{(m)}(y)}{\sum_{k=1}^{M} \rho^{(k)} f^{(k)}(y)} \quad (3.4)$$

are referred to as a posteriori weights.

The parameters of the GMM, the variances and the means, must be estimated offline with help of an estimation video database large enough to statistically represent the video sequences to be restored. The parameters can be calculated with the EM-algorithm (expectation maximization) [9].

Even though the concealment scheme has shown to increase PSNR compared to some other methods investigated in [17], the method has the following disadvantages:

- The iterative offline parameter estimation with the EM-algorithm yields a nondecreasing log-likelihood of the realizations from the video training database in each iteration. But an increase of log-likelihood of the vectors in this database is not necessarily linked to an increase in PSNR of the estimate (3.3).

- The estimator (3.3) has high computational complexity involving a large amount of floating point operations.

- The number of dimensions of $y$ can not be increased without increasing complexity both for the offline parameter estimation and the online pixel estimation with (3.3).

With the above method as background and its described disadvantages in mind, a new mixture-based spatio-temporal PLC scheme [16], which is described in the next section, was proposed.

3.2.2 Theory

With the GMM-based estimator described above as starting-point, the new estimator is described with the following changes to (3.3). The reason for the specific changes and the simplified approach is not explained in [16]. The means $\mu_x^{(m)}$ and $\mu_y^{(m)}$ are removed, the matrices $C_{xy}^{(m)} (C_{yy}^{(m)})^{-1}$ are replaced by $A^{(m)}$ and the posteriori weights $\phi^{(m)}$ are replaced by simpler functions $\pi^{(m)}$. The vector $y$ is divided into two subsets, $y_C$ and $y_P$, used for classification in the posteriori weights $\pi^{(m)}$ and prediction of $\hat{x}$ respectively. With this division of $y$, more pixels can be used in the posteriori weights than in the prediction. The new estimator is written

$$\hat{x}(y, \theta) = \sum_{m=1}^{M} \pi^{(m)}(y_C) A^{(m)} y_P, \quad (3.5)$$
Packet loss concealment schemes

\( \pi^{(m)}(y_C) = \frac{\rho^{(m)}h^{(m)}(y_C)}{\sum_{k=1}^{M} \rho^{(k)}h^{(k)}(y_C)} \),

(3.6)

\( h^{(m)}(y_C) = \exp \left( -c^{(m)} \frac{\| y_C^{(m),1} - y_C^{(m),2} \|_2^2}{D^{(m)}} \right) \),

(3.7)

where

\[ \theta = \left\{ \rho^{(1)}, \ldots, \rho^{(M)}, c^{(1)}, \ldots, c^{(M)}, A^{(1)}, \ldots, A^{(M)} \right\} , \]

\( y_C^{(m),1} \) and \( y_C^{(m),2} \) are vectors containing elements of \( y_C \), \( c^{(m)} > 0 \) is scalar, \( D^{(m)} \) is the dimension of the vectors \( y_C^{(m),1} \) and \( y_C^{(m),2} \).

The parameters list in \( \theta \) are estimated with a least squared (LS) approach

\[ \hat{\theta} = \arg \min_{\theta} E_{X,Y} \left( \| x - \hat{x}(y, \theta) \|_2^2 \right) , \]

(3.8)

which results in maximized PSNR in contrast to the parameter estimation to (3.3) which results in maximized log-likelihood. Since this LS problem does not have a closed form solution an iterative algorithm is used as follows:

1) The matrices \( A^{(m)} \) are determined by minimizing the MSE as a function of one matrix \( A^{(r)} \) at a time, while keeping all other parameters constant. This minimization is the standard linear MMSE problem with the solution

\[ A^{(r)*} = R_1(R_2)^{-1} , \]

(3.9)

\[ R_1 = E \left[ \pi^{(r)}(y_C)x y_P^T \right] - \sum_{m=1, m \neq r}^{M} A^{(m)} E \left[ \pi^{(r)}(y_C)\pi^{(m)}(y_C)y_P y_P^T \right] , \]

(3.10)

\[ R_2 = E \left[ \left( \pi^{(r)}(y_C) \right)^2 y_P y_P^T \right] . \]

(3.11)

Where in practice the expectations are replaced by empirical means of realizations from a video training database.

2) The parameters \( \rho^{(m)}, m = 1, \ldots, M \), are updated by minimizing the MSE while all other parameters are constant. The update is done by comparing three different parameter vectors \( \rho = [\rho^{(1)}, \ldots, \rho^{(M)}]^T \); \( \rho \) as obtained from previous update and \( \rho + \alpha [\rho^{(1)}b(1), \ldots, \rho^{(M)}b(M)] \) where \( \alpha \) is some scalar and \( b(1) \) to \( b(M) \) are elements of a normalized random vector. The vector that minimizes the MSE is kept. The same strategy is used for updating the parameters \( c^{(m)}, m = 1, \ldots, M \).

To summarize, the parameters \( \hat{\theta} \) can be obtained offline by using a training video database and an iterative algorithm maximizing PSNR in every iteration. The online estimator uses \( y_C \) and \( y_P \), which have dimensions that can be scaled independently, and where the vector \( y_C \) can have a high number of dimensions without a large impact on the complexity.

The vectors \( y_C^{(m),1} \) and \( y_C^{(m),2} \) should be chosen so that different situations in the video is classified properly. For example, to setup a mixture component that focuses on spatial correlation, \( y_C^{(m),1} \) and \( y_C^{(m),2} \) should be chosen so that (3.7)
contains differences of many spatial neighbors. The mixture components used in this thesis, and the choices of $y^{(m),1}_C$ and $y^{(m),2}_C$ is described in greater detail in the next section.

3.2.3 Parameters and estimators

The function of the estimator in (3.5) will differ depending on the choice of mapping between the pixels and the vectors $x$ and $y$. Different choices will give different estimators. Here we construct two estimators, for use in different scenarios depending on available surrounding pixels. Both estimators are heavily inspired by the ones presented in the experiments section of the original paper [16]. The estimators conceal lost blocks of $4 \times 4$ pixels at a time, where the lost pixels are represented by a vector $x$. The available surrounding pixels used in the estimation are represented by the vectors $y_C$ and $y_P$. Figure 3.1(a) and 3.1(b) illustrate the mapping between the pixels and the vectors for the two estimators respectively. The "larger" estimator shown in figure 3.1(a) is used when more surrounding pixels are available, and the "smaller" estimator in figure 3.1(b) is used if the first estimator is unusable due to lack of enough available surrounding pixels.

Five mixture components, $M = 5$, are selected for each estimator, as a result of the presented trade-off between PSNR and complexity (in terms of number of floating point operations) in [16], with respect to mixture components. These five mixture components focus on pixel correlation in the southwest, west, northwest, north and time direction respectively. This means that the mixture components include differences in pixel values for their respective direction in the exponent of (3.7).

The estimators are rotated to better adjust the concealment to available surrounding pixels. Some surrounding pixels may be more reliable for the estimation than others, see section 3.6 for further discussion. If neither of the two estimators, with any rotation, can be employed due to lack of surrounding pixels (lost block may be adjacent to a border or the surrounding pixels may also be lost) the estimator is replaced by motion copy concealment. Motion copy concealment is performed by copying motion-compensated temporally neighboring pixels, i.e. the corresponding pixels of the lost block in the previous frame with movement taken account for is used as estimation.

The two estimators are trained using a video training database with a total of 1400000 realizations. Each realization is a pair of two blocks from two consecutive frames. The block corresponding to the earlier frame in a realization is motion-compensated so that the squared difference between the two blocks are minimized. This is not ideal in the actual concealment process since the MV of a lost block only can be estimated (in our case a lost block also results in a lost MV for that block). However this simplification is assumed to be of little significance and is not further investigated. The location of the pair of blocks within the two frames is random and independent with uniform distribution.
3.3 Packet loss concealment in the H.264 JM reference software

The first concealment scheme for the MMLS-based method to be compared against is the PLC scheme implemented in the JM reference software, version 14.0. It is not a part of the H.264 standard itself, but a method implemented to test the standard in error-prone environments, and to be used as comparison for any new PLC scheme [24]. The method was included in the reference software because of its simplicity but has shown good performance especially for scenes with static objects such as backgrounds [13]. For a more detailed description of the method see [24] and [13].

The scheme chooses one of two methods depending on what type of frame to conceal. If the frame is intra coded (I-frame) the scheme estimates the lost blocks with a spatial interpolation algorithm and if the frame is inter coded (P-frame) a temporal approach is taken. In this thesis however we are only concerned with the predictively encoded frames and thus exclude the spatial algorithm from this description.

3.3.1 Temporal error concealment

The temporal error concealment is focused on finding estimates for the lost MVs to the MBs. When a MV estimate is found, the pixels are estimated as the pixels which the MV points to in the reference frame. This concealment concept is known as motion copy concealment or motion copying. If the average MV calculated over the entire frame, is below a threshold $T$, the zero MV is used as estimate, resulting in copy concealment of the entire frame. If the average MV is equal to or above the threshold $T$, each lost MV is estimated as the available spatially surrounding MV that minimizes the side match distortion $d_{sm}$, when used with motion copying. The side match distortion is the sum of pixel differences over the block borders for the luminance component of the estimated MB. The side match distortion measure is shown in equation (3.12), where $\hat{Y}(mv_{dir})_{IN}$ is the j-th concealed $Y$ value in the IN-blocks using the MV from the spatially neighboring MB in direction $dir$, and $Y_{OUT}^{j}$ is the j-th reconstructed $Y$ value in the OUT-blocks, and $N$ is the total number of the calculated boundary pixels. Concealed neighbor blocks are only included in the side match distortion measure if no correctly received neighbor blocks exist. The IN and OUT-pixels are illustrated in figure 3.2. The candidate MVs are those of the adjacent blocks together with the zero MV.

$$\begin{align*}
    \text{dir}^* = \min_{\text{dir}\in\{\text{top,bottom,left,right,zero}\}} \frac{1}{N} \sum_{j=1}^{N} |\hat{Y}(mv_{dir})_{IN}^{j} - Y_{OUT}^{j}| 
\end{align*}
$$

3.3.2 Parameters

The threshold parameter $T$ is set to 4, which preliminary investigations using the video training database has shown to be a good value.
3.4 Q.-F. Zhu’s spatio-temporal packet loss concealment scheme

Q.-F. Zhu’s spatio-temporal PLC scheme based on the paper *Coding and Cell-Loss Recovery in DCT-Based Packet Video* [27] is the second method for comparison with the MMLS-based scheme. The scheme was originally intended for applications where a part of the transform coefficients in a block may be lost, but our situation where all the coefficients of a block are lost is treated as a special case. The method has shown good performance compared to more recent and complex PLC schemes [16].

### 3.4.1 Theory

Let \( f \) be a vector consisting of the pixels from an image block and \( f_p \) be the corresponding motion-compensated block from a previous frame represented as a vector. Then let the residual be \( e = f - f_p \), which the encoder transmits after transformation and further compression. When a block is lost we have to estimate the residual \( e \) by \( \hat{e} \) which gives an estimate \( \hat{f} = \hat{e} + f_p \) for the original pixels \( f \). To calculate the estimate \( \hat{e} \) we introduce a smoothness measure

\[
\psi(\hat{e}) = \frac{1}{2} \left[ w \left( \| S_w \hat{f} - b_w \|^2 + \| S_e \hat{f} - b_e \|^2 + \| S_n \hat{f} - b_n \|^2 \right) + (1 - w)\hat{e}^T \hat{e} \right]
\]

The first term of the last line of (3.13) measures the spatial smoothness and the second term measures the temporal smoothness. A weighting factor \( w \) is introduced to control the relative contribution from the spatial part and the temporal part. The matrices \( S_w, S_e, S_n \) and \( S_s \) depend on the amount of smoothing to impose between every two pixels in the west, east, north and south directions. The vectors \( b_w, b_e, b_n \) and \( b_s \) consist of the one pixel wide border in the four directions outside the block to be concealed.

The optimum for \( \hat{e} \) is found when \( \frac{\partial \psi}{\partial \hat{e}} = 0 \), i.e. when the smoothing measure is minimized. Solving the equation for \( \hat{e} \) gives

\[
\hat{e} = \left( \frac{1}{w} I + S \right)^{-1} \left( b - S f_p \right).
\]

An iterative approach is used if any adjacent block is lost. The iteration begins with initializing the lost pixels with a "middle" value (gray), then iterates over the lost blocks in the frame, using the reconstructed pixel values from the previous iteration as boundary pixels in the current iteration. If a lost block is adjacent to a border, so that some boundary pixels do not exist, the method falls back to motion copy concealment.
3.4.2 Parameters

The matrices $S_w, S_e, S_n$ and $S_s$ are chosen as suggested in [23], resulting in that a smoothing constraint is imposed between pixels in the direction towards its nearest boundary. For a more detailed description of the structure of the matrices, we refer to [23]. The block size to be concealed is set to 16 × 16 pixels for the luminance component and 8 × 8 pixels for the chroma components, thus separate matrices are stored for luminance and chroma. The matrices are sparse such that the matrix multiplications in (3.14) can be computed rapidly. The weight factor $w$ is determined by maximizing the PSNR for the video training database, giving a value of $w = 0.05$. The number of iterations when adjacent blocks are lost is set to 20, which simulations in [27] has found to yield satisfactory results.

3.5 W. Zhu’s spatial packet loss concealment scheme

The purely spatial method in “Second-Order Derivative-Based Smoothness Measure for Error Concealment in DCT-Based Codecs” [28] is chosen as a third method for comparison to the MMLS-based scheme. This third method is targeted towards recovering of lost or erroneous transform coefficients, both for image and video communications. As with the spatio-temporal scheme in section 3.4, our case is equal to losing all transform coefficients and thus all equations below have been modified to not include any transformations.

3.5.1 Theory

For an image block $B$ with $m \times m$ pixels and where pixel $(i, j)$ has value $f(i,j)$ we can construct a quadratic variation measure

$$
\Psi_1(f) = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} (f(i-1,j) - 2f(i,j) + f(i+1,j))^2
+ \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} (f(i,j-1) - 2f(i,j) + f(i,j+1))^2
+ 2 \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} (f(i,j) - f(i+1,j) - f(i,j+1) + f(i,j+1))^2,
$$

(3.15)

where the expression includes second order derivative in the horizontal direction and the vertical direction as well as a cross term.

To emphasize the stronger smoothing required by horizontal and vertical edges a Laplacian-based measure is also defined

$$
\Psi_2 = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} (f(i+1,j) + f(i-1,j) + f(i,j+1) + f(i,j-1) - 4f(i,j))^2.
$$

(3.16)

The two measures are combined to form a hybrid measure

$$
\Psi = \Psi_1 + \lambda \Psi_2,
$$

(3.17)
where $\lambda$ is a balance parameter that controls their relative contribution.

Further let $\hat{f}(i,j)$ represent the estimated pixels for a lost block $B$ that we wish to conceal. The summation in (3.17) includes boundary pixels outside the actual block. These pixel values are assumed to be known and are denoted by $\hat{f}(i,j) = b_{i,j}$ (3.18)

where $b_{i,j}$ represent the boundary pixels.

We write (3.17) in matrix-vector representation

$$\Psi(f) = \|A_h f - b_h\|^2 + \|A_v f - b_v\|^2 + 2\|A_d f - b_d\|^2 + \lambda\|A_L f - b_L\|^2$$  (3.19)

where the matrices $A_h, A_v$ and $A_d$ represent the differential operations in the quadratic variation measure and $A_L$ represent the Laplacian operation. The vectors $b_h, b_v, b_d$ and $b_L$ consist of zeros except at the locations corresponding to boundary pixels.

The optimal solution for $\hat{f}$ can then be found when $\frac{\delta \Psi(f)}{\delta f} = 0$ [23], which gives us

$$\hat{f} = A^{-1}b,$$  (3.20)

where

$$A = A_h^T A_h + A_v^T A_v + 2A_d^T A_d + \lambda A_L^T A_L,$$

$$b = A_h^T b_h + A_v^T b_v + 2A_d^T b_d + \lambda A_L^T b_L.$$  (3.21)

If the border pixels are unavailable the pixels are initialized to a “mid-level” (gray) value and an iterative solution is used, where the border values in the current iteration are the values found in the previous iteration.

### 3.5.2 Parameters

The lost pixels are concealed in blocks of $16 \times 16$ luminance components and their corresponding $8 \times 8$ chroma components. The balance factor $\lambda$ is selected by maximizing PSNR for the video training database, and is found to be optimal for $\lambda = 2.5$. The number of iterations, when border pixels are also lost, is set to 20.

### 3.6 Common properties and strategies

The order in which the lost blocks of a frame are concealed, is not defined for the spatio-temporal method, the MMLS-based method and the spatial method. Thus a common concealment order for these methods is proposed and implemented in this thesis, and described in section 3.6.1. The technique for concealing entirely lost frames, which is implemented in the reference software, and used when necessary for all the schemes, is presented in section 3.6.2. Further a more sophisticated method, for MV estimation, than median MV and average MV is adopted for the MMLS-based scheme and the spatio-temporal scheme. This method is the one used by the JM concealment scheme and described in section 3.6.3. Special cases and scenarios are covered in section 3.6.4.
3.6.1 Concealment block order

The order in which the lost blocks of a frame are concealed is important for the quality of the estimation as a whole. One obvious reason is that correctly received pixels are more reliable as prediction source in the concealment process, than previously estimated pixels. Another reason is that estimated pixels will in many cases be used for further prediction, due to lack of surrounding correctly received pixels, and thus any concealment mistakes may propagate through out the frame [24]. One can further argue that central blocks of the frame are "difficult" to conceal, with discontinuous motion and large prediction error, and therefore should be concealed in the end of the concealment process to prevent errors from propagating to the "easier" parts of the frame [24].

Where one frame is to be concealed, all lost blocks of $8 \times 8$ pixels are searched for in raster scan order, from left to right and downwards. The block size of $8 \times 8$ pixels is imposed by the current implementation of the reference software. It makes use of a status map for all $8 \times 8$ blocks in the frame telling whether the block is received, concealed or lost. When the first lost block is found a new search is initiated to find the continuous area of lost blocks. This search is done by finding the next row with a correctly received block in the same column as the lost block and the next column with a correctly received block in the same row as the lost block. The search for this continuous area is straight forward but may not be optimal for all cases, though early tests indicate that it works very well for our purposes and codec settings.

A continuous area of lost blocks is estimated from the correctly received areas, by alternating first row-wise towards the middle of the lost area then column-wise towards the middle of the lost area. See figure 3.3 for an illustration of the concealment order. If the continuous lost area is adjacent to one or more borders the concealment process works towards the borders rather than the middle of the area.

This concealment order is based on [24] with some minor simplifications. All methods except the JM, which is based on the exact description in [24], use the concealment order described above.

3.6.2 Concealment of entire frame losses

When multiple consecutive packets are dropped the situation where an entire frame is lost may occur. Since all the schemes presented in this thesis needs to have some available spatial information to be applied, a method already implemented in the reference software is used to handle such situations.

The method is proposed in [6] and is executed in two steps. First the MVs is copied from the last decoded frame and then motion-compensation is performed to conceal the lost frame.

3.6.3 Motion vector estimation

Since all MVs are lost together with their corresponding MB a technique for estimating the MVs is required. The MMLS-based scheme as presented in [16] uses the
median of the surrounding MVs as an estimate and the spatio-temporal scheme as presented in [27] uses the average of the surrounding MVs as an estimate. Those are simple and straightforward estimations of the MV, but as stated in [24], median or averaging does not give better results than the MV estimation included in the JM reference software. Based on this statement the MV estimation technique used in the JM reference software and presented in section 3.3.1 is adopted for both the MMLS-based scheme and the spatio-temporal method.

3.6.4 Special cases

Preliminary investigations show that using copy concealment, by copying spatially corresponding pixels from the previously decoded frame into the lost blocks, when the average frame motion is low can improve PSNR. This modification is suggested in [24] and implemented in the JM concealment scheme. A threshold parameter $T$ (same parameter as for the concealment scheme in JM) is thus introduced for both the MMLS-based scheme and the spatio-temporal scheme. If the average frame motion is below $T = 4$, regular copy concealment will be performed.

Also motion copy concealment is used in all special cases for all methods where the concealment scheme in question can not be applied. For example when blocks to be concealed are adjacent to a border. This approach includes the spatial method as well.
Figure 3.1. Illustration of the variables used by the two estimators.
Figure 3.2. Illustration of the IN and OUT pixels in the side match distortion measure for the MV estimation.

Figure 3.3. Illustration of the concealment order.
Chapter 4

Evaluation

The following chapter presents the results from the objective and subjective evaluation of the PLC schemes treated in this thesis, and the settings and packet loss models under which the evaluation is preformed. An objective measure is introduced, and a study which compares the different schemes is presented together with observations of the concealment processes. Complexity analysis is not a part of the evaluation but some rough estimates based on calculation times can be found in Appendix C.

4.1 Packet loss models

To evaluate the PLC schemes we need models to simulate packet losses in a controlled and repeatable manner. In [7] Bolot reports that packet losses are independent if only a small amount of the available bandwidth is used. In [15] however Paxson concludes that the probability of a lost packet is considerably higher if the previous packet was lost than if it was correctly received. As a result two models are adopted in this thesis to evaluate the PLC schemes, an independent model and a simple Gilbert model. In both models, loss rate refers to the expected number of lost packets divided by the number of sent packets.

4.1.1 Independent packet loss model

The independent model simulates packet loss by dropping packets at random independently of each other. It is easily implemented and is commonly used, see for example [16] and [8].

4.1.2 Simple Gilbert model

The Gilbert model is a 2-state Markov process used to model bursty packet loss scenarios. The two states, called good state (G) and bad state (B), have different probabilities for independently causing a dropped packet upon arrival to the
state. In the simple Gilbert model however, which we will use in this thesis, the probability for a dropped packet in state (G) is zero and in state (B) is one.

The transition probabilities for the process is derived from the steady state error probability $\pi_B$ and a burstiness parameter denoted by $k$. The burstiness parameter $k$ is defined as the quota between the probability of leaving state (G) and staying in state (B). The state diagram for the model is shown in figure 4.1 and the relation between the steady state error probability $\pi_B$, the burstiness parameter $k$ and the transition probability $p$ is shown in equation (4.1).

$$p = \frac{\pi_B}{1 - \pi_B + k\pi_B}.$$  (4.1)

The same error pattern is used for all video sequences and all concealment schemes, to make sure the evaluation is fair.

4.2 Encoder settings

The H.264 standard gives the possibility to control a large set of parameters. A short description of the values of the most important parameters used in our simulations are presented in Appendix D. Only P-frames are used, except for the very first frame, and the reference frame is always the previously encoded frame.

Three packetization techniques of slices and slice groups are considered:

**Slices** Two slices but no slice groups. The first slice contains the MBs of the upper part of the frame and the second slice contains the MBs of the lower part of the frame. See figure 2.1(a).

**Interleaved** Two slice groups and one slice in each group. The MBs are interleaved into the two slice groups such that the first slice group contains the slice with MBs from even MB-rows and the other slice group contains the slice with MBs from the odd MB-rows. See figure 2.1(b).

**Dispersed** Two slice groups and one slice in each group. The MBs in the two slices are dispersed throughout the frame in a checkerboard like pattern. See figure 2.1(c).
Each slice is put into one IP-packet (indirectly via a protocol called RTP), and it contains its MVs, the residuals of its MBs and additional parameters (like encoding modes etc) specific to the slice.

4.3 Objective evaluation

The objective evaluation is measured in peak signal-to-noise ratio (PSNR) in decibel scale for the luminance component

$$\text{PSNR}_{dB} = 10 \log_{10} \frac{255^2}{\text{MSE}},$$

where MSE is the mean squared error between the original samples and the decoded samples, and 255 is the maximum pixel value due to eight bit quantization resolution. It is the most commonly used objective measure even though it does not always correlate well with other subjective video quality measures [18].

4.3.1 Video sequences

As simulation data a video evaluation database consisting of 88 randomly chosen video sequences from [1], with an average of 385 frames per sequence, is built.

4.3.2 Results with independent packet loss model

The measurements for the independent packet loss model are evaluated for 1%, 5%, 10%, 20% and 30% loss rates. The results are presented in figures 4.2-4.4 and tables E.1-E.3 in appendix E.

From the figures it is apparent that the performance for the purely spatial method is poor for all packetization settings and loss rates compared to the other concealment schemes. The proposed MMLS-based scheme has the highest PSNR for the dispersed and the interleaved mapping, and the spatio-temporal method has the highest PSNR for regular slices with the MMLS-based method at second place.

4.3.3 Results with simple Gilbert packet loss model

The measurements for the gilbert packet loss model are evaluated for 1%, 5%, 10%, 20% and 30% loss rates with a burstiness factor, $k$, of seven. The results are presented in figures 4.5-4.7 and tables E.4-E.6 in appendix E.

The relative positions of the graphs are the same as with the independent packet loss model. With the MMLS-based method in first place, for the dispersed and interleaved mapping, and at second place for regular slices, after the spatio-temporal method.
4.4 Subjective evaluation

The subjective evaluation is performed with the purpose of finding differences, in estimation artifacts, between the concealment schemes, and for internal grading. It is not the purpose to calculate a comparable score, defined by a video testing standard. The evaluation is achieved in two parts. By a study with a forced two choice test method, with a group of test persons, and by carefully observing restored video sequences with different content and try to identify artifacts.

4.4.1 Video sequences

Three video sequences are analyzed and used in the subjective evaluation. They are chosen to represent three different types of content and scenarios. The sequences are commonly used in the video coding society and can be downloaded from [2], [3], [4]. The sequences are in color but only the luminance component are presented in the figures in this chapter. The video sequences, together with brief descriptions of their distinguish characteristics are given below:

Paris A video sequence of two persons sitting in front of a table, recorded with a stationary camera. The two persons are having a friendly discussion moving
4.4 Subjective evaluation

Figure 4.3. Performance of the different PLC schemes for varying loss rates and with interleaved slice groups and independent packet loss model.

slightly and "working" with their hands. The background is static through the entire sequence.

Soccer A video sequence of people playing soccer, though it is not a regular broadcasted recording. The camera has a static position but is panning horizontally and zooming out.

City+Mobile (C+M) A combination of two video sequences resulting in a change of scene in the middle of the sequence. The first part is a recording of a city from the air and the second part is of a moving calendar and a moving toy train. Both parts include a lot of details.

4.4.2 Forced two choice test study

A forced two choice test is "invented" for this thesis with the above mentioned goals in mind. It is based on the standardized tests defined in *ITU-R Recommendation 500-11* [12] but does not claim to follow all its considerations and principles.

A test is executed by showing the same erroneous video sequence concealed with two different methods to a test person. The two methods, called method A and method B, are shown one after another, first A then B and then A again and
then B again and after that the test person is "forced" to chose one of the methods as the preferred one. The video sequence is approximately 10 seconds long with a pause of 1 second between every presentation of the methods under test.

Three test movies, one loss pattern of 5% independent packet loss and two packetization techniques are used, giving a total of 6 test video sequences. Every test video sequence is used to compare every combination of two concealment methods, giving a total of 36 combinations of sequences and concealment methods to test. The tests are generated and put in a randomized list. For every test person ten tests are performed, with a total duration of approximately ten minutes per person. Each test is picked and removed from the list of tests and if the list becomes empty a new one is created as described above.

### 4.4.3 Study results

The results of the study is presented in table 4.1 as a tournament score table, where every "win" is rewarded with one point. The result is also illustrated in figure 4.8 as % wins over the corresponding tests. 95%-confidence intervals are included, with the assumption that the % wins are normally distributed. A total of nine test persons has participated in the test.
4.4 Subjective evaluation

Figure 4.5. Performance of the different PLC schemes for varying loss rates, simple Gilbert model and with dispersed slice groups.

The MMLS-based method has a few more wins in total than the JM method and the spatio-temporal method, which only differ with one point. The spatial method is outperformed with only two wins in total. It is with the dispersed mapping that the MMLS-based is clearly better according to the study, for regular slices there is no apparent difference.

<table>
<thead>
<tr>
<th>Method</th>
<th>Tests</th>
<th>Paris</th>
<th>Soccer</th>
<th>C+M</th>
<th>Paris</th>
<th>Soccer</th>
<th>C+M</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JM</td>
<td>45</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>MMLS</td>
<td>43</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td>Spatio-temp.</td>
<td>46</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>28</td>
</tr>
<tr>
<td>Spatial</td>
<td>46</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.1. Scores for the PLC schemes for the different video sequences.
4.4.4 Observations

By carefully observing the video sequences a couple of differences between the concealment schemes can be noted. All the observations are presented with 5% independent packet loss and in grayscale.

The most protruding effect is observed for the spatial method which gives a very blurry image with dispersed mapping, figure 4.9(a), and a gray layer over the lost blocks for regular slices, figure 4.9(b). The entire frames are blurred or look gray due to errors propagating through the sequence. The good quality at the borders are an effect of the motion copy concealment of the border blocks.

In the City+Mobile sequence, with dispersed mapping, we can observe a difference in the concealment of the blocks lost right after the scene change. See figure 4.10 for a visual comparison. The spatial method in figure 4.10(e) gives the blurry effect over the lost blocks as noted above, for the MMLS-based method in figure 4.10(c) we can see the remains of the previous scene but blurred to some extent and the JM and spatio-temporal method in figure 4.10(d) and 4.10(f) looks like a combination between the current frame and the previous frame.

In scenes with fast motion, such as in the soccer sequence, we can observe blocking artifacts. The effect is most visible for the JM and the spatio-temporal method, see figure 4.11(a) and 4.11(b), with obvious edges between the blocks.
4.4 Subjective evaluation

![Figure 4.7. Performance of the different PLC schemes for varying loss rates, simple Gilbert model and with half frame slices.](image)

The MMLS-based method, see figure 4.11(c), has smoother transitions between the block borders.

One effect that is barely visual in the video sequences, and can not be illustrated here, is a blurring effect of the background. It occurs over time for static backgrounds in a sequence concealed with the MMLS-based and spatio-temporal method when the threshold parameter $T$ is small enough for the concealment to never or seldom fallback to copy concealment. The effect is more distinct for the MMLS-based scheme than for the spatio-temporal method. It is also observed that when the $T$ parameter is used, and results in copy concealment, small but fast moving objects will give artifacts.
**Evaluation**

Figure 4.8. Bar graph over the study results in terms of % wins of the corresponding tests. 95%-confidence intervals are calculated with the assumption that the % wins are normally distributed.

Figure 4.9. Paris sequence with dispersed mapping (a) and slices (b) concealed with the spatial method.
Figure 4.10. Visual comparison of the concealment methods with a packet loss right after a scene change in the sequence. The packetization technique is dispersed mapping and the sequence is C+M.
Figure 4.11. Visual comparison of a concealed sequence with fast motion using dispersed slice group mapping. Differences in results can be seen in the circumscribed area.
Chapter 5

Conclusions

This thesis evaluates the MMLS-based scheme by comparisons with three well known concealment schemes. The comparisons are evaluated by means of PSNR, observations of artifacts and a study with human observers. The schemes are implemented in the H.264 JM reference codec.

For the more sophisticated packetization techniques, namely dispersed and interleaved mapping, the objective measures and the observations show an overall higher performance and better video quality for the MMLS-based scheme than for the competitors, see figures 4.2, 4.3, 4.5, 4.6, 4.10 and 4.11. The result is also verified by the study in the case of dispersed mapping, see table 4.1. Without slice groups however, when the frames are divided into two half’s and sent in separate packets, the spatio-temporal method shows the highest PSNR values, but with the MMLS-based scheme at a second place.

The model used in the MMLS-based scheme is trained for the case when the surrounding pixels are reliable, which can be considered a good model when interleaved or dispersed slice groups are used, but a bad model without slice groups. Thus it can be explained why the MMLS-based scheme has worse performance without slice groups and one can argue that the method is less robust than the competitors because of the importance of the training process.

From the observations of the MMLS-based scheme it is shown that the method has the ability to classify the situation and emphasize temporal or spatial concealment accordingly, see figure 4.10. The study results for the C+M video sequence with dispersed mapping confirms this observation, see table 4.1. The observation together with the study results indicates that the MMLS-based scheme is a good choice of method when the content of a sequence changes rapidly between the frames.

Further the observations also indicate that the MMLS-based scheme works better than the other methods when concealing sequences with fast motion, see figure 4.11. This conclusions are however not confirmed by the study results and can thus be said to be less significant.

The most obvious conclusion is that the purely spatial method does not work very well under the conditions and scenarios evaluated in this thesis, see figures
4.2-4.7. The method can not restore details within a lost area of pixels, resulting in a blurry image. Only one situation can be identified where the method gives a comparable result, and that is when concealing a lost packet corresponding to the first packet after a change of scene in a sequence, see figure 4.10.

The overall impression is that the MMLS-based scheme performs equally good or better than the competitors in most cases when the dispersed mapping is used. However it is not a significant result, and perhaps other factors like packetization technique versus bit rate and implementation complexity should have a greater impact depending on the application.

5.1 Future work

Implementation of concealment schemes in the JM reference software is somewhat error prone due to lack of comments, global and unused variables, and one has to deal with the limitations in the software or spend valuable time to work around them. The advantage of using the JM reference software is that the results are true, in the sense that the context is realistic in real world applications. A good contribution to further concealment research would be an implementation of a common test framework, including different channel models, loss scenarios, packetization techniques etc. Such a framework, in conjunction with common video evaluation databases, would build a good basis for practically relevant results and open up for better comparison between different concealment schemes.

The GMM-based scheme and the closely related MMLS-based scheme both show good performance. It would be interesting to combine these good statistical models with some kind of adaptivity, either by adapting the parameters of the models or by image processing techniques which can choose between a larger set of models.

The estimation of the MV is critical, as concluded in preliminary investigations for this thesis, and thus further research on a combined approach to the pixel and MV estimation problem could be of interest.
Chapter 6

A new packet loss concealment scheme

A new PLC method is investigated by considering a joint model for MVs and pixels. The model is using GMMs for both pixels and MVs.

Let a group of lost pixels be denoted by the stochastic vector $x$ and the surrounding available pixels, both temporally and spatially, be denoted by the stochastic vector $y$. Further let the corresponding MV to $x$ be denoted by the stochastic vector $v$ and the spatially surrounding available MVs to $v$ be denoted by the stochastic vector $w$. The joint model is then defined by the pdf $p(x, y, v, w)$.

We form an estimator for the lost MV as

$$v^* = \arg \max_v p(x, y, v, w), \quad (6.1)$$

where $x$ is subject to

$$x = g(y, v, w), \quad (6.2)$$

which is the GMM-based pixel estimator presented in [17]. We further assume partial independence for the pdf and write it as

$$p(x, y, v, w) = p_1(x, y_{\text{spatial}})p_2(v, w), \quad (6.3)$$

where $y_{\text{spatial}}$ consists of the spatially surrounding pixels to $x$ from $y$. The two underlying pdf:s, $p_1$ and $p_2$, are modeled by two GMMs. The expression in (6.1) is maximized through testing all offsets $(dx, dy), \forall dx, \forall dy \in \{1, 0, -1\}$ to the previous estimate of $v$, which is initialized as the median of the surrounding MVs.

In the simulations the maximization process is repeated four times for every frame and the concealment size of $x$, as well as the block size for motion compensation is set to $8 \times 8$ pixels. Due to complexity concerns we simplify the pixel model to work with $4 \times 4$ pixels and repeat the estimation and pdf calculation four times, to obtain $x = g(y, v, w)$ and $p_1(x, y_{\text{spatial}})$, until all $8 \times 8$ pixels have been covered. Packet loss is generated by using a simple Gilbert model with expected error length of 20, which is equal to $k = 361$ (the burstiness parameter), and an
error rate of 5%. A lost packet corresponds to a lost block of $8 \times 8$ pixels, and the errors are introduced after the decoding process so that error propagation may not occur.

The GMMs are trained with the EM-algorithm [9] and the video training database. The evaluation is performed with the video evaluation database. Regular motion copying and a GMM-based approach for $4 \times 4$ pixels without MV-searching are used for comparison.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motion copying</td>
<td>43.67</td>
</tr>
<tr>
<td>GMM</td>
<td>44.19</td>
</tr>
<tr>
<td>Proposed method</td>
<td>43.96</td>
</tr>
</tbody>
</table>

Table 6.1. PSNR results i dB.

We can conclude that the new PLC scheme, built on a joint model, does not improve performance compared to the GMM-based estimator. By carefully looking at the algorithm while concealing lost pixels, we can note that the MV model can not properly decide which is the best MV when the surrounding MVs are non-zero. The reason could possibly be that the MV model is inaccurate, due to the appearance of the distribution for MVs, where most MVs are zero. No further investigations are performed.
Bibliography


Appendix A

Getting started with the JM reference software

This short guide will give a crash course in how to encode raw video sequences, drop packets and decode the disturbed video stream. The original reference software can be found at: http://iphome.hhi.de/suehring/tml/ which differs slightly from the modified version, which is a result from this thesis.

The reference software is constituted of essentially two parts, the encoder and the decoder but does also contain a simple application for simulation of packet loss. They all have separate binaries named: lencod.exe, ldecod.exe and rtp_loss.exe respectively. The encoder and decoder are controlled via input arguments and configuration files and the packet loss application is controlled solely via input arguments.

A.1 How to encode

The encoding process is started by the command: lencod.exe -d <configuration file> where the configuration file contains all the parameters necessary to encode the sequence. The application encodes a raw YUV-file (420p), which is specified in the configuration file. To create a configuration file it is recommended that one starts with a copy of the default configuration file, delivered with the software.

The following list will highlight the properties of the standard configuration file that needs to be changed in order for the encoding to work properly:

**InputFile:** Path to the raw video sequence to be encoded (YUV 420p format)

**FramesToBeEncoded:** The number of frames from the input sequence that should be encoded

**SourceWidth:** The horizontal resolution of the input sequence

**SourceHeight:** The vertical resolution of the input sequence
OutputFile: The encoded output file

The following list will highlight some of the interesting settings that can be changed depending on the desired type of encoding:

ProfileIDC: Change profile to give access to desired functionality (66=baseline, 77=main, 88=extended)

IntraPeriod: Intra-frame period, the number of P-frames between every I-frame

QPISlice/QPPSlice: Quantization step, great impact on the bit-rate

FrameSkip: Number of frames to skip from the input sequence between every encoded frame

SearchRange: Search range for the motion-compensation

NumberReferenceFrames: Number of previous encoded frames that can be used for motion-compensation

Log2MaxFNumMinus4: Should be set to -1 for the error concealment to work properly

SliceMode: Slice mode (0=off, 1=fixed #MBs in slice, 2=fixed #bytes in slice)

SliceArgument: Argument to slice mode (ex. if SliceMode=1 and SliceArgument=100 then 100 MBs will be put into every slice)

num_slice_groups_minus1: # of slice groups minus 1

slice_group_map_type: Set the type of slice group mapping to use, see configuration file for further instructions

OutFileMode: Output file mode (0:file format according to annex B of the H.264 standard, 1:RTP (RTP should be used if packet loss is to be applied))

A.2 How to drop packets

When the OutFileMode of the encoder is set to RTP (1), the output file will be a “stream” of RTP-packets. The packets within this file can be dropped independently with the following command: rtp_loss.exe <input file> <output file> <loss rate in %> [skip packets] (original reference software) or rtp_loss.exe <input file> <output file> indep <loss rate in %> [skip packets] (modified reference software). Where the last argument can be used to make sure that the specified number of first packets will not get dropped. Packets can also be dropped according to the simple Gilbert model with the command rtp_loss.exe <input file> <output file> gil <consecutive loss probability> <loss probability> [skip packets], with the modified version of the application.
A.3 How to decode

The decoding is initiated by the following command: `ldecod.exe <configuration file>`, where the configuration file contains the parameters used to decode a file. It is recommended that the file is created by copying the default decoder configuration file, distributed with the reference software. The configuration file does not contain any names for the parameters, instead each parameter is written on its own parameter line according to a predefined order which is documented in the default configuration file.

The first three parameters of the configuration file should be set according to the comments of the default configuration file. They are needed to specify the input file, the output file and an original raw file for SNR computation. The NAL mode parameter line should be set to 1 if RTP-packets has been used in the encoding process. The error concealment parameter line refers to the method which should be used for picture error concealment (when loss of entire frames occur) and should not be confused with the PLC method used when a part of the frame is correctly received. In the modified version of the software an additional parameter line commented as Error Concealment Method is introduced to switch between different concealment methods (0: (default) JM, 1: MMLS-based, 2: spatial method, 3: spatio-temporal method).

The output is a raw video sequence in the YUV420 format.
Appendix B

Source code manual

This source code manual gives pointers to the new files, the modified files, important functions and parameters which constitutes the new modified version of the JM reference software. The code is well commented and should be sufficient to describe the inner details of the implementation, in conjunction with this report.

The new and the changed files in the modified JM reference software are listed below, together with sublists over the important content. The source files are found in the subfolders `ldecod/inc` (header files) and `ldecod/src` (code files).

**erc_mmls.[h|c]** code, definitions and precalculated values for the MMLS-based concealment scheme.

- **ercMMLSConcealFrame()** entry function for MMLS-based concealment of a partly received frame
- **ercMMLSConcealLostFrame()** entry function for MMLS-based concealment of an entirely lost frame
- **concealByTrial2()** a modified version of the JM MV estimation function **concealByTrial()**

**erc_mmls_model.c** matrices and vectors $\rho, c$ and $A$, used by the MMLS-based estimators

**erc_orange.[h|c]** code for orange error concealment (debug purpose)

**erc_spatial.[h|c]** code for the spatial error concealment method

- **ercSpatialConcealFrame()** entry function for the spatial PLC method
- **erc_spatial_matrix.c** matrices $A_h, A_d, A_v, A_L$ and $A^{-1}$, used in the spatial PLC method

**erc_spatio_temp.[h|c]** code for the spatio-temporal PLC scheme

- **ercSpatialTemporalConcealFrame()** entry function for the spatio-temporal PLC scheme
erc_spatio_temp_matrix.c matrices use in the spatio-temporal PLC scheme

erc_api.h file from the original software which contains definitions of various data structures and assignment of parameters

MVPERMB_THR the threshold parameter (T) used to fallback to copy concealment when the average frame motion is low

image.c code file from the original software which has been modified to call additional PLC methods

exit_picture() this function decides which concealment method to use, if packet loss has occurred, and calls the corresponding method
Appendix C

Complexity measures

This section presents the calculation time for the four PLC schemes when concealing a 50 frames long sequence. The numbers are rough but give a good indication of the relative complexity. Two packetization settings are evaluated, regular slices and dispersed mapping, and packets are dropped independently with a rate of 25%.

<table>
<thead>
<tr>
<th>Method</th>
<th>Slices</th>
<th>Dispersed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>4.2</td>
<td>6.0</td>
</tr>
<tr>
<td>JM</td>
<td>2.3</td>
<td>2.9</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td>11.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Spatial</td>
<td>40.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Table C.1. Calculation time in seconds for the PLC schemes with a packet loss rate of 25%.

When regular slices are used the spatial and the spatio-temporal method need to iterate every lost frame 20 times each resulting in a long computation time, but for the dispersed mapping no iterations are needed.
Appendix D

Encoder settings

The baseline profile is used with a quantization step of 28. The search range for the motion-compensation is set to 12 pixels and subpixel precision is allowed. Also sub-macroblock motion-compensation is allowed but only the last encoded frame may be used as reference frame.
Appendix E

Objective evaluation results

<table>
<thead>
<tr>
<th>Packet loss rate</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>32.36</td>
<td>29.85</td>
<td>26.29</td>
<td>23.03</td>
<td>20.97</td>
</tr>
<tr>
<td>JM</td>
<td>31.59</td>
<td>28.83</td>
<td>25.35</td>
<td>21.78</td>
<td>19.68</td>
</tr>
<tr>
<td>Spatial</td>
<td>26.95</td>
<td>24.10</td>
<td>22.13</td>
<td>19.92</td>
<td>18.16</td>
</tr>
</tbody>
</table>

Table E.1. Performance of the different PLC schemes for varying loss rates and with dispersed slice groups and independent loss model.

<table>
<thead>
<tr>
<th>Packet loss rate</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>29.35</td>
<td>27.80</td>
<td>24.69</td>
<td>21.26</td>
<td>19.70</td>
</tr>
<tr>
<td>JM</td>
<td>31.90</td>
<td>27.17</td>
<td>24.07</td>
<td>20.62</td>
<td>19.04</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td>31.94</td>
<td>27.23</td>
<td>24.20</td>
<td>20.79</td>
<td>19.21</td>
</tr>
<tr>
<td>Spatial</td>
<td>25.90</td>
<td>22.16</td>
<td>18.09</td>
<td>15.30</td>
<td>14.97</td>
</tr>
</tbody>
</table>

Table E.2. Performance of the different PLC schemes for varying loss rates and with interleaved slice groups and independent loss model.

<table>
<thead>
<tr>
<th>Packet loss rate</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>29.91</td>
<td>27.23</td>
<td>24.00</td>
<td>20.63</td>
<td>18.86</td>
</tr>
<tr>
<td>JM</td>
<td>29.52</td>
<td>26.78</td>
<td>23.66</td>
<td>20.11</td>
<td>18.28</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td>30.64</td>
<td>27.84</td>
<td>24.31</td>
<td>20.71</td>
<td>18.73</td>
</tr>
<tr>
<td>Spatial</td>
<td>18.73</td>
<td>16.67</td>
<td>16.02</td>
<td>15.21</td>
<td>14.86</td>
</tr>
</tbody>
</table>

Table E.3. Performance of the different PLC schemes for varying loss rates and with half frame slices and independent loss model.
Table E.4. Performance of the different PLC schemes for varying loss rates, simple Gilbert model and with dispersed slice groups.

<table>
<thead>
<tr>
<th>Method</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>32.52</td>
<td>27.44</td>
<td>24.91</td>
<td>21.85</td>
<td>20.06</td>
</tr>
<tr>
<td>JM</td>
<td>31.74</td>
<td>26.78</td>
<td>24.51</td>
<td>21.28</td>
<td>19.79</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td>31.79</td>
<td>26.83</td>
<td>24.56</td>
<td>21.34</td>
<td>19.84</td>
</tr>
<tr>
<td>Spatial</td>
<td>27.16</td>
<td>23.75</td>
<td>22.68</td>
<td>20.43</td>
<td>19.51</td>
</tr>
</tbody>
</table>

Table E.5. Performance of the different PLC schemes for varying loss rates, simple Gilbert model and with interleaved slice groups.

<table>
<thead>
<tr>
<th>Method</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>32.56</td>
<td>26.30</td>
<td>23.28</td>
<td>21.03</td>
<td>19.50</td>
</tr>
<tr>
<td>JM</td>
<td>32.13</td>
<td>25.89</td>
<td>22.77</td>
<td>20.54</td>
<td>19.10</td>
</tr>
<tr>
<td>Spatio-temporal</td>
<td>32.17</td>
<td>25.92</td>
<td>22.85</td>
<td>20.68</td>
<td>19.26</td>
</tr>
<tr>
<td>Spatial</td>
<td>26.54</td>
<td>21.78</td>
<td>20.45</td>
<td>17.48</td>
<td>16.89</td>
</tr>
</tbody>
</table>

Table E.6. Performance of the different PLC schemes for varying loss rates, simple Gilbert model and with half frame slices/loss pattern.

<table>
<thead>
<tr>
<th>Method</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMLS</td>
<td>30.18</td>
<td>26.05</td>
<td>24.15</td>
<td>21.36</td>
<td>19.96</td>
</tr>
<tr>
<td>JM</td>
<td>29.73</td>
<td>25.67</td>
<td>23.90</td>
<td>21.14</td>
<td>19.75</td>
</tr>
<tr>
<td>Spatial</td>
<td>18.84</td>
<td>16.78</td>
<td>16.48</td>
<td>15.79</td>
<td>15.57</td>
</tr>
</tbody>
</table>
Upphovsrätt

Detta dokument hålls tillgängligt på Internet — eller dess framtida ersättare — under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för icke-kommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All anmänande användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

For ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

Copyright

The publishers will keep this document online on the Internet — or its possible replacement — for a period of 25 years from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for his/her own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/

© Henrik Cooke