Chap 2 MRF-Based ISCD for Uncompressed Video

% preface

Before introducing our proposed UEP scheme, the operation theory of MRF-based iterative joint source-channel decoder (ISCD) in \cite{3d\_mrf} works firstly is presented in this chapter. In the beginning, since this decoder is based on a MRF model, the basic concept of Markov random field (MRF) is introduced in section \ref{bb:3d\_mrf\_model}. Then, section \ref{bb:sisodecoder} shows the key component of MRF-based ISCD, the soft-in-soft-out (SISO) source decoder based on MRF model. Finally, the overall structure of MRF-based ISCD actually operates in a iterative process between SISO channel decoder and SISO source decoder. Thus, the structure diagram is depicted in section \ref{bb:3d\_mrf\_iscd} as well as the operation steps of such decoder.

%===========================================================

\section{3D-MRF Model}

\label{bb:3d\_mfr\_model}

The Markov random field (MRF) is a probability model, consisting of elements and the neighborhood system describing the relationship between those elements within the MRF. Due to the MRF's local properties and the multi-dimensional features, it is widely used to describe the image properties. For low-level MRF models, it is often used to solve the pixel-based problems such as image reconstruction, edge detection and texture synthesis. For high-level MRF models, object-based problems such as object detection, tracking and recognition is usually considered. Since the uncompressed video decoding is actually similar with typical image reconstruction problem when the video is seen as multiple images, the low-level MRF models are usually applied in the related problems. Among these low-level MRF models, the spatial relationship between the image pixels can be described by a kind of two-dimensional (2D) MRF model.

\cite{ISCD\_MRF} is a typical iterative source-channel decoding (ISCD) based on 2D-MRF model which relates a pixel to its spatial neighborhood pixels and models the image as a piecewise continuous surface. In \cite{ISCD\_MRF}, the frames are decoded by maximizing the {\textit a posteriori} probability of the 2D-MRF configuration in each iteration. Furthermore, \cite{3d\_mrf} proposes a 3D-MRF model which relates a bit to its spatial neighborhood bits in the same bit-plane as well as its adjacent bits in its temporal neighborhood system. The proposed temporal neighborhood system can exploit the temporal redundancy in video sequence well since the temporal redundancy has been a key component in video compression techniques.

%===========================================================

\subsection{Overview of Low-Level MRF Model}

\label{bbb:low\_level\_mrf}

The basic two elements of an MRF are sites and the labels corresponding to those sites. For example, in most image reconstruction MRF applications, a site usually refers to the position/index of a pixel, and the label is usually the corresponding pixel value in that position. For a site denoted as $i$, the corresponding label $q\_i$ which takes value in $\mathcal{Q}$ is assumed to be a realization of the random variable $q\_i$. Let $\mathit{S}$ be a set consisting of multiple sites with size $|\mathit{S|}$. The neighborhood system used to relate each site in $\mathit{S}$ to others is defined as

\begin{equation}

\mathit{N} = \{\mathit{N}\_i | \forall i \in \mathit{S}\},

\end{equation}

where $\mathit{N}\_i$ is defined as the set consisting of an arbitrary site $i$'s neighborhood sites with respect to the neighborhood system. Furthermore, take

\begin{equation}

Q = \{Q\_1, Q\_2,...,Q\_{|\mathit{S}|}\}

\end{equation}

as a random field consisting of all the labels of sites in $S$. A configuration $q = \{q\_1,q\_2,...q\_{|\mathit{S}|}\} \in \mathcal{Q}^m$ is a joint realization with respect to $Q$. Given a configuration $q$, the joint event $\{Q\_1=q\_1, Q\_2=q\_2,...,Q\_{|\mathit{S}|} = q\_{|\mathit{S}|}\}$ is denoted as $Q=q$, and the joint probability $P(Q=q)$ is simply represented by $P(q)$. $Q$ with the neighborhood system $\mathit{N}$ is said to be an MRF if and only if positivity and Markovianity appear in $F$. The positivity is defined by

\begin{equation}

P(q) > 0, \forall f \in \mathcal{Q},

\end{equation}

and the Markovianity says that

\begin{equation}

P(q\_i | q\_{\mathit{S} - \{i\}}) = P(q\_i|q\_{\mathit{N}\_i}),

\end{equation}

where $q\_{\mathit{S} - \{i\}}$ are the set of labels at all the sites in $S$ except the site $i$. The positivity as a global property implies that given an arbitrary joint realization (configuration) of $L$, there is an unique non-zero joint probability for this configuration. In other hands, Markovianity provides the local property of an MRF by saying that a site's label only depends on its neighborhood's labels. If a set of sites$\mathit{S}$, a neighborhood system $\mathit{N}$ and the conditional probability distribution $P(q\_i|q\_{N\_i})$ of any arbitrary site's label in $\mathit{S}$ are specified, an MRF is uniquely defined.

%===========================================================

\subsection{Neighborhood System and Conditional Distribution of 3D MRF Model}

\label{bbb:3d\_mrd\_model}

As mentioned above, in order to define the proposed 3D-MRF model in ~\cite{3d\_mrf}, a set of sites $\mathit{S}$, a neighborhood system $\mathit{N}$ and the conditional probability distribution $P(q\_i|q\_{N\_i})$ of any arbitrary site's label in $\mathit{S}$ should be specified. 3D-MRF model is a low-level MRF model, linking each bit of a transmitted video frame to its surrounding bits both in spatial and temporal direction. Let $\mathit{u}^{(x,y)}\_{k,n}$ denotes the $n$-th bit of a pixel located at $(x,y)$ in $k$-th frame of a video sequence. For simplicity, $\mathit{u}^{(x,y)}\_{k,n}$ is denoted as $\mathit{u}\_{i}$, where a single index $i$ comprises the indices $(k,n,x,y)$.

Different from conventional MRF model which usually takes a single pixel as a site ~\cite{ISCD\_MRF}, 3D-MRF sets a single bit $u\_{i}$ as a single site. Then the neighborhood sites of $u\_{i}$ can be separated into two subsets as

\begin{equation}

\mathit{N}\_{i}=\mathit{N}\_{i}^{[s]}\cup\mathit{N}\_{i}^{[t]},

\end{equation}

where $\mathit{N}\_{i}^{[s]}$ is the spatial neighborhood subset, and $\mathit{N}\_{i}^{[t]}$ is the temporal neighborhood subset. As illustrated in Fig.~\ref{fig\_3DMRF\_s}, only four nearby bits are chosen in the spatial domain due to the usually higher similarity to $u\_{i}$ at center, and thus the spatial neighborhood subset is defined as

\begin{equation}

u\_{\mathit{N}\_{i}^{[s]}}=\{ u\_{k,n}^{(x-1,y)}, u\_{k,n}^{(x+1,y)}, u\_{k,n}^{(x,y-1)}, u\_{k,n}^{(x,y+1)} \}.

\end{equation}

\begin{figure}[htb]

\centering

\includegraphics[width=\textwidth]{fig/mrf\_spatial\_v3.eps}

\caption{The spatial neighborhood system of 3D-MRF.}

\label{fig\_3DMRF\_s}

\end{figure}

Furthermore, temporal subset $\mathit{N}\_{i}^{[t]}$ takes only one single site in each reference frame as shown in Fig.~\ref{fig\_3DMRF\_t}. The site related to some site $u\_{i}$ is chosen through the motion vector from the decoder-side motion estimation (ME). Through local motion vector, a pixel $f\_k^{(x,y)}$ is related to another pixel $f\_{k'}^{(x',y')}$, where reference frame $k'$ is usually previous frame or next frame $k'=k\pm1$ for higher temporal redundancy. As a result, $n$-th bit in $f\_{k'}$ becomes the temporal neighborhood site of $n$-th bit in $f\_k$ in this MRF model, and thus the temporal neighborhood subset is

\begin{equation}

u\_{\mathit{N}\_{i}^{[t]}}=\{ u\_{k',n}^{(x',y')},\ \ x'=x-v\_i^x,\ \ y'=y-v\_i^y,\ \ k'\in\mathbb{F}\_k\},

\end{equation}

where $(v\_i^x,v\_i^y)$ is the local motion vector for $u\_i$, and $\mathbb{F}\_k$ is the set of reference frames for certain frame $k$.

\begin{figure}[htb]

\centering

\includegraphics[width=\textwidth]{fig/mrf\_temporal\_v3.eps}

\caption{The temporal neighborhood system of 3D-MRF.}

\label{fig\_3DMRF\_t}

\end{figure}

Finally, in order to specify the conditional distribution $P(u\_i,u\_{N\_i})$ of any arbitrary site's label in $\mathit{S}$, Gibbs distribution ~\cite{besag1974spatial} is utilized. A Gibbs distribution is defined as

\begin{equation} \label{eq\_Gibbs}

P(u\_i|u\_{N\_i}) = Z^{-1}\times e^{-U(u\_i,u\_{N\_i})},

\end{equation}

where

\begin{align}

\label{eq\_Gibbs\_Z}

Z &= \sum\_{u\_i \in \{0,1\}}e^{-U(u\_i, u\_{N\_i})}\\

&= e^{-U(u\_i=0,u\_{N\_i})}+ e^{-U(u\_i=1,u\_{N\_i})},

\end{align}

where $U(u\_i,u\_{N\_i})$ is the energy function. For a given site $u\_i$, the energy computed by the energy function $U(u\_i, u\_{N\_i})$ shows the local probability of the configuration $u$. In addition, since 3D-MRF model only considers the first-order neighborhood relationship (i.e., the pairs of sites where each pair is constructed by a site $i$ and one of its neighborhood $i' \in \mathit{N}\_i$), the energy function $U(u\_i,u\_{N\_i})$ can be expanded as

\begin{align}

\label{eq\_energy\_U}

U(u\_i, u\_{N\_i}) &= \sum\_{i' \in N\_i} V\_{c}(u\_i,u\_{i'})\\

&= \sum\_{ i' \in N\_i }V\_{c}^{[s]}(u\_i,u\_{i’})+\sum\_{ i' \in N\_i }V\_{c}^{[t]}(u\_i,u\_{i'}),

\end{align}

where $V\_{c}(u\_i,u\_{i'})$ is the potential function which computes the potential between $(u\_i, u\_{i'})$, and the form of potential function $V\_{c}(u\_i,u\_{i'})$ is the summation of both spatial and temporal neighborhood pairs’ potential. By assuming each bit-plane is a piecewise constant surface, the two potential functions take the form

\begin{align}

\label{eq\_spatial\_potential\_s}

V\_c^{[s]}(u\_i,u\_{i'})=\beta\_s(1-\delta(u\_i-u\_{i'}))\\

V\_c^{[t]}(u\_i,u\_{i'})=\beta\_t(1-\delta(u\_i-u\_{i'})),

\label{eq\_spatial\_potential\_t}

\end{align}

where $\delta(u)$ is the Dirac delta function, and $\delta(u)=1$ only when $u=0$. In other words, for each neighborhood pair $(u\_i,u\_{i'})$, $i'\in\mathit{N}\_i$, if $u\_i\neq u\_{i'}$, a penalty term is added to the energy function $U(u\_i,u\_{i'})$, where the penalty term is $\beta\_s$ and $\beta\_t$ according to whether $u\_{i'}$ is a spatial neighborhood site or a temporal one. Therefore, the energy function in (\ref{eq\_energy\_U}) can be written into an easily understandable form, which is

\begin{equation}\label{eq\_energy\_beta}

U(u\_i, u\_{N\_i}) = \beta\_s(\#\{u\_{N\_i^{[s]}}\neq u\_i\})+\beta\_t(\#\{u\_{N\_i^{[t]}}\neq u\_i\}),

\end{equation}

where $\#\{\cdot\}$ is the counting function.

The penalty coefficient $\beta\_s$ and $\beta\_t$ represent the homogeneity of the transmitted video sequence, where a high value of a penalty coefficient implies that there is a significant similarity between the neighborhood sites in 3D-MRF.

%===========================================================

\section{Overview of 3D-MRF based SISO decoding}

\label{bb:sisodecoder}

Based on the 3D-MRF model introduced above, the soft-in soft-out (SISO) source decoding for the uncompressed video can offer the extrinsic information according to the input \textit{apriori} information, which is usually provided by the SISO channel decoder. However, at the receiver side, the temporal neighborhood system and the MRF parameter $(\beta\_s,\beta\_t)$ is unknown. Therefore, it is necessary to estimate these essential elements at the decoder side with the received soft bits information.

For the temporal neighborhood system, decoder-side motion estimation is executed to locate the temporal neighborhood site of each site in the 3D-MRF model. This approach is to temporarily recover current frame and reference frame with their received soft information, and to implement the conventional block-based ME to obtain the temporal neighborhood system.

%==========================================================

\subsection{Parameter Estimation in Decoder}

\label{bbb:parameter\_est}

Since both the spatial and temporal neighborhood system can be constructed in decoder, the parameters $(\beta\_s,\beta\_t)$ of the potential functions is estimated in each bit-plane with the least-square fit solution first proposed by ~\cite{Gibbs\_par}. Consider the conditional probability $P(u\_i | u\_{N\_i})$ in (\ref{eq\_Gibbs}) where the left-hand side is modified by Bayes’ Theorem as

\begin{equation}\label{eq\_Gibbs\_bayes}

\frac{P(u\_i , u\_{N\_i})}{P(u\_{N\_i})} = \frac{e^{-U(u\_i,u\_{N\_i})}}{e^{-U(u\_i=0,u\_{N\_i})}+e^{-U(u\_i=1,u\_{N\_i})}}.

\end{equation}

Rearranging the above equation (\ref{eq\_Gibbs\_bayes}) as

\begin{equation}\label{eq\_Gibbs\_permute}

\frac{e^{-U(u\_i=0,u\_{N\_i})}+e^{-U(u\_i=1,u\_{N\_i})}}{P(u\_{N\_i})}=\frac{e^{-U(u\_i,u\_{N\_i})}}{P(u\_i , u\_{N\_i})}.

\end{equation}

Note that the left-hand side in (\ref{eq\_Gibbs\_permute}) is independent to whether $u\_i$ is 0 or 1. Therefore, substitute $u\_i=0$ and $u\_i=1$ into the right-hand side of (\ref{eq\_Gibbs\_permute}) and link them through the left-hand side of (\ref{eq\_Gibbs\_permute}), we have

\begin{equation}\label{eq\_Gibbs\_ind}

\frac{e^{-U(u\_i=0,u\_{N\_i})}}{P(u\_i=0,u\_{N\_i})}=\frac{e^{-U(u\_i=1,u\_{N\_i})}}{P(u\_i=1 , u\_{N\_i})}.

\end{equation}

Again we rearrange and take the logarithm of two sides in (\ref{eq\_Gibbs\_ind}) as

\begin{equation}\label{eq\_Gibbs\_ln}

\ln{\frac{e^{-U(u\_i=1,u\_{N\_i})}}{e^{-U(u\_i=0,u\_{N\_i})}}}=\ln{\frac{P(u\_i=1,u\_{N\_i})}{P(u\_i=0 , u\_{N\_i})}}.

\end{equation}

Substitute (\ref{eq\_energy\_beta}) into the left-hand side of (\ref{eq\_Gibbs\_ln}), and finally we have

\begin{equation}\label{eq\_Gibbs\_est}

\beta\_sx\_i^{[s]}+\beta\_tx\_i^{[t]}=y\_i,

\end{equation}

where

\begin{align}

x\_i^{[s]} & = \#\{u\_{N\_i^{[s]}}=1\}-\#\{u\_{N\_i^{[s]}}=0\} \\

x\_i^{[t]} & = \#\{u\_{N\_i^{[t]}}=1\}-\#\{u\_{N\_i^{[t]}}=0\} \\

y\_i &=\ln{\frac{P(u\_i=1,u\_{N\_i})}{P(u\_i=0, u\_{N\_i})}}=\ln{\frac{H(u\_i=1,u\_{N\_i})}{H(u\_i=0, u\_{N\_i})}},

\end{align}

where $H(u\_i,u\_{N\_i})$ is the histogram count of the local configuration $(u\_i,u\_{N\_i})$ occurs in one sample set. Once the local joint probability ratio $\frac{P(u\_i=1,u\_{N\_i})}{P(u\_i=0,u\_{N\_i})}$ is solved by histogram technique, $\beta\_s$ and $\beta\_t$ can be obtained from linear equation (\ref{eq\_Gibbs\_est}) as following

\begin{equation}\label{eq\_linear\_equation}

\begin{pmatrix}

x\_1^{[s]} & x\_1^{[t]} \\

x\_2^{[s]} & x\_2^{[t]} \\

\vdots & \vdots \\

x\_N^{[s]} & x\_N^{[t]} \\

\end{pmatrix}

\begin{pmatrix}

\beta\_s \\

\beta\_t

\end{pmatrix}

=

\begin{pmatrix}

y\_1 \\

y\_2 \\

\vdots \\

y\_N

\end{pmatrix},

\end{equation}

where $x\_i^{[s]}$, $ x\_i^{[t]} and y\_i$ are extracted from a sample set (i.e. a bit-plane). Note that (\ref{eq\_linear\_equation}) is an overdetermined system simply denoted as

\begin{equation}

\mathbf{X}

\begin{pmatrix}

\beta\_s \\

\beta\_t

\end{pmatrix}

= \mathbf{Y}.

\end{equation}

Thus, $\beta\_s$ and $\beta\_t$ can be estimated with a conventional least-square solution :

\begin{equation}

\begin{pmatrix}

\beta\_s \\

\beta\_t

\end{pmatrix}

= \mathbf{X}^{\dagger}\mathbf{Y},

\label{eq\_beta\_est}

\end{equation}

where $\mathbf{X}^{\dagger}$ is the Moore-Penrose pseudoinverse of matrix $\mathbf{X}$.

%===========================================================

\subsection{SISO source Decoder}

\label{bbb:siso\_s\_decoder}

The SISO source decoder based on 3D-MRF model executes in term of log likelihood ratio (LLR) of each bit in the transmitted video sequences. For a single site $u\_i$, the input {\it a priori} LLR of $u\_i$ is denoted as

\begin{equation}

L\_i^{(I)} = L\_s^{(I)}(u\_i) = \ln \frac{P(u\_i = 1)}{P(u\_i = 0)}.

\end{equation}

For each site, the {\it a posteriori} probability output can be represented by the {\it a posteriori} LLR

\begin{equation}

L\_i^{(O)} =\ln \frac{P(u\_i = 1 | L^{(I)}\_i, L^{(I)}\_{N\_i})}{P(u\_i = 0 | L^{(I)}\_i, L^{(I)}\_{N\_i})},

\label{eq\_LLR\_app}

\end{equation}

where $L\_{N\_i}^{(I)}$ is the set composed of input {\it a priori} LLRs of $N\_i$ which is the neighborhood of $u\_i$. From Bayes' theorem and the fact that $P(u\_i) = 1/2$ and $\ln(P(u\_i=0)/P(u\_i =1)) = 0$ when the probability is not conditional on the input {\it a priori} information $L^{(I)}\_i$, (\ref{eq\_LLR\_app}) can be modified as

\begin{align}

L\_i^{(O)} &= \ln \frac{P(u\_i = 1, L^{(I)}\_i, L^{(I)}\_{N\_i})}{P(u\_i = 0, L^{(I)}\_i, L^{(I)}\_{N\_i})}\nonumber \\

&= \ln \frac{P(L^{(I)}\_i | u\_i = 1, L^{(I)}\_{N\_i})}{P(L^{(I)}\_i | u\_i = 0 L^{(I)}\_{N\_i})} + \ln \frac{P(u\_i = 1 | L^{(I)}\_{N\_i})}{P(u\_i = 0 | L^{(I)}\_{N\_i})}.

\label{eq\_LLR\_app\_bayes}

\end{align}

Also, since that $L^{(I)}\_i$ and $L^{(I)}\_{N\_i}$ are conditionally independent when $u\_i$ is given, (\ref{eq\_LLR\_app\_bayes}) is further derived as

\begin{equation}

L\_i^{(O)} = \ln \frac{P(u\_i = 1| L^{(I)}\_i)}{P(u\_i = 0|L^{(I)}\_i)} + \ln \frac{P(u\_i = 1| L^{(I)}\_{N\_i})}{P(u\_i = 0|L^{(I)}\_{N\_i})},

\end{equation}

where the first term are simply $L\_i^{(I)}$ itself, and the second term is the probabilities conditional on $N\_i$ described from the 3D-MRF model. Because the uncertainty of $u\_{N\_i}$ in decoder, a novel solution which exploits only LLRs of each site with a low computation overhead is proposed in ~\cite{3d\_mrf}. The conditional probability of a site $u\_i$ derived in (\ref{eq\_Gibbs}) is modified based on the expectation of the energy function $U(u\_i,u\_{N\_i})$ over $L\_{N\_i}$ as

\begin{align}

P(u\_i|L\_{N\_i}) &= Z^{-1}\times e^{-E[U(u\_i, u\_{N\_i})| L\_{N\_i}]},

\label{eq\_app\_exp}

\end{align}

where $E[\cdot]$ is the expectation function based on a given probability distribution. From (\ref{eq\_energy\_U}-\ref{eq\_energy\_beta}), the expectation of energy function is

\begin{align}

\label{eq\_energy\_function\_exp}

E[U(u\_i, u\_{N\_i})| L\_{N\_i}] &= \sum\_{i' \in N\_i^{[s]}}E[V\_c^{[s]}(u\_i, u\_{i'})|L\_{i'}] + \sum\_{i' \in N\_i^{[t]}} E[V\_c^{[t]}(u\_i, u\_{i'})|L\_{i'}] \\

& = \sum\_{i' \in N\_i^{[s]}}\beta\_s (P(u\_{i'} \neq u\_i | L\_{i'}))+\sum\_{i' \in N\_i^{[t]}}\beta\_t (P(u\_{i'} \neq u\_i | L\_{i'})),

\label{eq\_energy\_function\_exp\_beta}

\end{align}

where $\beta\_s$ and $\beta\_t$ are the estimated 3D-MRF parameters in decoding.

By combining (\ref{eq\_LLR\_app})-(\ref{eq\_energy\_function\_exp\_beta}), the {\it a posteriori} LLR in (\ref{eq\_LLR\_app}) can be shown as

\begin{align}

L\_i^{(O)} = & \sum\_{i' \in N\_i^{[s]}} \beta\_s (P(u\_{i'} = 1 | L\_{i'}) - P(u\_{i'} = 0 | L\_{i'})) \nonumber \\

& + \sum\_{i' \in N\_i^{[t]}} \beta\_t(P(u\_{i'} = 1 | L\_{i'}) - P(u\_{i'} = 0 | L\_{i'})) + L\_i,

\label{eq\_app\_LLR\_final}

\end{align}

where the first term and second term correspond to the extrinsic information of spatial part and temporal part in 3D-MRF respectively, and the third term is simply the {\it a priori} information. Thus, we reformulate the 3D-MRF decoding output as

\begin{equation}\label{eq\_app\_LLR\_final\_short}

L\_i^{(O)} = L\_i^{(E,s)} + L\_i^{(E,t)} + L\_i^{(I)},

\end{equation}

where

\begin{align}\label{eq\_app\_LLR\_spatial}

L\_i^{(E,s)} = \sum\_{i' \in N\_i^{[s]}} \beta\_s (P(u\_{i'} = 1 | L\_{i'}) - P(u\_{i'} = 0 | L\_{i'})),\\

L\_i^{(E,t)} = \sum\_{i' \in N\_i^{[t]}} \beta\_t(P(u\_{i'} = 1 | L\_{i'}) - P(u\_{i'} = 0 | L\_{i'})),

\label{eq\_app\_LLR\_temporal}

\end{align}

which are the first term and second term in (\ref{eq\_app\_LLR\_final}). It shows that we can separately compute the extrinsic information $L\_i^{(E,s)}$ and $L\_i^{(E,t)}$ solely from the LLR of $u\_i$'s neighborhood without $L\_i$ involved, and the sum-up of two extrinsic information and the {\it a priori} information is the output of the decoder. Therefore, the SISO source decoder can be divided into two sub-decoder the spatial decoder and the temporal decoder, to compute the respective extrinsic information. Moreover, this source decoder can implement only one sub-decoder for less computation overhead depending on the characteristic of the video.

%===========================================================

\section{3D-MRF based ISCD structure}

\label{bb:3d\_mrf\_iscd}

An iterative joint source-channel decoding (ISCD) utilizes the SISO channel decoder, usually using BCJR algorithm ~\cite{BCJR} for convolutional channel encoding ~\cite{proakis}, and SISO source decoder at the same time, and iteratively exchanges the soft bit information to improve the correctness. The structure of ISCD is a turbo-like ~\cite{turbocode} process where in each iteration the soft output of the source decoder is fed back to the channel decoder as the updated {\it a priori} information. Through the iterative process, the mutual information between the decoders are supposed to increase and converge.

\begin{figure}[htb]

\centering

\includegraphics[width=\textwidth]{fig/block\_diagram\_s.pdf}

\caption{Spatial MRF based iterative source-channel decoder structure.}

\label{fig\_iscd\_s}

\end{figure}

Fig. \ref{fig\_iscd\_s} shows the decoder structure of ISCD with only spatial MRF source decoder. extrinsic information $L\_s^{(E,s)}$ from the output of the spatial source decoder is passed through the interleaver and is used as the input {\it a priori} information $L\_c^{(I)}$ for channel decoder in next iteration. Also, to prevent from the information overlapping, the channel decoder output LLR values $L\_c^{(O)}$ is subtracted by the input $L\_c^{(I)}$ and thus obtain the extrinsic information output of the channel decoder for source decoder, which is denoted by $L\_c^{E}$.

\begin{figure}[!hbt]

\centering

\includegraphics[scale=0.25]{fig/block\_diagram\_t.pdf}

\caption{Temporal MRF based iterative source-channel decoder structure.}

\label{fig\_iscd\_t}

\end{figure}

Likewise, Fig. \ref{fig\_iscd\_t} shows the implementation of the ISCD structure with only temporal MRF based source decoder. Since the temporal source decoding requires to utilize the extrinsic information $L\_s^{(E,t)}(\cdot)$ and temporal neighborhood sites $u\_{N\_i}$ passed from the reference (nearby) video frame, at least two frames would be decoded together. In Fig. \ref{fig\_iscd\_t}, the output of two channel decoders $L\_c^{(O)}(\cdot)$ are first used to perform motion estimation to locate the temporal neighborhood system, and then the extrinsic information $L\_c^{(E)}$ from the channel decoders are used to perform parameter estimation using the least square solution derived in Section. \ref{bbb:parameter\_est}. For the iterative process, the output extrinsic information of the temporal source decoder $L\_s^{(E,t)}$ is fed back to be the {\it a priori} information of the channel decoder. Furthermore, in this structure, an extra buffer to restore the temporal extrinsic information of the previous decoding frame, which is denoted by $L\_s^{(E,t’)}$, is added to propagate the temporal extrinsic information throughout the video decoding process. In other words, each frame can have two temporal extrinsic information source from both its previous and next frames.

\begin{figure}[!hbt]

\centering

\includegraphics[width=\textwidth]{fig/block\_diagram\_test.pdf}

\caption{3D-MRF based iterative source-channel decoder structure.}

\label{fig\_iscd\_all}

\end{figure}

Finally, combining the spatial and temporal source decoding according to (\ref{eq\_app\_LLR\_final\_short}), the overall 3D-MRF based iterative source-channel decoding scheme is shown in Fig. \ref{fig\_iscd\_all}, where the output extrinsic information from both the spatial and temporal decoders is added to form the output extrinsic information of 3D-MRF decoder, which is denoted by $L\_c^{(E)}$. Besides, in parameter estimation block, $(\beta\_s,\beta\_t)$ is jointly estimated as derived in (\ref{eq\_beta\_est}).

The ISCD shown in Fig. \ref{fig\_iscd\_all} has the properties of both serial concatenated codes (SCCC) and parallel concatenated codes (PCCC). The spatial MRF redundancy can be seen as an outer code in SCCC, and thus the channel code is seen as the inner code in SCCC. Besides, in the parallel direction, the temporal MRF redundancy of consecutive previous and following video frames is exploited by 3D-MRF temporal decoder, acting like the conventional PCCC iterative decoder. Consequently, the 3D-MRF based ISCD combines the spatial (serial) and temporal (parallel) extrinsic information and therefore significantly enhances the video decoding quality.