# Auto Regressive Model and Weighted Least Squares based Packet Video Error Concealment

Yongbing Zhang[1], Xinguang Xiang[1], Siwei Ma[2], Debin Zhao[1], Wen Gao[2]
[1]*Harbin Institute of Technology, Harbin, China*
[2]*Peking University, Beijing, China*
*{ybzhang,xgxiang,swma,dbzhao,wgao}@jdl.ac.cn*

***Abstract***

*In this paper, auto regressive (AR) model is applied to error concealment for block-based packet video encoding. Each pixel within the corrupted block is restored as the weighted summation of corresponding pixels within the previous frame in a linear regression manner. Two novel algorithms using weighted least squares method are proposed to derive the AR coefficients. First, we present a coefficient derivation algorithm under the spatial continuity constraint, in which the summation of the weighted square errors within the available neighboring blocks is minimized. The confident weight of each sample is inversely proportional to the distance between the sample and the corrupted block. Second, we provide a coefficient derivation algorithm under the temporal continuity constraint, where the summation of the weighted square errors around the target pixel within the previous frame is minimized. The confident weight of each sample is proportional to the similarity of geometric proximity as well as the intensity gray level. The regression results generated by the two algorithms are then merged to form the ultimate restorations. Various experimental results demonstrate that the proposed error concealment strategy is able to increase the peak signal-to-noise ratio (PSNR) compared to other methods.*

## 1. Introduction

State-of-the-art video coding standard H.264/AVC [1] significantly outperforms the previous coding standards, such as MPEG-1, H.262/MPEG-2 and H.263. Although many newly adopted techniques lead to the improvement of compression efficiency, the highly compressed bit stream is susceptible to transmission errors due to the limited bandwidth constraint. Consequently, packet errors, which will severely degrade the display quality at the decoder side, are unavoidable.

Error concealment (EC) is a post processing technique which hides the packet errors utilizing the correctly received information at the decoder side without modifying source and channel coding schemes. According to the information utilized, EC can be categorized into spatial approaches that employ spatially adjacent pixels for recovering the lost blocks, and the temporal approaches that fill the lost blocks utilizing the pixels in the previous frames.

Smoothness constraint is commonly utilized by the spatial methods. Y. Wang [2] proposed a spatial error concealment method by minimizing a first-order derivative-based smoothness measure. To suppress the induced blurring artifacts, second-order derivatives were considered in [3]. Although smoothness constraint achieves good results for the flat regions, it may not be satisfied in areas with high frequency edges. To tackle such shortcomings, an edge-preserving algorithm [4] was proposed to interpolate the missing pixels. Spatial methods may yield better performance than temporal methods in scenes with high motion, or after a scene change [5]. However, they can not restore the detail textures of the corrupted blocks. In this case, the information from the past frames (temporal methods) may improve the quality of the concealed blocks.

1068-0314/10 $26.00 © 2010 IEEE
DOI 10.1109/DCC.2010.100

Different from spatial methods, temporal methods restore the corrupted blocks by exploiting temporal correlation between successive frames. An important issue with this method is to find the most suitable substitute blocks from the previous frame, i.e., selecting the optimal motion vectors (MVs) for the lost blocks. If the MV of the corrupted block is available at the decoder, it can be utilized directly to motion-compensated the corrupted block. When the MV is also lost, it has to be estimated. Many pioneering works have been done on recovering of MVs. In [6], zero MV, the MV of the collocated block in the reference frame, and the average or the median of the MVs from the spatially adjacent blocks are selected as the candidate MVs for the corrupted blocks. The well known Boundary matching algorithm (BMA) proposed in [7] selected the MV that minimizes the total variation between the internal boundary and the external boundary of the reconstructed block as the optimal one to recover the corrupted block. There are also some more sophisticated algorithms [8-12] to obtain better MVs for the corrupted blocks. All these methods attempt to find the best MV in the previous quarter-pel resolution frame, which are interpolated by the fixed filter tap. However, the fixed interpolation filter can not capture the local region property quite well. To adaptively tune the interpolation filter coefficient, we propose an auto-regressive (AR) based EC scheme by extending our previous work [13] in this paper. In [13], the AR coefficients are derived by the least squares method assigning the same confident weight for each sample. However, [13] does not consider the influence of noisy samples. To better inhibit the influence of noisy samples, we assign a confident weight for each sample under the spatial continuity constraint in this paper. Besides, we also propose a weight derivation algorithm under the temporal continuity constraint. The regression results generated by the two algorithms are merged to form ultimate restorations for the corrupted blocks.

## 2. AR model based EC

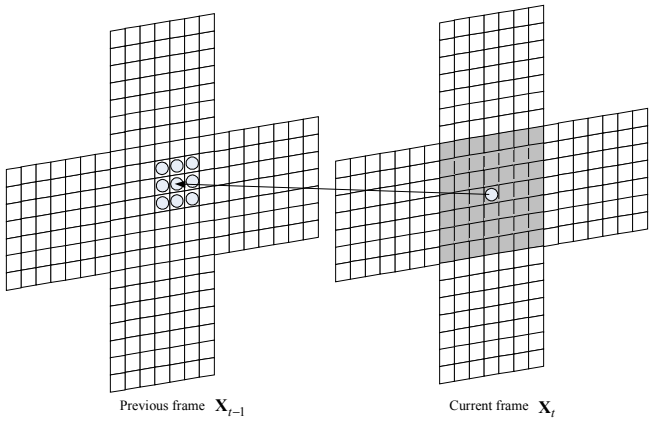


Fig. 1 Auto-regressive Model

Fig. 1 shows the restoring process by AR model. For each corrupted pixel, the corresponding pixel along the integer-pel accuracy motion vector in the previous reconstructed frame is first found, and then all the pixels within a square centered at the corresponding pixel are combined in a linear regression form to recover the corrupted pixel. As shown in Fig.1, the linear regression can be expressed as

$$\hat{x}_t(i,j) = \sum_{k=-R}^{R} \sum_{l=-R}^{R} \alpha(k,l)\, x_{t-1}(i+k, j+l), \tag{1}$$

where $\hat{x}_t(i,j)$ represents the corrupted pixel located at $(i,j)$ within the current frame $\mathbf{X}_t$, $x_{t-1}(i,j)$ represents the corresponding pixel (pointed by the integer-pel accuracy motion

vector) within the previous reconstructed frame $\mathbf{X}_{t-1}$ , $\alpha(k,l)$ represents the desired coefficients, and *R* represents the radius of the AR model.

Defining $\Psi_{k,l,R}$ as an operator that extracts a patch of a fixed and predetermined size (centered at $(k,l)$ and with $(2R+1)\times(2R+1)$ pixels) from an image, the expression $\Psi_{k,l,R}\mathbf{X}_{t-1}$ ( $\mathbf{X}_{t-1}$ is represented as a vector by lexicographic ordering) results with a vector of length $(2R+1)^2$ being the extracted patch.Thus, the linear regression in Eq. (1) can also be expressed as

$$\hat{x}_t(i,j) = \Psi_{i,j,R}\mathbf{X}_{t-1}\boldsymbol{\alpha}^T , \tag{2}$$

where $\boldsymbol{\alpha}$ represents the coefficient vector of the AR model. The summed square error between the corrupted and the actual pixel is

$$e^2 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(x_t(i,j)-\hat{x}_t(i,j)\right)^2 = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(x_t(i,j)-\Psi_{i,j,R}\mathbf{X}_{t-1}\boldsymbol{\alpha}^T\right)^2 , \tag{3}$$

where *N* represents the width of the corrupted block. To minimize $e^2$, the first derivative of $e^2$ to $\boldsymbol{\alpha}$ should be equal to zero according to the least squares algorithm, i.e.,

$$\frac{\partial e^2}{\partial \boldsymbol{\alpha}} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)\right)\boldsymbol{\alpha}^T - \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}x_t(i,j)\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T = 0 . \tag{4}$$

By solving the above equation, we will get the optimal coefficients as

$$\boldsymbol{\alpha}^T = \left[\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)\right)\right]^{-1}\left[\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}x_t(i,j)\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T\right]. \tag{5}$$

However, since the actual pixel $x_t(i,j)$ is not available at the decoder side, we can not directly obtain the AR coefficients according to Eq. (5). Instead, we have to estimate the optimal AR coefficients according to the adjacent spatial and temporal information of the lost blocks.

## 3. Weighted least squares algorithm

### 3.1. Weight derivation under spatial continuity constraint

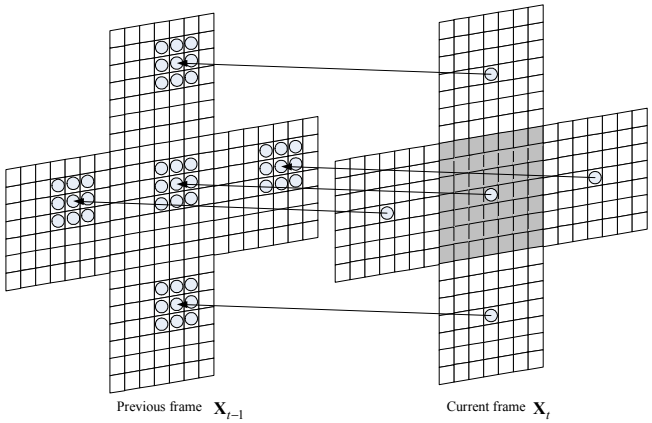


Fig. 2 Spatial continuity constraint

Pixels within adjacent blocks have a high possibility of belonging to the same object, which can be reflected by the phenomenon that adjacent blocks possess similar motion trends. This property is called spatial continuity constraint in this paper, based on which we can

derive the AR coefficients of the corrupted block. It is noted that under spatial continuity constraint, all the pixels within the lost block have the same AR coefficients, just like all the pixels within the corrupted block have the same MV in the traditional EC method. As shown in Fig.2, each pixel within the corrupted block and the available neighboring blocks can be regressed by the corresponding pixels within the previous reconstructed frame utilizing the same AR coefficients.

It should be noted that during the weight derivation process, different training samples should be assigned different probabilistic confidence to inhibit the influence of noisy sample. For example, the pixels that are closer to the corrupted block or with similar textures should be assigned a larger probabilistic confidence. Let $\mathbf{B}_t$ be available neighboring blocks within the current frame, i.e., $\mathbf{B}_t \subset \mathbf{X}_t$. In addition, let $b_t(i,j)$ be an arbitrary pixel within $\mathbf{B}_t$, i.e., $b_t(i,j) \in \mathbf{B}_t$, and the corresponding probabilistic confidence of $b_t(i,j)$ under the spatial continuity constraint is $w_{\boldsymbol{\alpha}}(i,j)$, with $0 \le w_{\boldsymbol{\alpha}}(i,j) \le 1$ and $\sum_{(i,j)\in\mathbf{B}_t} w_{\boldsymbol{\alpha}}(i,j) = 1$. $b_t(i,j)$ can also be represented by the regression function of $\mathbf{X}_{t-1}$ and $\boldsymbol{\alpha}$ as

$$\hat{b}_t(i,j) = \Psi_{i,j,R}\mathbf{X}_{t-1}\boldsymbol{\alpha}^T. \tag{6}$$

Assume the regression error of $b_t(i,j)$ induced by $\boldsymbol{\alpha}$ and $\mathbf{X}_{t-1}$ follows independent Gaussian distribution, i.e.,

$$P\left(\mathbf{X}_{t-1}, b_t(i,j) \mid \boldsymbol{\alpha}\right) = \frac{1}{\sqrt{2\pi}\sigma_{\boldsymbol{\alpha}}} \exp\left\{-\frac{\left\|b_t(i,j) - \hat{b}_t(i,j)\right\|^2}{{\sigma_{\boldsymbol{\alpha}}}^2}\right\}, \tag{7}$$

where $\sigma_{\boldsymbol{\alpha}}$ represents the variance of the error between $b_t(i,j)$ and $\hat{b}_t(i,j)$. Since the probabilistic confidence of $b_t(i,j)$ is $w_{\boldsymbol{\alpha}}(i,j)$, the weighted error is termed as $\left(b_t(i,j) - \hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)$, and accordingly, the corresponding weighted error distribution is represented as

$$P\left(\mathbf{X}_{t-1}, b_t(i,j) \mid \boldsymbol{\alpha}\right) = \frac{1}{\sqrt{2\pi}\sigma'_{\boldsymbol{\alpha}}} \exp\left\{-\frac{\left\|\left(b_t(i,j) - \hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)\right\|^2}{{\sigma'_{\boldsymbol{\alpha}}}^2}\right\}, \tag{8}$$

where $\sigma'_{\boldsymbol{\alpha}}$ represents the variance of the weighted error between $b_t(i,j)$ and $\hat{b}_t(i,j)$. Based on the observations of $\mathbf{B}_t$ and $\mathbf{X}_{t-1}$, the AR coefficients can be obtained via a maximum *a posteriori* (MAP) as follows.

Under MAP, the coefficient $\hat{\boldsymbol{\alpha}}$ is given by

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} P\left(\boldsymbol{\alpha} \mid \mathbf{X}_{t-1}, \mathbf{B}_t\right), \tag{9}$$

where $P\left(\boldsymbol{\alpha} \mid \mathbf{X}_{t-1}, \mathbf{B}_t\right)$ represents the occurring probability of $\boldsymbol{\alpha}$ given $\mathbf{X}_{t-1}$ and $\mathbf{B}_t$. Using Bayesian law, we have

$$P\left(\boldsymbol{\alpha} \mid \mathbf{X}_{t-1}, \mathbf{B}_t\right) = \frac{P\left(\mathbf{X}_{t-1}, \mathbf{B}_t \mid \boldsymbol{\alpha}\right) P(\boldsymbol{\alpha})}{P\left(\mathbf{X}_{t-1}, \mathbf{B}_t\right)}, \tag{10}$$

where $P\left(\mathbf{X}_{t-1}, \mathbf{B}_t \mid \boldsymbol{\alpha}\right)$ is the observation conditional probability, $P(\boldsymbol{\alpha})$ is the prior probability for AR coefficients. Since $P\left(\mathbf{X}_{t-1}, \mathbf{B}_t\right)$ is not a function of $\boldsymbol{\alpha}$, it can be ignored when maximizing $P\left(\boldsymbol{\alpha} \mid \mathbf{X}_{t-1}, \mathbf{B}_t\right)$. Based on the assumption that $\boldsymbol{\alpha}$ obeys uniform distribution, maximizing $P\left(\boldsymbol{\alpha} \mid \mathbf{X}_{t-1}, \mathbf{B}_t\right)$ equals to maximizing $P\left(\mathbf{X}_{t-1}, \mathbf{B}_t \mid \boldsymbol{\alpha}\right)$. Consequently, we have

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \arg\max_{\boldsymbol{\alpha}} P\left(\mathbf{X}_{t-1}, \mathbf{B}_t \mid \boldsymbol{\alpha}\right) \\
&= \arg\max_{\boldsymbol{\alpha}} \prod_{(i,j)\in \mathbf{B}_t} P\left(\mathbf{X}_{t-1}, b_t(i,j) \mid \boldsymbol{\alpha}\right) \\
&= \arg\max_{\boldsymbol{\alpha}} \prod_{(i,j)\in \mathbf{B}_t} \frac{1}{\sqrt{2\pi}\sigma'_{\boldsymbol{\alpha}}} \exp\left\{-\frac{\left\|\left(b_t(i,j)-\hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)\right\|^2}{\sigma'^{2}_{\boldsymbol{\alpha}}}\right\}, \\
&= \arg\max_{\boldsymbol{\alpha}} C_{\boldsymbol{\alpha}} \exp\left\{-\sum_{(i,j)\in \mathbf{B}_t} \left\|\left(b_t(i,j)-\hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)\right\|^2\right\}
\end{aligned}
\tag{11}
$$

with $C_{\boldsymbol{\alpha}} = \left(\sqrt{2\pi}\sigma'_{\boldsymbol{\alpha}}\right)^{-|\mathbf{B}_t|} \exp\left\{-\frac{|\mathbf{B}_t|}{\sigma'^{2}_{\boldsymbol{\alpha}}}\right\}$.

From Eq. (11), it is obvious that maximizing $P(\mathbf{X}_{t-1}, \mathbf{B}_t \mid \boldsymbol{\alpha})$ equals to minimizing $\sum_{(i,j)\in \mathbf{B}_t} \left\|\left(b_t(i,j)-\hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)\right\|^2$, i.e.,

$$
\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{(i,j)\in \mathbf{B}_t} \left\|\left(b_t(i,j)-\hat{b}_t(i,j)\right) w_{\boldsymbol{\alpha}}(i,j)\right\|^2 . \tag{12}
$$

Since the correlation between pixels decreases with the increase of distance, $w_{\boldsymbol{\alpha}}(i,j)$ is set to be inversely proportional to the distance between $b_t(i,j)$ and the corrupted block. That is to say

$$
w_{\boldsymbol{\alpha}}(i,j) = \begin{cases} \frac{1}{N-i}, & \text{if } b_t(i,j) \in \textit{up neighboring block} \\ \frac{1}{N-j}, & \text{if } b_t(i,j) \in \textit{left neighboring block} \\ \frac{1}{i+1}, & \text{if } b_t(i,j) \in \textit{bottom neighboring block} \\ \frac{1}{j+1}, & \text{if } b_t(i,j) \in \textit{right neighboring block} \end{cases}, 0 \le (i,j) \le N-1 . \tag{13}
$$

By setting the first derivative of the weighted square errors in Eq. (12) to zero, the AR coefficients under spatial continuity constraint is computed as

$$
\boldsymbol{\alpha}^T = \left[\sum_{(i,j)\in \mathbf{B}_t} \left(\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T \left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)\right)\right]^{-1} \left[\sum_{(i,j)\in \mathbf{B}_t} b_t(i,j)\left(\Psi_{i,j,R}\mathbf{X}_{t-1}\right)^T\right]. \tag{14}
$$

With the obtained AR coefficient $\boldsymbol{\alpha}$, the lost block is restored according to Eq. (2).

### 3.2. Weight derivation under temporal continuity constraint

Besides spatial continuity constraint, video sequence also has temporal continuity constraint, which can be embodied by the observation that the same object among adjacent frames is usually threaded by the same motion trajectory. If the current frame has two previous frames, the temporal continuity constraint can also be utilized to derive the AR coefficients. As shown in Fig. 3, for each corrupted pixel $x_t(k,l)$, the corresponding pixel $x_{t-1}(k,l)$ in the closest reconstructed frame is first found, and the corresponding pixel $x_{t-2}(k,l)$ in the second closest reconstructed frame is also found. After that, $x_{t-1}(k,l)$ is regressed by $x_{t-2}(k,l)$ as well as the pixels surround it as

$$
\hat{x}_{t-1}(k,l) = \Psi_{k,l,R}\mathbf{X}_{t-2}\boldsymbol{\beta}^T , \tag{15}
$$

where $\boldsymbol{\beta}$ represents the AR coefficients derived by the temporal continuity constraint. The derived coefficient $\boldsymbol{\beta}$ is then utilized to restore the corrupted pixel $x_t(k,l)$.

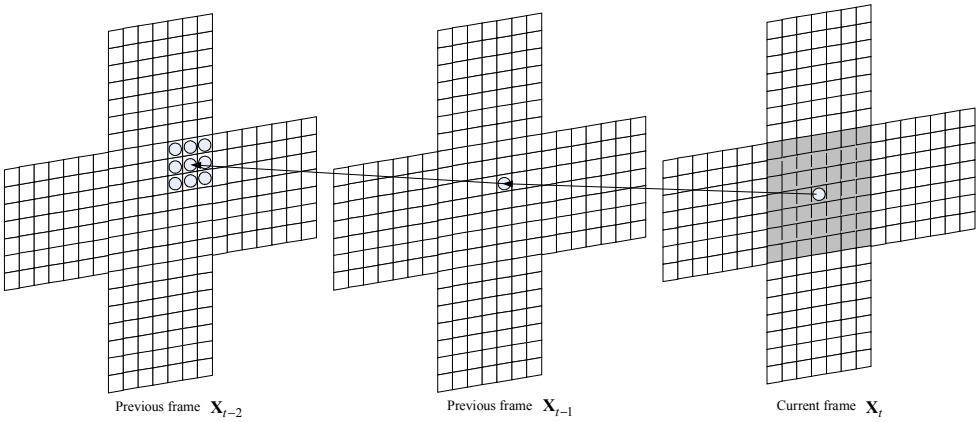


Fig. 3 Temporal continuity constraint

It is noted that under spatial continuity constraint, all the pixels within the corrupted block have the same AR coefficients. Obviously, it can not reflect the local features of each corrupted pixel. To remedy the deficiency of this method, we assign different AR coefficients for different pixel under the temporal continuity constraint. To compute the appropriate AR coefficients for $x_t(k,l)$, we need to find some training samples around $x_{t-1}(k,l)$. In this paper, the training sample is defined to be $\Psi_{k,l,M}\mathbf{X}_{t-1}$, with $M \geq R$.

Based on the observation of $\Psi_{k,l,M}\mathbf{X}_{t-1}$ and $\mathbf{X}_{t-2}$, the coefficient $\hat{\boldsymbol{\beta}}$ can be computed as

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} P\left(\boldsymbol{\beta} \mid \Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2}\right). \tag{16}$$

Using the Bayesian law, we have

$$P\left(\boldsymbol{\beta} \mid \Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2}\right) = \frac{P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right)P(\boldsymbol{\beta})}{P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2}\right)}, \tag{17}$$

where $P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right)$ is the observation conditional probability, $P(\boldsymbol{\beta})$ is the prior probability for AR coefficients. Similar to the previous subsection, maximizing $P\left(\boldsymbol{\beta} \mid \Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2}\right)$ equals to maximizing $P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right)$. Again, assume the weighted error of $x_{t-1}(u,v)$ and $\hat{x}_{t-1}(u,v)$, $(u,v) \in \Psi_{k,l,M}$, follows independent Gaussian distribution, i.e.,

$$P\left(x_{t-1}(u,v), \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right) = \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left\{-\frac{\left\|\left(x_{t-1}(u,v) - \hat{x}_{t-1}(u,v)\right)w_\beta(u,v)\right\|^2}{\sigma_\beta^2}\right\}, \tag{18}$$

where $\sigma_\beta$ represents the variance of the weighted error between $x_{t-1}(u,v)$ and $\hat{x}_{t-1}(u,v)$, $w_\beta(u,v)$ represents the probabilistic confidence of $x_{t-1}(u,v)$, with $0 \leq w_\beta(u,v) \leq 1$ and $\sum_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} w_\beta(u,v) = 1$. Consequently, we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \arg\max_{\boldsymbol{\beta}} P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right) \\
&= \arg\max_{\boldsymbol{\beta}} \prod_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} P\left(x_{t-1}(u,v), \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right) \\
&= \arg\max_{\boldsymbol{\beta}} \prod_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left\{-\frac{\left\|\left(x_{t-1}(u,v)-\hat{x}_{t-1}(u,v)\right)w_\beta(u,v)\right\|^2}{{\sigma_\beta}^2}\right\}, \\
&= \arg\max_{\boldsymbol{\beta}} C_{\boldsymbol{\beta}} \exp\left\{-\sum_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} \left\|\left(x_{t-1}(u,v)-\hat{x}_{t-1}(u,v)\right)w_\beta(u,v)\right\|^2\right\}
\end{aligned}
\tag{19}
$$

with $C_{\boldsymbol{\beta}} = \left(\sqrt{2\pi}\sigma_{\boldsymbol{\beta}}\right)^{-\left|\Psi_{k,l,M}\mathbf{X}_{t-1}\right|} \exp\left\{-\frac{\left|\Psi_{k,l,M}\mathbf{X}_{t-1}\right|}{{\sigma_{\boldsymbol{\beta}}}^2}\right\}$.

It is obvious that maximizing $P\left(\Psi_{k,l,M}\mathbf{X}_{t-1}, \mathbf{X}_{t-2} \mid \boldsymbol{\beta}\right)$ equals to minimizing $\sum_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} \left\|\left(x_{t-1}(u,v)-\hat{x}_{t-1}(u,v)\right)w_\beta(u,v)\right\|^2$, i.e.,

$$
\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}} \left\|\left(x_{t-1}(u,v)-\hat{x}_{t-1}(u,v)\right)w_\beta(u,v)\right\|^2. \tag{20}
$$

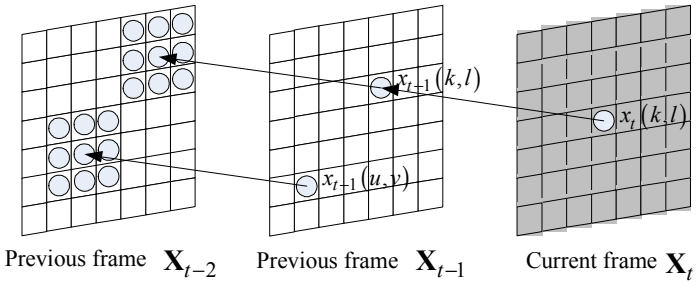


Fig. 4 Pixels involved during probabilistic confidence computation under temporal continuity constraint

Inspired by the nonlocal-means method in [14], each sample is assigned a confident weight that reflects the probability that this sample and the target sample have the same value. As shown in Fig. 4, when computing the AR weights for $x_t(k,l)$, we first find the corresponding pixel $x_{t-1}(k,l)$ (the target pixel sample) in the closest previous frame. And then we collect the training samples within $\Psi_{k,l,M}\mathbf{X}_{t-1}$. For each training sample $x_{t-1}(u,v)$ within $\Psi_{k,l,M}\mathbf{X}_{t-1}$, we will get the corresponding pixels $\Psi_{u,v,R}\mathbf{X}_{t-2}$ utilized during regressing. The similarity between $x_{t-1}(k,l)$ and $x_{t-1}(u,v)$ depends on the similarity of geometric proximity as well as the intensity gray level $\Psi_{k,l,R}\mathbf{X}_{t-2}$ and $\Psi_{u,v,R}\mathbf{X}_{t-2}$. The similarity is measured as a decreasing function of the weighed Euclidean distance between $\Psi_{k,l,R}\mathbf{X}_{t-2}$ and $\Psi_{u,v,R}\mathbf{X}_{t-2}$ as well as the geometric proximity, namely

$$
w_{\boldsymbol{\beta}}(u,v) = \frac{1}{S(k,l)} \exp\left\{-\frac{\left\|\left[\Psi_{k,l,R}\mathbf{X}_{t-2} - \Psi_{u,v,R}\mathbf{X}_{t-2}\right] \bullet K\right\|^2}{2\sigma^2}\right\}, (u,v) \in \Psi_{k,l,M}\mathbf{X}_{t-1} \tag{21}
$$

where $\bullet$ represents the inner product of two vectors, and $\sigma^2$ is a constant to control the decay of the exponential function. Here, the element of $K$ is computed as

$$
K(m,n) = \frac{1}{\gamma} \exp\left\{-\sqrt{(m-k)^2 + (n-l)^2}\right\}, \tag{22}
$$

and

$$S(k,l)=\sum_{(u,v)\in\Psi_{k,l,M}\mathbf{X}_{t-1}}\exp\left\{-\frac{\left\|\left[\Psi_{k,l,R}\mathbf{X}_{t-2}-\Psi_{u,v,R}\mathbf{X}_{t-2}\right]\bullet K\right\|^2}{2\sigma^2}\right\}. \tag{23}$$

It is noted that $\gamma$ is a normalizing constant and $K(m,n)$ is a weight value for each sample in similarity window, which decreases with the distance from the center of similarity window.

After have obtained the AR coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the ultimate restored pixel is merged as

$$\hat{x}_t(i,j)=\tau\bullet\Psi_{i,j,R}\mathbf{X}_{t-1}\boldsymbol{\alpha}^T+(1-\tau)\bullet\Psi_{i,j,R}\mathbf{X}_{t-1}\boldsymbol{\beta}^T. \tag{24}$$

where $0\le\tau\le1$ is the merging factor. In this paper, $\tau$ is set to be 0.5.

## 4. Simulation results

In this section, various experiments are conducted to verify the superior performance of the proposed EC scheme. Three different EC algorithms, including the EC method of JM (JM), our previous work in [13], and the proposed scheme, are simulated on the H.264 reference software JM 10.0. It is noted that the method in [13] is equivalent to our method under spatial continuity when the probabilistic confidences of all the training samples are set to be the same. The first 100 frames of CIF sequences Foreman, Mobile and Flower at 30 fps are encoded under the quantization parameters (QPs) of 24 and 28. Group of Picture (GOP) of IPPP…structure with one I frame inserted every 16 frames are considered. Two reference frames and CABAC are used during encoding. The de-compressed streams are dropped at the packet loss rates (PLRs) of 5%, 10%, and 20% respectively. The size of missing blocks is set to be 16x16 to coincide with the standard's macroblock (MB). Each row of MB is grouped as a slice and when one slice is lost, all the MBs within the same row are corrupted. In the experiments, parameters of R and M are set to be 1 and 3, respectively.

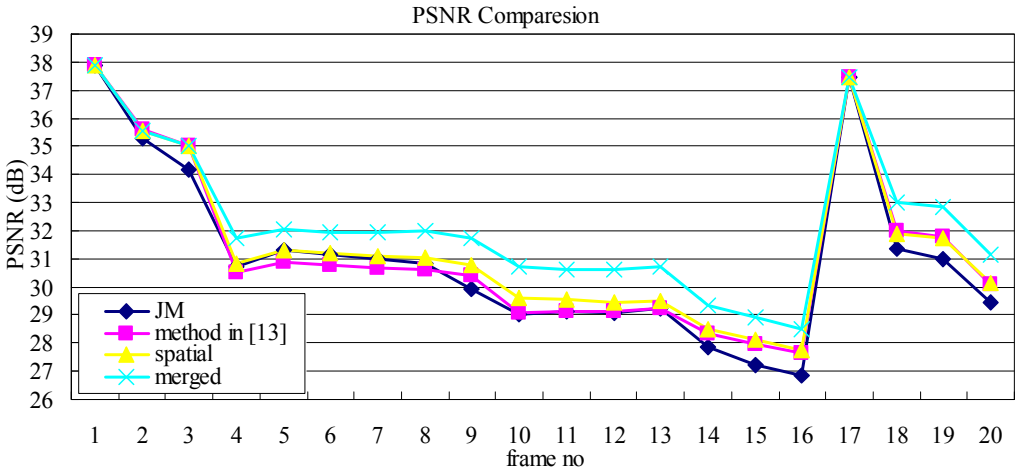


Fig. 5. PSNR (dB) comparison for Foreman with PLR=10%

The peak signal-to-noise ratio (PSNR) is used to quantitatively evaluate the recovered video quality. The average PSNR values of each compared method are shown in Table 1, where spatial represents the proposed method only under spatial continuity constraint, and merged represents the proposed method under spatial continuity constraint and temporal continuity. Spatial has a 0.49~2.74dB improvement than JM and 0~0.23dB improvement than [13], except that there is about 0.09dB loss for Flower when the PLR is 10%. When

combining the spatial and temporal continuity constraints, the merged result has a 0.76~3.36dB improvement than JM and 0.14~0.9dB improvement than [13].

Fig. 5 represents the PSNR performance of the first 20 frames for Foreman with PLR=10%. Since I frames are not corrupted in the simulation, the 1st and 17th frame have the same PSNR values. It reports that the proposed method significantly outperforms JM and the method in [13].

Table 1 Average PSNR performance comparison of different EC methods

| Video sequences | QP | Method | PLR | | |
|---|---|---|---|---|---|
| | | | 5% | 10% | 20% |
| Foreman | 24 | JM | 36.18 | 30.52 | 27.05 |
| | | [13] | 36.56 | 31.47 | 27.77 |
| | | spatial | 36.79 | 31.65 | 27.87 |
| | | merged | 37.46 | 32.25 | 27.91 |
| | 28 | JM | 34.76 | 29.93 | 26.90 |
| | | [13] | 35.25 | 30.75 | 27.45 |
| | | spatial | 35.25 | 30.97 | 27.51 |
| | | merged | 35.73 | 31.59 | 27.66 |
| Mobile | 24 | JM | 32.56 | 25.96 | 22.99 |
| | | [13] | 35.00 | 28.58 | 25.03 |
| | | spatial | 35.01 | 28.70 | 25.08 |
| | | merged | 35.29 | 29.32 | 25.48 |
| | 28 | JM | 31.11 | 25.54 | 22.81 |
| | | [13] | 32.64 | 27.78 | 24.66 |
| | | spatial | 32.66 | 27.91 | 24.72 |
| | | merged | 32.88 | 28.47 | 25.10 |
| Flower | 24 | JM | 34.04 | 26.13 | 22.95 |
| | | [13] | 35.56 | 28.32 | 24.39 |
| | | spatial | 35.63 | 28.42 | 24.48 |
| | | merged | 36.11 | 29.25 | 25.13 |
| | 28 | JM | 32.23 | 26.04 | 22.75 |
| | | [13] | 33.33 | 27.93 | 24.26 |
| | | spatial | 33.37 | 27.84 | 24.29 |
| | | merged | 33.71 | 28.69 | 24.89 |

## 5. Conclusion

This paper presents a novel error concealment scheme utilizing auto regressive model, where each corrupted pixel is restored as a weighted summation of corresponding pixels within the previous frame in a linear regression manner. Two algorithms utilizing weighted least squares method under the spatial continuity and temporal continuity constraints are proposed to derive the AR coefficients. Simulation results suggest that the proposed error concealment method is able to provide much better performance than other methods.

## Acknowledgement

This work was supported in part by National Science Foundation of China (60736043), and National Basic Research Program of China (973 Program, 2009CB320905).

## References


[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] Y. Wang, Q.-F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," IEEE Trans. Commun., vol. 41, no. 10, pp. 1544–1551, Oct. 1993.

[3] W. Zhu, Y. Wang, and Q.-F. Zhu, "Second-order derivative-based smoothness measure for error concealment in DCT-based codecs," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 6, pp. 713–718, Oct. 1998.

[4] S. D. Rane, G. Sapiro, and M. Bertalmio, "Structure and texture filling-in of missing image blocks in wireless transmission and compression," IEEE Trans. Image Process., vol. 12, no. 3, pp. 296–303, Mar. 2003.

[5] D. Persson, T. Eriksson, and P. Hedelin, "Packet video error concealment with Gaussian mixture models," IEEE Trans. Image Process., vol. 17, no. 2, pp. 145–154, Feb. 2008.

[6] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in Proc. ICASSP-92: 1992 IEEE Int. Conf. Acoustics, Speech, Signal Process., San Francisco, CA, 1992, vol. 3, pp. 545–548.

[7] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., 1993, vol. 3, pp. 417–420.

[8] S. Tsekeridou, F. A. Cheikh, M. Gabbouj, and I. Pitas, "Motion field estimation by vector rational interpolation for error concealment purposes,"in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.,Mar. 1999, vol. 6, pp. 3397–3400.

[9] M. Al-Mualla, N. Canagarajahm, and D. R. Bull, "Error concealment using motion field interpolation," in Proc. IEEE Int. Conf. Image Process., Oct. 1998, vol. 3, pp. 512–516.

[10] J. H. Zheng and L. P. Chau, "A temporal error concealment algorithm for H.264 using Lagrange interpolation," in Proc. IEEE Int. Symp. Circuits Syst., 2004, pp. 133–136.

[11] Z. W. Gao and W. N. Lie, "Video error concealment by using Kalmanfiltering technique," in Proc. IEEE Int. Symp.-Circuits Syst., 2004, pp. 69–72.

[12] W. N. Lie and Z. W. Gao, "Video error concealment by intergrating greedy suboptimization and Kalman filtering techniques," IEEE Trans. Circuits Syst. Video Technol., vol. 16, pp. 982–992, 2006.

[13] X. Xiang, Y. Zhang, D. Zhao, S. Ma, and W. Gao, "A High Efficient Error Concealment Scheme Based On Auto-Regressive Model For Video Coding", in Proceedings of Picture coding symposium ,PCS, Chicago, USA, May.06-08, 2009.

[14] M. Protter, M. Elad, H. Takeda and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction" IEEE Trans. on Image Processing, vol.18, No.1, pp. 36-51, Jan. 2009.