

OPTIMAL POWER ALLOCATION FOR MINIMIZING VISUAL DISTORTION OVER MIMO COMMUNICATION SYSTEMS

Jincheol Park[†], *Uk Jang*^{*}, *Taegeun Oh*[†], *Sanghoon Lee*[†] and *Alan C. Bovik*[‡]

[†]Wireless Network Lab., Center for IT of Yonsei University, Seoul, Korea, 120-749.

^{*} Electronics and Telecommunications Research Institute (ETRI), Taejeon, Korea, 305-700.

[‡]Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX

ABSTRACT

A recent dynamic increase in demand for wireless multimedia services has greatly accelerated the research on cross layer optimization techniques for transmitting multimedia data over wireless channel. In this paper, we explore a novel theoretical approach for joint optimization between the rate distortion (RD) of H.264/AVC video and the link-capacity of MIMO parallel subchannels. We obtain the optimal power level of subchannels through an optimization problem to minimize total visual distortion. In the simulation results, compared to the water filling (WF) method, the proposed scheme provides better results in aspects of visual quality in the face of sum rate loss.

1. INTRODUCTION

In the fourth-generation (4G) mobile communication networks, wireless multimedia applications (visual communications, multimedia streaming, interactive applications, etc.) require high data-rate, high-quality and real-time multimedia services anytime and anywhere over a broad bandwidth with sufficient robustness in spite of the inherently limited and unpredictable wireless channel. Due to the increasing demand for wireless multimedia data, cross layer optimization techniques for transmitting multimedia data over wireless channel are recently emerged as one of the most prominent research topics. Recent attempts allocate the radio resources of orthogonal frequency division multiplexing (OFDM) or multi-input multi-output (MIMO) systems for H.264/AVC video transmission [1]-[3].

In [2], a power allocation scheme is proposed for transmitting H.264/AVC video data over a MIMO-OFDM wireless channel, and in [3], an unequal error protection (UEP) algorithm is proposed to transmit scalable video over MIMO system, in a practical manner. However very few attempts have been made at theoretical analysis to obtain the performance upper or lower bound for the multimedia data transmission. In [1], we obtained a closed-form optimal loading ratio over

an OFDMA cellular network by means of the theoretical upper bound of a quality criterion. In this paper, we analyze the theoretical lower bound of visual distortion using a novel approach when allocating the power for H.264/AVC video transmission. The slice containing visually important data (i.e. region of interest, ROI) is transmitted by the principal eigen subchannel of the MIMO system, and the slices containing visually less important data are transmitted by the remaining subchannels. An optimal power level of each subchannel is determined by joint optimization between the rate distortion (RD) of slices in video frames and the link-capacity of corresponding MIMO subchannels to minimize the visual distortion. Finally, a closed-form solution of optimal power allocation is obtained by the Lagrangian method. Compared to the water filling (WF) algorithm, which is the optimal power allocation method in that it maximizes sum rate [6], the proposed scheme provides better results in aspects of visual quality in the face of sum rate loss.

2. VISUAL WEIGHT MODEL

Suppose that two users perform a visual communication over a capacity limited mobile channel. Generally, it is common knowledge that the face object or region is the most important visual information needed to be delivered. Thus, each frame can be spatially divided into the ROI region and the remaining non-ROI regions according to visual attention. The ROI region is determined by using an eye tracker or a face detection algorithm and so on. In the H.264/AVC, the ROI coding can be implemented by using the error resilience tool flexible macro-block ordering (FMO), which allows to group the ROI while only adding acceptable overhead [4]. The H.264/AVC specifies 7 different types of FMO, labeled Type 0 to Type 6. Among them, Type 2 divides one or more rectangular slice groups and a background as shown in Fig. 1 (a).

The relative difference in visual importance among slice groups can be obtained by the human visual system (HVS)[5]. When the visual stimulus is perceived by the HVS, the light coming from objects is projected on the retina. A point on the retina that the light is directly focused on is called fovea,

This work was supported in part by the MIC, Korea, under the ITRC support program supervised by the IITA (IITA-2005-C1090-0502-0012).

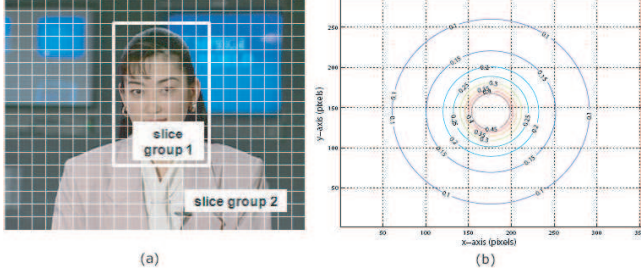


Fig. 1. (a) Slice grouping by Type 2 of FMO, (b) The local bandwidth

which lies on the visual axis. On the image, the point mapped into the fovea is called “foveation point”, which is generally in ROI. The retina possesses a nonuniform spatial distribution (resolution) of photoreceptor cells. It decreases exponentially with distance away from the fovea. Thus, the image is sampled non-uniformly by the cones. By the Shannon’s sampling theorem, the perceptible frequency bandwidth becomes locally different according to the sampling density, which is called the local bandwidth. Thus, the image information near the foveation points should be more importantly treated than further ones.

In [5], the fast foveated algorithm calculates the local bandwidths, shown in Fig.1 (b), by an exponential model of the spatial sampling grid of the fovea in the human retina, in conjunction with an assumed viewing distance, and by choosing the model parameters such that the human eye would be unable to discriminate the foveated image from the original, assuming a given fixation point in the image being observed. Using the model, the local bandwidths, f_n , of the n^{th} macro block are obtained. Finally, normalizing the sum of the local bandwidths over each slice group, the visual weight of the l^{th} slice, w_l , is calculated like below,

$$w_l = \frac{\sum_{n \in S_l} f_n}{\sum_{n \in A} f_n} \quad (1)$$

where S_l denotes the macro-block (MB) set contained in the l^{th} slice group, A is a set of whole MBs in a frame.

3. MIMO CHANNEL AND SINGULAR VALUE DECOMPOSITION

We consider a communication system with a transmitter (with M_T transmit antennas) and a receiver (with M_R receive antennas). The $M_R \times 1$ received signal vector, \mathbf{y} , is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (2)$$

where \mathbf{n} denotes the $M_R \times 1$ independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector with covariance matrix $N_o \mathbf{I}_{M_R}$. The signal vector to be sent from the transmitter is expressed as $\mathbf{x} = [s_1, \dots, s_{M_T}]^T$, with

$E[\mathbf{x}\mathbf{x}^H] = \text{diag}(p_1, p_2, \dots, p_{M_T})$ subject to $\sum_i^{M_T} p_i = P_T$, where P_T is the total transmit power. The channel response between the transmitter and the receiver is represented by an $M_R \times M_T$ MIMO channel matrix as

$$\mathbf{H} = \begin{pmatrix} h_{11} & \cdots & h_{1M_T} \\ \vdots & \ddots & \vdots \\ h_{M_R1} & \cdots & h_{M_R M_T} \end{pmatrix} \quad (3)$$

where h_{mn} ($1 \leq m \leq M_R, 1 \leq n \leq M_T$) is modeled as a complex Gaussian variable with zero-mean and unit variance representing the channel response between the n^{th} transmit antenna and the m^{th} receive antenna.

However the tangled MIMO channel is inadequate for the UEP scheme of weighted FMO slices. Incidentally, the eigenbeamforming matrix decomposes the original $M_R \times M_T$ channel matrix into $B = \text{rank}(\mathbf{H})$ decoupled eigen subchannels [6], so that information carried along each eigen subchannel will not interfere with each other at the receiver. Transmitting the bit-stream of each FMO slice over the subchannel and controlling the power of each signal according to the visual weight and the subchannel gain, the UEP scheme can be naturally accomplished.

The optimal transmit beamforming scheme is a singular value decomposition (SVD) algorithm[6], where the transmit beamforming matrix is obtained through SVD of the channel matrix \mathbf{H} which is $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where \mathbf{U} and \mathbf{V} are unitary matrix and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_B)$ is the $B \times B$ diagonal matrix containing singular values. With the full channel state information at the transmitter, the signal vector \mathbf{x} of dimension $M_T \times 1$ is multiplied by the right singular matrix \mathbf{V} prior to transmission. At the receiver, the received signal vector \mathbf{y} is multiplied by the matrix \mathbf{U}^H . Then, the MIMO channels can be decomposed into B parallel spatial subchannels (eigen channels) given by $\tilde{\mathbf{y}} = \mathbf{U}^H \mathbf{H} \mathbf{V} \mathbf{x} + \mathbf{U}^H \mathbf{n} = \mathbf{\Lambda} \mathbf{x} + \tilde{\mathbf{n}}$, where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{n}}$ denote the transformed received signal vector and noise vector, respectively. Equivalently, decomposed spatial subchannels can be rewritten as

$$\tilde{y}_i = \lambda_i s_i + \tilde{n}_i, \quad i = 1, \dots, B \quad (4)$$

The capacity of the MIMO channel is the sum of the individual parallel subchannel capacities and is given by

$$\mathcal{C} = \sum_{i=1}^B \log_2 \left(1 + \frac{\lambda_i^2}{N_o} p_i \right). \quad (5)$$

The optimal power allocation for maximizing the data throughput, well-known WF algorithm [6] can be applied to each spatial subchannels.

4. VISUAL DISTORTION OPTIMIZATION

Using the visual weight in (1) and the RD model of Gaussian source in [7], the visual distortion-rate model for the l^{th} slice can be defined as

$$D_l^{w_l}(R_l) = w_l \cdot M_l \cdot \bar{\sigma}_l^2 \cdot 2^{-2R_l}, \quad (6)$$

where M_l and $\bar{\sigma}_l^2$ are the number of macro-block and the average standard deviation for the l^{th} slice, respectively.

Note that the total number of FMO slice groups is equal to $B = \text{rank}(\mathbf{H})$. Assuming that a source codec compresses the video source not to exceed link capacity, the rate of the l^{th} slice group becomes as below,

$$R_l = \psi \cdot \log_2 \left(1 + \frac{\lambda_l^2}{N_o} p_l \right), \quad (7)$$

where ψ ($0 \leq \psi \leq 1$) is used to make the ideal link capacity become the achievable throughput without bit error.

The power control among B grouped FMO slices can be formulated as the following constrained optimization problem:

$$(A) \quad \min_{R_l} \sum_{l=1}^B D_l^{w_l}(R_l) \quad \text{s.t.} \quad \sum_{l=1}^B R_l \leq R_{max}$$

If we consider the visual distortion rate model and the link capacity formula for the MIMO system simultaneously, in other words substituting (7) to (6), our optimization problem can be transformed as follows:

$$(B) \quad \min_{p_l} \sum_{l=1}^B w_l M_l \bar{\sigma}_l^2 \left(1 + \frac{\lambda_l^2}{N_o} p_l \right)^{-2\psi}$$

$$\text{s.t.} \quad \sum_{l=1}^B p_l \leq P_T, \quad p_l \geq 0, \quad \forall l$$

The solution of problem (B) is an optimal power set, $\{p_1, p_2, \dots, p_B\}$ for each spatial subchannel. Because problem (B) is a convex optimization, we can apply to the (Karush Kuhn Tucker) KKT condition with respect to p_k to obtain an optimal power set which is a globally optimum solution.

Using a Lagrangian relaxation,

$$L(p_k, \nu) = \sum_{l=1}^B w_l M_l \bar{\sigma}_l^2 \left(1 + \frac{\lambda_l^2}{N_o} p_l \right)^{-2\psi} + \nu \left(\sum_{l=1}^B p_l - P_T \right)$$

where ν is a nonnegative Lagrangian multiplier. Taking the derivatives with respect to p_k and ν , respectively, gives the KKT conditions as follows:

$$-2\psi w_l M_l \bar{\sigma}_l^2 \cdot \left(1 + \frac{\lambda_l^2}{N_o} p_l^* \right)^{-2\psi-1} \frac{\lambda_l^2}{N_o} + \nu = 0, \quad (8)$$

$$-P_T + \sum_{l=1}^B p_l \geq 0, \quad \nu \left(\sum_{l=1}^B p_l^* - P_T \right) = 0, \quad \nu \geq 0. \quad (9)$$

From (8), if power p_l is allocated to the l^{th} data stream (that is, $p_l^* \geq 0$), then the complementary slackness condition is

satisfied. As a result, the optimal values of p_l^* is given by

$$p_l^* = \frac{N_o}{\lambda_l^2} \left[\left(\frac{2w_l M_l \psi \bar{\sigma}_l^2 \lambda_l^2}{N_o} \nu^{-1} \right)^{\frac{1}{2\psi+1}} - 1 \right]. \quad (10)$$

To obtain multiplier ν , use (9), when $\nu > 0$,

$$\begin{aligned} \sum_{l=1}^B p_l^* &= \sum_{l=1}^B \frac{N_o}{\lambda_l^2} \left[\left(\frac{2w_l M_l \psi \bar{\sigma}_l^2 \lambda_l^2}{N_o} \nu^{-1} \right)^{\frac{1}{2\psi+1}} - 1 \right] \\ &= \sum_{l=1}^B \frac{N_o}{\lambda_l^2} \left(\frac{2w_l M_l \psi \bar{\sigma}_l^2 \lambda_l^2}{N_o} \right)^{\frac{1}{2\psi+1}} \nu^{-\frac{1}{2\psi+1}} - \sum_{l=1}^B \frac{N_o}{\lambda_l^2} \\ &= P_T \end{aligned} \quad (11)$$

Therefore, the multiplier ν can be calculated as below,

$$\nu = \left[\frac{\left(\sum_{l=1}^B \frac{N_o}{\lambda_l^2} \left(\frac{2w_l M_l \psi \bar{\sigma}_l^2 \lambda_l^2}{N_o} \right)^{\frac{1}{2\psi+1}} \right)^{2\psi+1}}{\left(P_T + \sum_{l=1}^B \frac{N_o}{\lambda_l^2} \right)} \right] \quad (12)$$

5. SIMULATION RESULTS

In the simulation, the CIF ‘‘Foreman’’ sequence with 15 fps is compressed using the H.264/AVC codec. The first frame is coded as I-frame and the remaining frames are coded as P-frame. The FMO of Type 2 [4] is utilized to packetize MBs into two slice groups ($B = 2$) according to the visual importance. Since we assumed that the foveation point is the center of the face, each frame can be divided into the ROI ($l = 1$) and non-ROI slices ($l = 2$). Also, we use a 2x2 MIMO system under an independent Rayleigh fading channel. The elements of the MIMO channel matrix \mathbf{H} are obtained from Clarke / Jakes’s model [8]. It is assumed that the channel is known perfectly at both the transmitter and the receiver.

To show that the proposed algorithm is excellent for minimizing visual distortion, we compare the proposed algorithm with the WF algorithm which is the optimal power allocation algorithm in the SVD-based MIMO spatial subchannel transmission [6]. Fig. 2 and Fig. 3 present the ergodic capacity¹ and the visual distortion according to the different SNRs, respectively. The proposed scheme is worse than the WF algorithm in terms of capacity. However, if we use the conventional WF algorithm, the visual importance of each slice cannot be considered and FMO is not applicable. It means that the WF algorithm can maximize system capacity, but it cannot minimize the distortion in aspect of visual quality. Especially, it is interesting that our proposed scheme can directly minimize the total visual distortion and can also improve overall video quality in the face of some system capacity loss at the low SNR region. This result is well represented in Fig. 3.

¹The term ‘‘capacity’’ means the tightest upper bound on the amount of information that can be achieved with arbitrarily very small error probability. In this simulation, we don’t consider the bit error but the quantization error.

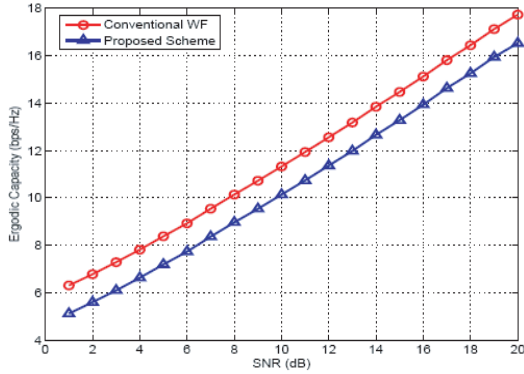


Fig. 2. Ergodic capacity as a function of SNR.

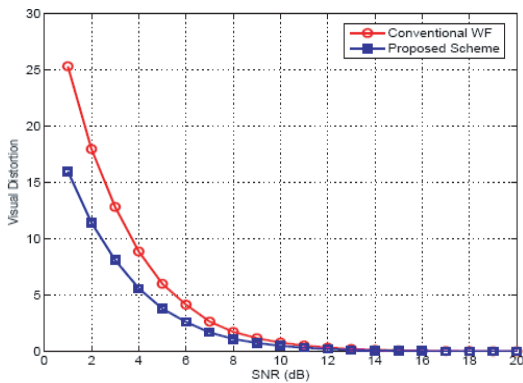


Fig. 3. Visual distortion as a function of SNR.

As aforementioned, despite the proposed scheme has loss in the capacity compared with the WF, the visual distortion of image can be minimized at low SNR conditions.

Fig. 4 shows the simulation results when the visual weights of ROI and non-ROI slice are 0.7 and 0.3 respectively. Fig. 4 (a) and (b) are the reconstructed image when the WF algorithm and the proposed scheme are used, respectively. In the WF algorithm, Fig. 4 (a), the ROI slice has severe distortion since every slice has the same visual importance. Even if the capacity is a bit larger, the overall visual quality is not better due to the distortion of ROI slice. Whereas, in the proposed scheme, the ROI slice is relatively well protected. At the same time, the overall quality is also fairly good due to properly allocated resources. So, it is easily observed that the visual quality of the proposed scheme, Fig.4 (b), is better than Fig.4 (a). In the objective quality assessment, SSIM and PSNR, the proposed scheme has also the best results.

6. CONCLUSION

This paper describes the optimal power level of each subchannel was determined by joint optimization between the rate distortion (RD) of the slices in the video frames and the link-



Fig. 4. Reconstructed Image of (a) WF (without FMO), (b) Proposed Scheme (with FMO). SNR=2dB.

capacity of the corresponding MIMO spatial subchannels to minimize the visual distortion. In the simulation, the proposed optimization technique markedly improved the video quality, particularly at the low SNR channel condition. Thus, we can conclude that if the source and channel are jointly considered for optimization, the QoS level of visual communications will be improved.

7. REFERENCES

- [1] U. Jang, H. Lee and S. Lee, "Optimal Carrier Loading Control for the Enhancement of Visual Quality over OFDMA Cellular Networks," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1181-1196, Oct. 2008.
- [2] S. K. Bandyopadhyay, S. Prasad D. G. and L. P. Kondi, "Cross-Layer Optimization for H.264/AVC Video Transmission Over Space-Frequency (SF) Coded MIMO-OFDM Systems," *IWCMC*, Aug. 2008, pp. 413-418
- [3] D. Song and C. W. Chen, "Scalable H.264/AVC Video Transmission Over MIMO Wireless Systems With Adaptive Channel Selection Based on Partial Channel Information," *IEEE Trans. Circuits Syst. Video Technol*, vol. 17, no. 9, pp. 1218-1226, Sept. 2007.
- [4] P. Lambert, W. De Neve, Y. Dhondt, and R. Van de Walle, "Flexible Macroblock Ordering in H.264/AVC," *Journal of Visual Communication and Image Representation*, Elsevier, pp. 358-375, Aug. 2005.
- [5] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, pp. 149-162, Feb. 2003.
- [6] I. Telatar, "Capacity of multi-antenna Gaussian channels", *Technical Report*, AT & T Bell Laboratories, 1995.
- [7] T. M. Cover, and J. A. Thomas, *Elements of Information Theory*, Philadelphia, PA:Wiley Interscience, 1991.
- [8] W. C. Jakes, *Microwave Mobile Communications*. New York:Wiley, 1974.