In Proc. of PW-EI-VCIP'98, San Jose, CA, USA, January 1998

Dependent Quantization for Stereo Image Coding

Woontack Woo and Antonio Ortega

Integrated Media Systems Center
Department of Electrical Engineering-Systems
University of Southern California,
Los Angeles, California 90089-2564
Email: wwoo@sipi.usc.edu and ortega@sipi.usc.edu

ABSTRACT

In this paper, we address the problem of optimal bit allocation for stereo images. Conventional rate-distortion (RD) based methods have mainly concentrated on minimizing total distortion within a given bit budget by independently encoding each image. However, stereo image coding, like video coding, requires a dependent bit allocation framework to further improve encoding performance because *binocular* and *spatial* dependencies are introduced by the disparity estimation and differential pulse coded modulation (DPCM) of the disparity vector field. We first formulate the dependent bit allocation problem for stereo image coding and extend it to blockwise dependent bit allocation. We then focus on the blockwise dependent quantization because using open-loop disparity estimation decouples the dependent bit allocation problem into two independent problems; *disparity estimation* and *dependent quantization*. The encoding complexity and delay in the dependent quantization framework can be significantly reduced by exploiting the unidirectional binocular dependency. An optimal set of quantizers can be selected using the Viterbi algorithm. For a given three quantization scales, the proposed scheme provides higher PSNR, about $3dB$ compared to JPEG without disparity compensation and $0.5dB$ compared to optimal independent blockwise quantization with disparity compensation. The proposed scheme can help develop a fast and efficient bit allocation strategy, be a benchmark of practical rate control schemes or be used in asymmetric applications, which may involve offline encoding, such as CD-ROM, DVD, video-on-demand, etc.

Keywords : stereo image coding, dependent bit allocation, blockwise quantization, Viterbi algorithm

1 INTRODUCTION

The usage of stereoscopic images and videos is becoming popular because of the increasing demand for more realistic 3D imaging systems. The 3D systems have a number of applications in visualization (CAD/CAM/medical data), telemedicine, telerobotics, telepresence, TV, cinema, and virtual reality (VR). The recent developments of autostereoscopic displays requiring no annoying glasses are accelerating the usage of stereoscopic images and videos. However, the price for this added realism is doubling of the data as compared to monocular cases. As a result, stereo image/video coding has attracted considerable attention over last few years because the channel bandwidth limitation, as for a monocular image/video, is the main bottleneck for realizing 3D systems.[1–7]

Disparity estimation (DE) and compensation (DC) in stereo image coding are similar to the motion estimation (ME) and compensation (MC) in video coding. In both cases, block-based methods have been widely used because they are simple to implement and exploit the spatial redundancy in the displacement (disparity or motion) vector (DV) field, though the true DV field is obviously not blockwise constant. In addition, both are used to exploit the similarity between two images and thus the predictive encoding schemes introduce dependencies; *binocular dependency* in stereo images and *temporal dependency* in video. For example, in video coding, quantization decisions on some reference frames affect other frames, due to the use of closed-loop (temporal) predictive coding. In dependent coding framework, the choice of RD point for the reference frame affects the operating RD points for others. Therefore, without loss of generality, we can presume that, in stereo image coding, an optimal independent bit allocation of the reference image may not be the optimal, as it affects the performance of the target image.

However, there are wide differences between DE and ME. While the temporal dependency has $2D$ direction, the binocular dependency exists only along the horizontal ($1D$) direction, if the epipolar constraint is satisfied as usually assumed in the stereo geometry.[8] In addition, the DV field and the disparity compensated difference (DCD) frame require relatively higher bit rates because the true DV field of a block is less uniform, with many pixels having different nonzero DVs, due to the stereo camera geometry. Meanwhile, a typical MV field may have either zero or near-zero MVs except moving objects. In addition, the energy of the motion compensated difference (MCD) frame is relatively lower, *i.e.*, the changes between consecutive images are relatively small, except for blocks containing boundaries of moving objects in video sequences. The DV field and the DCD frame are also subject to the influence of various noise effects, as well as occlusions, which increase high frequency components in the DCD, because stereo images are obtained from two cameras. Therefore, from a compression perspective, an efficient bit distribution scheme is even more important for coding of stereo images/video.

RD-constrained methods have been used in video coding to increase coding efficiency. However, conventional RD-based approaches have mainly concentrated on minimizing total distortion within a given bit budget by independently encoding each image.[9] Schuster *et al* addressed optimal bit allocation problem for the target image, *i.e*, between the MCD frame and the MV field in the frame level using the dynamic programming framework.[10] Chen *et al.* showed how to estimate an optimal RD-constrained MV field using the Viterbi algorithm.[11] Also, the dependent bit allocation framework was formalized for the temporal dependency in video by Ramchandran *et al.*[12] However, the blockwise dependency introduced by the ME is very complicated to exploit because an interframe coded block has a $2D$ dependency with respect to blocks in the previous frame. For example, a given block in the current frame may be predicted based on up to 4 blocks in the previous frame, and thus quantizer choices for those 4 blocks affect the performance of the current block. Moreover, some of 4 blocks in the previous frame may also affect other blocks in the current frame.

We extend the dependent bit allocation framework to blockwise dependent bit allocation problem of stereo image coding, where the goal is to allocate rates efficiently among the reference image, the DV field, the DCD frame and the quantizers such that the overall distortion is minimized. By exploiting the binocular dependency, *i.e.*, simple $1D$ dependency, we can construct a compact dependent tree, *trellis*, and then find optimal blockwise dependent quantizers for the stereo pair using Viterbi algorithm. The proposed scheme can help develop a fast and efficient bit allocation strategy, which is essential to maintain high (perceptual) image/video quality for the available bit budget, especially for low bit rates. The proposed scheme can also be a benchmark of practical rate control schemes or be used in asymmetric applications, which may involve offline encoding, such as CD-ROM, DVD, video-on-demand, etc. In particular, it can be useful for coding applications where encoding is done just once but many users will access and decode the data, *e.g.*, storage of stereo data in the WWW.

This paper is organized as follows. In Section 2, we first formulate a general dependent bit allocation problem for stereo image coding. Then, we extend the framework to blockwise dependent bit allocation and describe how to find an optimal quantizer assignment using the Viterbi algorithm, for a given arbitrary set of quantizer, in Section 3. As an example, the proposed framework is applied to a case of three quantization scales and experimental results are provided to support the efficiency of the proposed scheme in Section 4. Finally, we discuss the results and give directions for future work in Section 5.

2 DEPENDENT BIT ALLOCATION

We formulate a dependent bit allocation problem for stereo image coding aimed at properly distributing the given bits between stereo image pairs such that the overall distortion is minimized. Figure 1 shows the block diagram of a stereo image codec using closed-loop DE. The general procedure of disparity-compensated coding for stereo images is as follows. Let F_1 and F_2 be the stereo image pairs, *i.e.*, the reference image and the target image, respectively. First, F_1 is independently encoded by transform coding such as a JPEG-like codec[13] or a wavelet codec. Then, the DV field is estimated based on blocks or meaningful objects to exploit binocular redundancy between blocks in F_2 and $F_1(Q_1)$, the reconstructed reference image. The DV field is encoded using DPCM and followed by entropy coding such as Huffman coding. The shape information also has to be encoded as an additional side information, if an object or shape-based DE is adopted. The DC is also performed and the resulting DCD, the difference between F_2 and $F_1(Q_1)$ along the DV, can be encoded to increase the quality of the reconstructed target image. In general, the bit rate is controlled by adjusting the quantizer (quantization scales or factors). In the decoder, first the reference image is decoded and then the target image is reconstructed by adding the disparity compensated image and the decoded DCD.

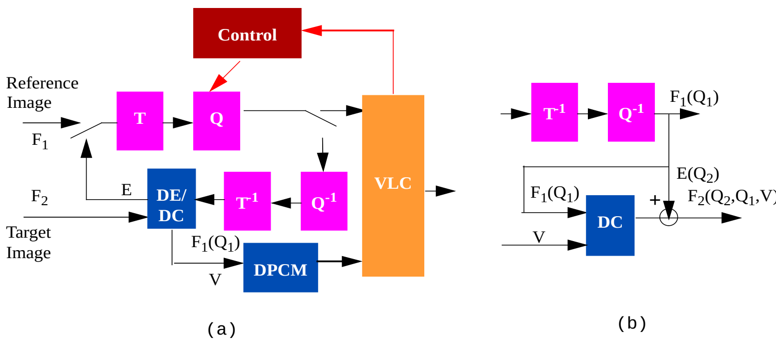

Figure 1: Block diagram of stereo image codec: (a) encoder. (b) decoder.

As shown in Figure 1, the general stereo coding problem is to estimate a DV field and to select sets of quantizers for the reference image and the DCD such that the total distortion is minimized with a given bit budget R_{budget}. Let D and R refer to the distortion and the bit rate, respectively. The problem of stereo image coding can be formulated as follows.

$$\begin{array}{ll} \text{Given} & F_1, F_2, R_{budget} \\ \text{find} & \hat{X} = (V, Q_1, Q_2) \\ \text{such that} & \hat{X} = \arg\min_X \{D_1(Q_1) + \alpha D_2(V, Q_1, Q_2)\} \\ \text{subject to} & R_1(Q_1) + R_2(V, Q_1, Q_2) \leq R_{budget} \end{array}$$

where V denotes a DV field and Q represents a set of quantizers. The weight between D_1 and D_2 can be controlled by the weighting constant α which supports two different views of the depth perception process: *fusion theory* and *suppression theory*.[14] Fusion theory claims that both images in a stereo pair equally contribute in 3D perception while suppression theory indicates that the highest quality image (or region) dominates the perception. We set α equal to one during our experiments.

The distortions of F_1 and F_2 are measured using mean squared error (MSE), *i.e.*, $D_1 = (F_1 - F_1^{Q_1})^2$ and $D_2 = (F_2 - F_2^{Q_1,Q_2,V})^2$, where F^Q denotes the decoded image. The decoded target image, $F_2^{Q_1,Q_2,V}$, can be reconstructed by adding up the decoded reference image and the decoded DCD, *i.e.*, $F_2^{Q_1,Q_2,V} = F_1^{Q_1}(V) + E^{Q_2}$, where E is defined as the difference between the target image and the compensated image from the reference image with DV. Note that while the closed-loop DE minimizes $E = F_2 - F_1^{Q_1}(V)$, the open-loop DE minimizes $F_2 - F_1(V)$.

In general, the constrained RD optimization problem can be transformed into an unconstrained problem using a Lagrangian method. The Lagrangian cost for the dependent bit allocation can be defined as follows

$$\begin{aligned} J(\lambda) &= J_1(Q_1) + J_2(Q_1, Q_2) = D + \lambda R \\ &= D_1(Q_1) + \alpha D_2(Q_1, Q_2) + \lambda\{R_1(Q_1) + R_2(Q_1, Q_2, V)\} \end{aligned} \quad (1)$$

where λ is a Lagrange multiplier. In a practical lossy data compression scheme, only a finite number of operational RD pairs are possible for a given source because only a finite set of quantizers is available. Under this assumption, the optimal operating RD points can be searched for the fixed slope λ. Note that the independent bit allocation problem for the stereo images can be considered as a special case of the dependent bit allocation, *i.e.*, $D_2(Q_1, Q_2) = D_2(Q_2)$ and $R_2(Q_1, Q_2, V) = R_2(Q_2, V)$.

Figure 2 shows an operational RD plot to help motivate the need for a dependent bit allocation framework. Note that for a given λ and three operational RD points, the RD optimal quantizer for the reference image is Q_{1b} because the Lagrangian cost $J_1(Q_{1b})$ is smaller than others. However, when stereo images are considered simultaneously, Q_{1a} can be an optimal quantizer for the reference image because the summed Lagrangian cost $J_1(Q_{1a}) + J_2(Q_{1a}, Q_{2b})$ can be smaller than $J_1(Q_{1b}) + J_2(Q_{1b}, Q_{2b})$. This assertion can similarly be extended to the blockwise dependent quantization in stereo image coding.

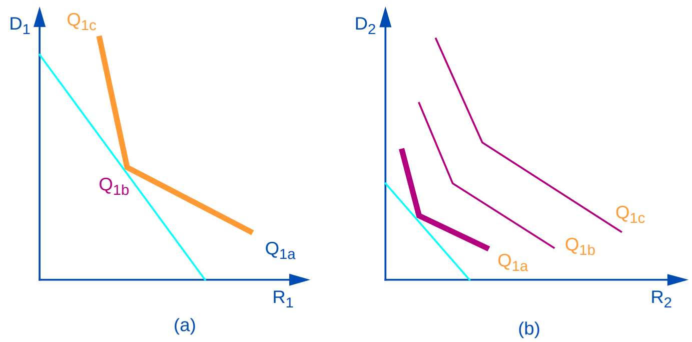

Figure 2: Dependent bit allocation in operational RD plot. (a) independent bit allocation: a given λ, the quantizer Q_{1b} is an optimal because the Lagrangian cost $J_1(Q_{1b})$ is smaller than others. (b) dependent bit allocation: if stereo pairs are considered together, there is a chance for the quantizer Q_{1a} to be an optimal, because the summed Lagrangian cost $J_1(Q_{1a}) + J_2(Q_{1a}, Q_{2b})$ can be smaller than $J_1(Q_{1b}) + J_2(Q_{1b}, Q_{2b})$.

3 DEPENDENT QUANTIZATION

3.1 Blockwise Dependent Bit Allocation

Figure 3 shows that each block in the image has different importance according to its contents, *i.e.*, different spatial activity that requires a different amount of bits. Thus, equal bit allocation may result in discontinuity of (perceptual) quality between adjacent blocks. Therefore, we need to allow blockwise bit rate control to optimize overall quality of the image/video. We assume that, for a given finite set of quantizers, a different quantizer (or quantization scale) can be assigned per block, as is the case in current standards such as H.263x.

Let the target image be segmented into square blocks, $F_2 = \{B_n, 0 \le n \le N-1\}$ where N is the total number of blocks. Then, each block B_n has a disparity vector v_n and a quantizer q_n. The total rate and distortion for the independent bit allocation can be represented as the sum of the block rate, $R = \sum_{n=0}^{N-1} r_n(v_n, q_n)$, and the block distortion $D = \sum_{n=0}^{N-1} d_n(q_n)$ where r_n, d_n and q_n denote the rate, the distortion and the quantizer of the

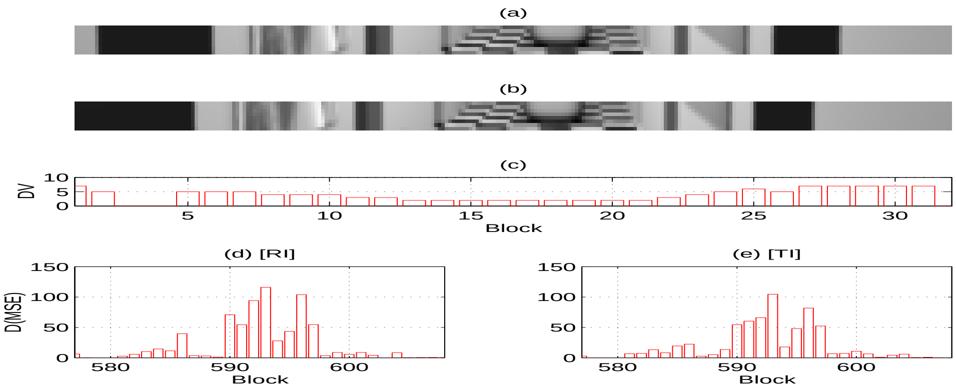

Figure 3: Binocular dependency and different spatial activities in (19*th*) GOBs: we denote blocks in each row a group of blocks (GOB) (a) a GOB in the reference image. (b) a GOB in the target image. (c) two GOBs have binocular dependency along the DV. (d) MSE vs. GOB in the reference image (RI). (e) MSE vs. GOB in the target image (TI).

block, respectively. By incorporating binocular dependency introduced by the DE, the dependent bit allocation problem is to estimate the DV field and a set of the quantizers, (Q_1, Q_2), for the reference image and the DCD frame such that the overall distortions, $(D_1 + D_2)$, are minimized with a given bit budget, R_{budget}.

In the blockwise dependent bit allocation framework, the operational RD curve of the current block in the target image depends on that of the blocks in the reference image. DPCM of the quantizer indices and the *DV* field also introduces *spatial* dependency between consecutive (or neighboring) blocks. The resulting total Lagrangian cost for the simple block matching can be represented as a sum of the Lagrangian cost of each block.

$$\begin{aligned} J(\lambda) &= D_1 + \lambda R_1 + D_2 + \lambda R_2 \\ &= \sum_{m=0}^{N-1} \{d_m(q_m) + \lambda r_m(d_m, \nabla q_m)\} + \sum_{n=0}^{N-1} \{d_n(q_n, q^\eta(v_n)) + \lambda r_n(d_n, \nabla v_n, \nabla q_n)\} \end{aligned} \quad (2)$$

where d, r and q denote distortion, rate and quantizer of a block, respectively. A vector of quantizer q^η corresponds to the quantizers of the blocks in the reference image, which are used to predict the current block in the target image.

3.2 Open-loop DE and Dependent Quantization

Using open-loop DE, we can reduce the complexity of the dependent bit allocation problem, *i.e.*, decouple the bit allocation problem into two independent problems; (*i*) RD-constrained DV field estimation[11] and (*ii*) dependent quantization.[12] Figure 4 shows the block diagram of a stereo image encoder using open-loop DE. The same decoder can be used as shown in Figure 1 (b).

The open-loop DE does not depend on the quantizer choices because it estimates the DV field minimizing the difference between F_2 and $F_1(V)$ instead of $F_1(V, Q_1)$, the reconstructed reference image. The closed-loop DE can reduce further the DCD energy because it reduces the difference between the original target image and the reconstructed reference image. However, the open-loop DE may estimate a more consistent DV field than the closed-loop DE does. The Lagrangian cost for RD-constrained DE based on the open-loop approach can be

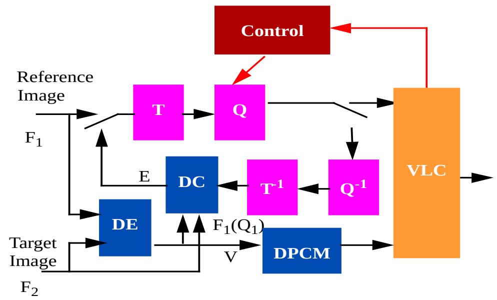

Figure 4: Block diagram of stereo image encoder using open-loop DE.

defined as follows

$$J(\lambda) = \sum_{n=0}^{N-1} \{\tilde{d}_n + \lambda r_n(\nabla v_n)\} \tag{3}$$

where $\tilde{d}$ denotes the difference between blocks in F_2 and $F_1(V)$. Note that the rate r_n is not a function of $\tilde{d}$ or q_n. The optimal DV field can be estimated using the algorithm developed for ME by Chen *et al.*[11] and encoded independently.

In this framework, we assume the quantizer indices are encoded without DPCM and entropy coding, *i.e.*, there is a constant number of overhead bits per block. Note that $1D$ dependency could be incorporated into our scheme. With a given DV field (from the open-loop DE), the distortion and the bit rate of a block in the DCD frame depend upon the block(s) in the reference image along the DV. Then, the Lagrange cost function for the dependent quantization can be reduced from (2) as follows

$$J(\lambda) = \sum_{m=0}^{N-1} \{d_m(q_m) + \lambda r_m(q_m)\} + \sum_{n=0}^{N-1} \{d_n(q_n, q^{\eta}(v_n)) + \lambda r_n(q_n)\} \tag{4}$$

where η denotes (at most) two consecutive blocks in the reference image as shown in Figure 5. The selection of a quantizer for the *nth* block in the DCD frame has to be decided according to the selection of quantizers of the blocks in the reference image and vice versa. For example, a block B_n in the DCD frame is compensated from (at most) two consecutive blocks, B_m and B_{m+1}, in the reference image along the disparity vector v_n. Therefore, the rate r_n and the distortion d_n of the B_n in the DCD frame depend only on the quantizers $q^{\eta}(v_n) = \{q_m, q_{m+1}\}$ of corresponding blocks (selected by a v_n) in the reference image, *i.e*, $d_n = d_n(q_n, q_m, q_{m+1})$.

3.3 Viterbi Algorithm for Blockwise Dependent Quantization

With a given arbitrary set of quantizers, the blockwise dependent quantization problem for stereo images can be solved using a forward deterministic dynamic programming scheme, *i.e.*, the Viterbi algorithm (VA).[15] The first step is to construct a dependent tree, *trellis*, to represent all the viable allocations for each row of blocks, which we call a group of block (GOB). With a given finite set of admissible quantizers, we define the trellis as follows to assign quantizers to all blocks in the DCD frame and the reference image.

Trellis: trellis is made of all possible paths linking the nodes in the first stage and the nodes in the last stage, *i.e.*, all possible concatenated choices of quantizers for the stereo pair.

Path: a concatenation of branches from the first stage to the final stage in the trellis. Each path corresponds to a set of quantization choices for the DCD frame and the reference image.

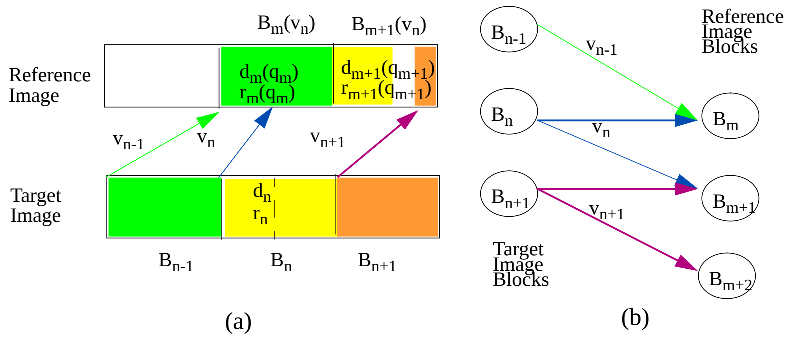

Figure 5: Binocular dependency between corresponding blocks along the disparity vector. At most two consecutive blocks in the reference image are related to a block in the target image. For example, a block B_n in the target image is compensated from (at most) two consecutive blocks, B_m and B_{m+1}, in the reference image along the disparity vector v_n. Therefore, d_n is a function of q_n, q_m and q_{m+1}.

Stage: kth stage in the trellis corresponds to the kth block in the reference image. Each stage has $L = |q_1|$ admissible decision vector set, *i.e*, L state nodes, where $|q_1|$ denotes the number of available quantizers for the reference image.

Node: each node corresponds to a possible quantizer choice for the kth block in the reference image. The choices are ordered from top to bottom in order of the finest to the coarsest. Each node i in the $(k-1)th$ stage has a Lagrangian cost, $J_i(k-1) = d_1(q_1) + \lambda r_1(q_1)$, for the block in the reference image.

Branch: An optimal quantizer of the $(k-1)th$ block in the DCD frame is assigned to a branch, connecting nodes in the $(k-1)th$ and the $(k)th$ stages. The optimal quantizer is determined by the two quantizer choices of the successive blocks in the reference image, corresponding to the connected nodes by the branch. The attached Lagrangian cost, $J_{ij}(k-1,k)$, and the total transition cost of the branch, $J^t_{ij}(k-1,k)$, is defined in (5) and (6), respectively. The total transition cost is additive over the path.

Given a finite set of quantizers the compact dependency tree can be constructed due to the *binocular dependency*, the horizontal ($1D$) dependency. As shown in Figure 3, the binocular dependency allows performing independently the blockwise dependent quantization in each GOB, because a GOB in the target image depends only on a GOB in the reference image. Therefore, the optimal dependent bit allocation scheme can be applied independently to each GOB instead of the whole image, *i.e*, for a given arbitrary set of quantizers, a trellis can be constructed independently for each GOB, without loss of generality. By comparison, in the case of video coding an optimal blockwise dependent bit allocation would require that the whole image be considered. As a result, the binocular dependency significantly reduces the encoding complexity and delay which are dramatically increased along with the dimension of the trellis, *i.e*, both the number of nodes per stage and the number of stages. Figure 6 shows the state diagram and corresponding consecutive stages in trellis for the case of three different quantizers.

The binocular dependency is also restricted to only three blocks as shown in Figure 6 (a). Let q_{11} and q_{12} be the quantizers of the consecutive blocks in the reference image. Given two quantizers, q_{11} and q_{12}, the corresponding residue along the disparity vector is completely determined, *i.e*, we can choose the optimal quantizer, q_{21}, for the residue block in the DCD frame. As shown in Figure 6 (c), two quantizers, q_{11} and q_{12}, corresponds to ith node in the $(k-1)th$ stage and the jth node in the $(k)th$ stages, respectively. The quantizer q_{21}, determined by the (q_{11}, q_{12}), is assigned to the branch connecting two nodes. The corresponding cost of the branch, $J_{ij}(k-1,k)$, is defined as follows

$$J_{ij}(k-1,k) = d_{21}(q_{21}, q_{11}, q_{12}) + \lambda r_{21}(q_{21}, q_{11}, q_{12}) \quad (5)$$

where d, r, and q correspond to the distortion, the rate, and the quantizer of a block, respectively. Therefore, each branch in the trellis corresponds to the consecutive choices of quantizers, (q_{21}, q_{11}, q_{12}), for the block in the

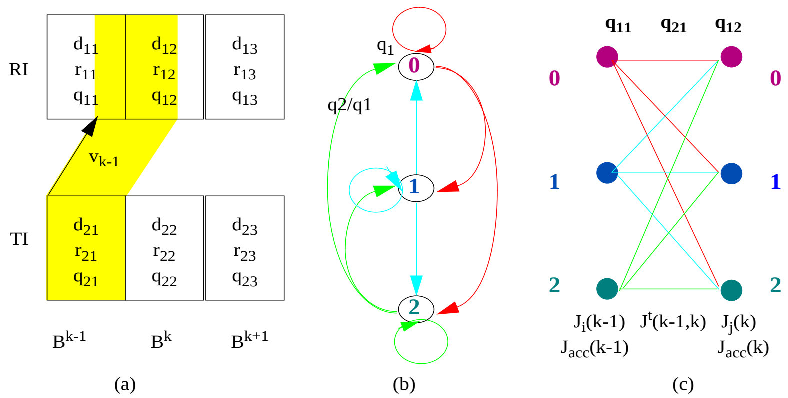

Figure 6: Example of blockwise dependent quantization for a given three different quantizers: (a) binocular dependency along the disparity vector. (b) state diagram: state node corresponds to the quantizer choice for the reference image (RI) and the transition branch to the quantization choice for the block in the DCD frame. (c) possible connections between the (k-1)th and the (k)th stages in the trellis.

DCD and the consecutive blocks the reference image. The total transition cost from the ith node in the $(k-1)th$ stage to the jth node in the (k)th stage, $J^t_{ij}(k-1,k)$, is defined as follows

$$
\begin{aligned}
J^t_{ij}(k-1,k) &= J_i(k-1) + J_{ij}(k-1,k) \\
&= d_{11}(q_{11}) + d_{21}(q_{21}, q_{11}, q_{12}) + \lambda\{r_{11}(q_{11}) + r_{11}(q_{21}, q_{11}, q_{12})\} \qquad (6)
\end{aligned}
$$

where d, r, and q correspond to the distortion, the rate, and the quantizer of a block, respectively.

Now, the blockwise dependent quantization problem is equivalent to finding the smallest cost path from a node in the first stage to a terminal node in the last stage of the trellis. A path in a trellis corresponds to the quantizer choices assigned to the two GOBs: one in the DCD frame and the other in the reference image. The Viterbi algorithm can be used in searching the minimum cost path through the trellis. The algorithm provides a globally optimal solution with a given transition cost because it sequentially compares possible paths from the initial stage to successive stages in the trellis, while pruning suboptimal paths. With a given λ, the optimal dependent quantizers, (Q_1, Q_2), can be found by applying repeatedly the following procedure to each GOB.

Step 0 Initialization: let K and L be the number of stages and nodes per stage, respectively. Add an initial node B_{-1} and a final node B_K. Select a λ and set $k = 0$ & $Jacc_0(-1) = 0$.

Step 1 At a stage $k-1$, permissible branches are added to the end of each node i (of all surviving paths) and Lagrangian costs are assigned to the node for the block in the reference image and to the branch for the block in the DCD frame. Thus the cost defined in (6), $J^t_{ij}(k-1,k)$, is the total transition cost corresponding to the branch, connecting the ith node in the $(k-1)th$ stage to jth node in the $(k)th$ stage.

Step 2 At a stage k, for each node j, an accumulated transition cost, $J_{tmp}(i)$, is calculated by summing the accumulated cost, $Jacc_i(k-1)$, and the total transition cost, $J^t_{ij}(k-1,k)$. The one with the lowest accumulated-transition-cost is chosen of all arriving branches (at most L). The resulting cost is assigned to the accumulated cost, $Jacc_j(k)$ and the rest branches are pruned.

$$J_{tmp}(i) = Jacc_i(k-1) + J^t_{ij}(k-1,k)$$
$$Jacc_j(k) = \min\{J_{tmp}(i)\}_{i=0}^{L-1}$$

Step 3 go to *step 1* and repeat until the end node is met, *i.e.*, $k = K$.

Step 4 the path with minimum total cost across all paths can be found by back tracking the surviving path.

4 EXPERIMENTAL RESULTS

The proposed blockwise quantization framework is evaluated using the synthesized image of size 256×256 in Figure 7 (a). The block size is 8×8 and the size of the search window for disparity estimation is 1×8. The disparity estimation is performed independently using fixed size block matching (FSBM) between the target image and the reference image within the search window. The DV field has to be encoded using a lossless encoding scheme because lossy coding can lead to higher distortion in the reconstructed target image or degrade 3D perceptual quality. Therefore, the resulting DV field is encoded using DPCM with a median predictor based on its neighboring DVs to further reduce the remaining spatial redundancy among neighboring DVs. The reference image and the DCD are encoded using JPEG-like coder. As shown in Figure 7 (b) & (d), DPCM significantly reduces the entropy of the DV field.

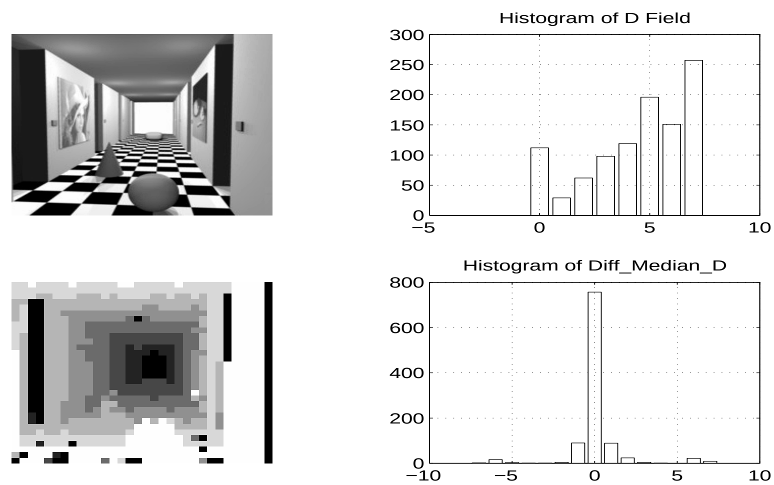

Figure 7: Comparison of histogram of the DV field. The entropy of the DV field reduced using DPCM: (a) the reference image (*provided by CVPR group at University of Bonn, Germany*) (b) histogram of the DV (c) the DV field (d) histogram of the DPCM of the DV

In this experiment, we assume that three different quantization scales (QS) for DCT coefficients are available, *i.e.*, $QS = \{70, 50, 30\}$ corresponds to the quantizer index, $q = \{0, 1, 2\}$. Each block in a GOB corresponds to a stage in the trellis. A possible combined choice of QSs for the DCD and the reference image is assigned to the state node. The number of stages, K, is Nx/B where Nx and B denote the width of the image and the block, respectively. The number of nodes per stage is $L = |QS|$, *i.e.*, $L = 3$. As a result, the dimension of the trellis for a GOB is $(L, K) = (3, 32)$, *i.e.*, 3 state nodes per a stage and 32 stages in a trellis. We repeatedly use the same structure of the trellis for each of 32 GOBs. The Lagrangian cost in (6) is assigned to the branch as a transition cost. Figure 6 shows state diagram and consecutive stages in trellis for the case of three different quantizers. Note that the size of header of the JPEG image is ignored.

The performance is measured in terms of the bit rate and the peak signal to noise ration (PSNR). The mean PSNR is used to evaluate the performance of stereo images together. The mean PSNR is defined as follows

$$PSNR_{mean} = 10 \times \log_{10}\{\frac{255^2}{(D_1 + D_2)/2}\} \tag{7}$$

where D_1 and D_2 represent MSE of the reference image and the target image, respectively.

Figure 8 shows the selected indeces of quantization scales (Q) for the reference image and the target image. The darker intensity, the finer quantizer. Figure 8 (a) & (b) denote the selected Q using the blockwise independent quantization scheme and (c) & (d) correspond to the selected Q using the proposed blockwise dependent quantization scheme, respectively. A typical bit distribution strategy (of video coding) is as follows. The given bits are allocated to the intra-frame first and the remaining bits are distributed to the displacement vector and the resulting compensated difference. This strategy generally works when the portion of bitrate for the inter-frame is relatively smaller than that of intra-frame. Note that, however, as shown in Figure 8, in the stereo case, finer quantizers are not always assigned to blocks in the reference image, which means that the quality of the reference image may not directly control the total PSNR performance. As shown, according to contents of the blocks, the DCD also can play an important role in improving the RD performance.

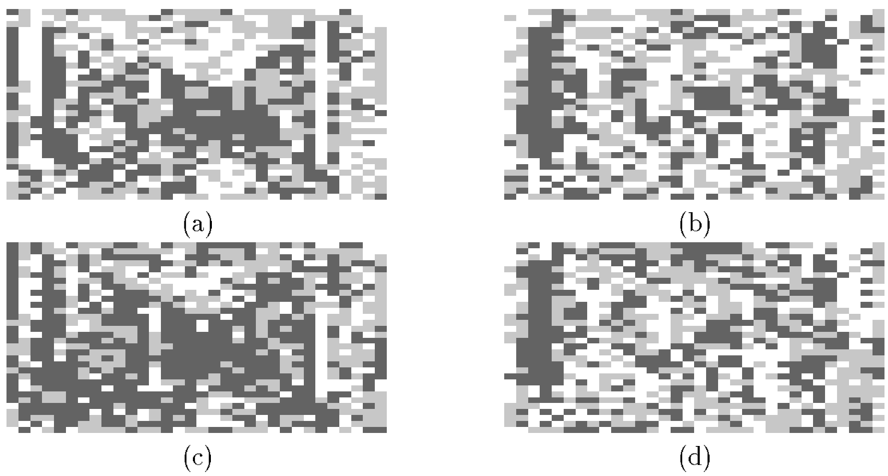

Figure 8: The resulting quantization scales ($\lambda = 0.5$): the darker intensity, the finer quantizer. (a) independent quantization (Q) selection of the reference image. (b) independent Q selection of the DCD. (c) dependent Q selection of the reference image. (d) dependent Q selection of the DCD.

Figure 9 (a) & (b) compare the achieved R-D performances for the reference image and the DCD, respectively. The star-line denotes the results of the independent JPEG without DC and the x-mark-line the results of blockwise independent quantization with DC. The circle-line corresponds to the achieved RD plot using the proposed blockwise dependent quantization scheme. Note that the RD performances for the reference image are similar in both cases, *i.e.*, the dependent quantization and the independent quantization. However, by selecting QS_1 taking dependency into account, better RD performance for the target image can be achieved than those of the independent bit allocation as shown in Figure 9 (b). Figure 9 (c) shows the comparison of RD performance in terms of the bit rate and the mean PSNR. The star-line corresponds to the independent JPEG and plus-lines to the resulting RD plot when the quantization scales are fixed for the reference image, *i.e.*, $QS_1 = \{70, 50, 30\}$, respectively. The x-mark-line and the circle-line denote the RD performance achieved using the blockwise independent and the blockwise dependent quantization scale selection for different λ. According to experimental results, the proposed blockwise dependent quantization method resulted in 3-4 dB improvement in PSNR at given bit rates compared to a fixed quantization without DC (JPEG) and 0.5 dB improvement compared to the independent blockwise quantization with DC. Note that the PSNR gain can be increased as the number of available quantizers increases.

5 DISCUSSION

We have formulated the dependent bit allocation problem for stereo image coding and extended the dependent bit allocation framework developed in video coding to blockwise dependent bit allocation of stereo images. Using open-loop DE decoupled the blockwise dependent bit allocation problem into RD-constrained DE and dependent quantization. We have shown, for a given arbitrary set of quantizers, how to find a set of optimal quantizers using Viterbi algorithm. An optimal dependent quantization leads to an efficient bit allocation between the reference image and the DCD. According to experimental results, the proposed scheme outperformed in terms of PSNR at given rates, *i.e.*, about $3dB$ compared to JPEG without DC and $0.5dB$ compared to the independent blockwise bit allocation with DC. In addition, by exploiting the unidirectional binocular dependency the computational complexity and the encoding delay of the Viterbi algorithm were significantly reduced, *e.g.*, for a given arbitrary set of quantization scales, a trellis, dependency tree, was constructed for a GOB instead of the whole image. Adopting reasonable RD models can further reduce the computational complexity of the proposed scheme. This framework can be directly extended to multiview image coding. Further research, however, is required to reduce its complexity (in the generation of operational RD data), to include spatial dependency of Qs and DVs, and to incorporate $2D$ temporal dependency. The dependency introduced by region or object-based DC is another challenging redundancy to exploit.

6 REFERENCES

[1] M. E. Lukacs. Predictive coding of multi-viewpoint image sets. In *Proc. on ICASSP 86*, pages 521–524, October 1986.

[2] I. Dinstein, G. Guy, and J. Rabany. On the compression of stereo images: Preliminary results. *Signal Processing*, 17(4):373–381, August 1989.

[3] M. G. Perkins. Data compression of stereopairs. *IEEE Tran. on Comm*, 40:684–696, April 1992.

[4] M.W. Siegel, P. Gunatilake, S. Sethuraman, and A.G. Jordan. Compression of stereo image pairs and streams. In *SPIE Proc. on SDVRS-94*, volume 2177, pages 258–268, February 1994.

[5] H. Aydinoglu, F. Kossentini, Q. Jiang, and M. H. Hayes. Region based stereo image coding. In *Proc. IEEE ICIP'95*, pages 57–60, Washington, October 1995.

[6] W. Woo and A. Ortega. Stereo image compression based on the disparity compensation using the MRF model. In *SPIE Proc. on VCIP 96*, volume 2727, pages 28–41, March 1996.

[7] W. Woo and A. Ortega. Stereo image compression based on the disparity field segmentation. In *SPIE Proc. on EI-VCIP 97*, volume 3024, pages 391–402, February 1997.

[8] B.K.P. Horn. *Robot Vision.* The MIT Press, 1986.

[9] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE ASSP*, 36:1445–1453, September 1988.

[10] G. M. Schuster and A. K. Katsaggelos. A video compression scheme with optimal bit allocation between displacement vector field and displaced frame difference. In *IEEE Proc. on ICASSP 96*, volume 4, pages 1966–1969, Atlanta, May 1996.

[11] M.C. Chen and A.N. Willson Jr. R-D optimal motion estimation algorithm for video coding. In *Proc. IEEE ICASSP'96*, volume 4, pages 2096–2099, Atlanta, May 1996.

[12] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Tr. on IP*, 3(5):533–545, September 1994.

[13] W. Pennebaker and J. Mitchell. *JPEG Still Image Compression Standard.* Van Nostrand Rheinhold, 1994.

[14] B. Julesz. *Foundations of Cyclopeon Perception.* The University of Chicago Press, 1971.

[15] G.D. Forney. The Viterbi algorithm. *The Proc. IEEE*, 61:268–278, March 1973.

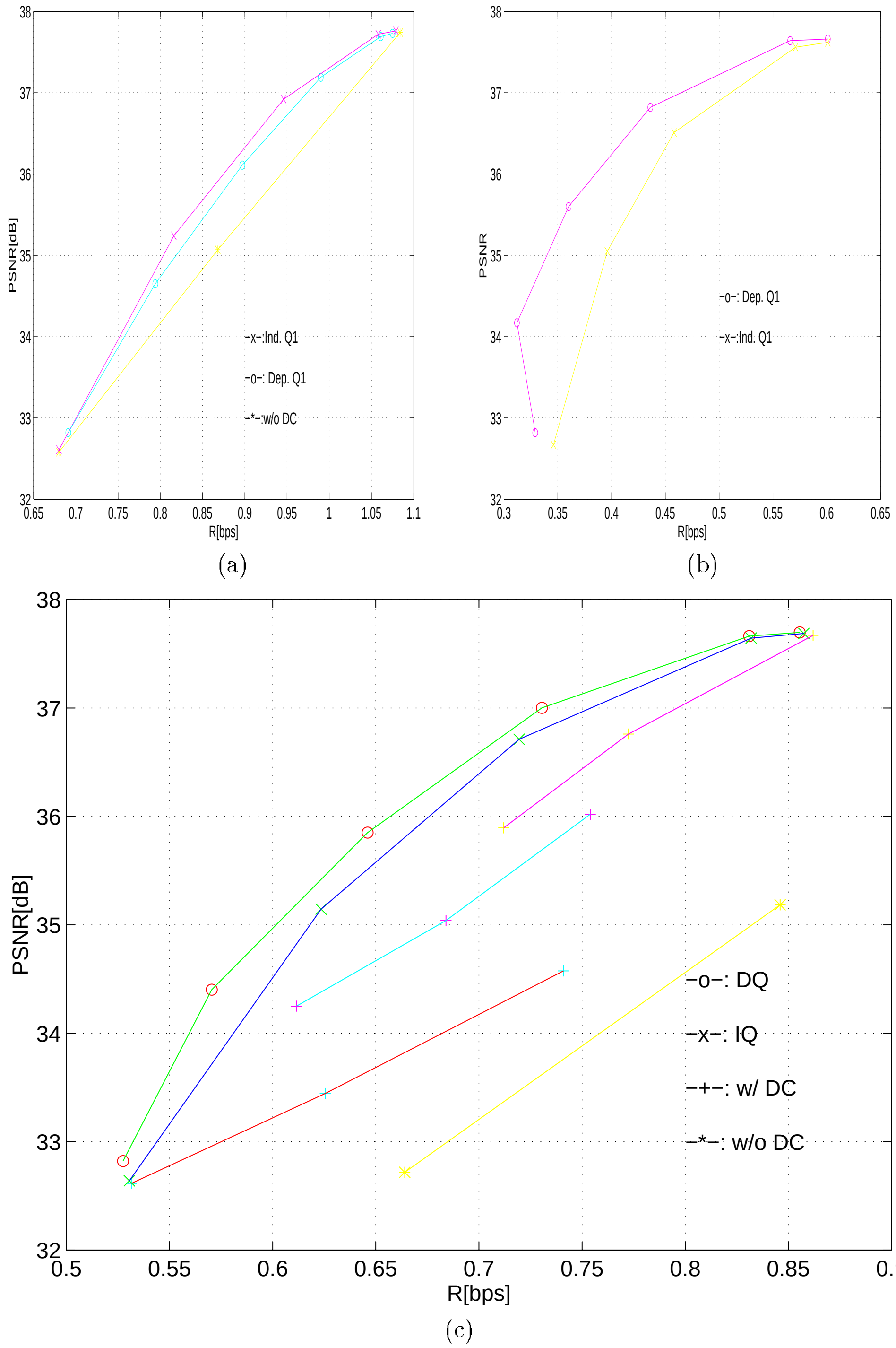

Figure 9: RD performance comparison ($\lambda = \{0, 0.1, 0.5, 1, 2, 100\}$). The star-line denotes the results of the JPEG without DC and the x-mark-line blockwise independent QS selection. The circle-line corresponds to the results of blockwise dependent QS selection. (a) blockwise dependent quantization of the reference image. (b) blockwise dependent quantization of the reference image. (c) mean PSNR vs. rate. The plus-lines denote the resulting RD plot when the quantization scales are fixed for the reference image, *i.e.*, $QS_1 = \{70, 50, 30\}$, respectively.