Optimal Blockwise 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

CORRESPONDING AUTHOR:
Antonio Ortega
Phone: (213) 740-2320
Fax: (213) 740-4651
E-mail: ortega@sipi.usc.edu

This work was supported in part by the National Science Foundation under grant MIP-9502227 (CAREER) and the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency. Parts of this work appeared in [1].

Abstract

Many stereo image compression algorithms exploit the redundancy between the two images in a stereo pair by using disparity compensated prediction. Thus, research in coding of stereo images has focused mostly on the issue of disparity estimation, with less attention being devoted to the equally important problem of allocating bits among the two images. This bit allocation problem is complicated by the dependencies arising from using a prediction based on the quantized reference images. In this paper, we address the problem of blockwise bit allocation for coding of stereo images. The goal is to select the quantization parameters for each block in the reference and difference images so as to minimize some averaged distortion measure, while meeting any applicable bit budget constraints. In this paper we show how, given the special characteristics of the disparity field, one can achieve an optimal solution with reasonable complexity, whereas in similar problems in motion compensated video only approximate solutions are feasible. The key observation is to note that the disparity field is composed of vectors with only (or predominantly) horizontal components. Thus, unlike the motion compensated case, blockwise dependencies are limited to a single dimension, i.e. the blocks in one line in the target image depend only on blocks in the same line in the reference image. Under these conditions we present algorithms based on dynamic programming that provide the optimal blockwise bit allocation. With our experiments based on a modified JPEG coder we show gains over standard, independent, bit allocation techniques. For example, the proposed scheme provides higher PSNR, about 1-2dB compared to constant quantization in the whole frame and 0.2-0.5dB compared to disparity compensation with independent blockwise quantization. We also propose a fast algorithm that provides most of the gain at a fraction of the complexity.

Keywords

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

I. Introduction

The usage of stereoscopic images/video is becoming increasingly popular as demand grows for more realistic 3D imaging systems in a variety of applications such as visualization (CAD/CAM/medical data), telecommunication (telemedicine, telepresence), telerobotics (remote control, autonomous navigation, surveillance), entertainment (interactive HDTV and cinema) or Virtual Reality. A wider deployment of stereo systems has always been limited by the requirement of inconvenient stereo glasses. Thus recently introduced technologies for autostereoscopic displays are likely to contribute to a wider usage of stereo techniques. As in the case of monocular images bandwidth or storage limitations have to be taken into account (with in this case a doubling of the data rates) thus requiring efficient compression techniques [2], [3], [4].

As in other coding scenarios, compression can be achieved by taking advantage of redundancies in the source data (e.g., spatial and temporal redundancies for monocular images and video). In the case of stereo images and video an additional source of redundancy stems from the similarity between the images in a stereo pair. In this paper, we will assume that "generic" transform coding and motion estimation are used to exploit the spatial and temporal redundancies, and will focus on the issues that are specific of disparity compensated coding.

As shown in Figure 1 the basic idea in block-based disparity estimation and compensation (DE/DC) is to use one of the images in the stereo pair as a reference (F_1) and to try to estimate the other image (the target, F_2) by finding for each block in the target image the block in the reference that best matches it [5], [6]. Note that the principle is analogous to that behind motion estimation and compensation and thus many of the intuitions and techniques used in motion estimation are directly applicable to disparity estimation [7]. Since the goal is not to estimate the true disparity but rather to achieve a high compression ratio it may not be worthwhile to compute a dense disparity field if the cost of transmitting the disparity vector (DV) field is too high. For this reason, and due to their comparative simplicity and robustness, we are focusing on block-based, rather than segmentation based, techniques.

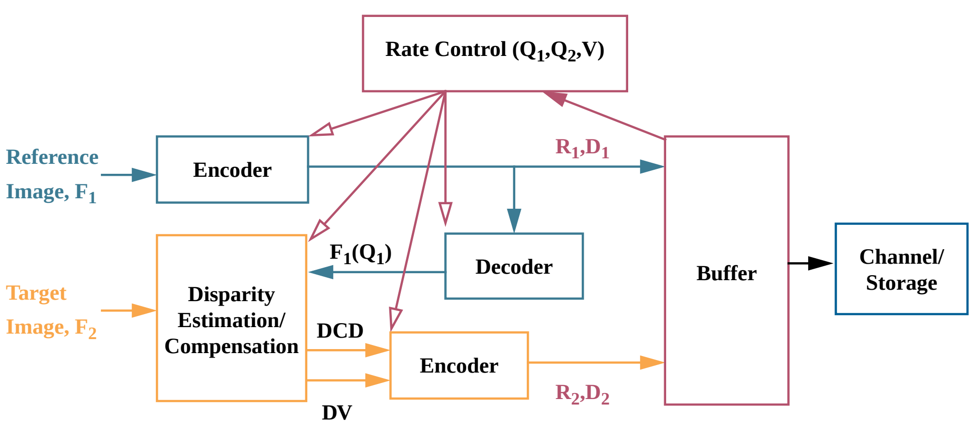


Fig. 1. Block diagram of a general encoder for stereo images, where encoder consists of disparity estimation/compensation, transform/quantization and entropy coding.

There is one significant difference between motion and disparity compensation, though. This difference lies in the fact that, if the cameras meet the epipolar constraint[1], disparity occurs only along the horizontal direction, i.e. a particular object will appear in the two images with only a horizontal shift between its respective positions. By comparison, in the motion compensation case one can observe motion vectors with any direction in the 2D plane. This property not only simplifies the matching process but, as will be shown, also enable us to find optimal solutions to our allocation problem.

Various stereo coding techniques have been proposed since Lukacs introduced block-based DE/DC [5]. While Dinstein *et al.* proposed a compression method based on the frequency domain relationship without DE [8], [9], most research efforts in stereo image/video coding, have been based on DE and thus a great deal effort has been devoted to investigating efficient DE schemes. Examples include DE in DCT domain [6] or sub-band domain [10], DE using Markov Random Fields (MRF) models [11], [12], hierarchical segmentation-based DE [13], multiresolution-based DE [14], [15], pixel-based DE with object-based coding [16], [17]. There is of course a wealth of work in motion estimation, which has relevance to the DE/DC problem. In particular, techniques developed in rate-distortion (RD) based ME in video coding [18], [19], [20], [21], [22], [23] can be used to estimate optimal (in an RD sense) choices of disparity.

With few exceptions (e.g. [24]) quantization and bit allocation issues specific to stereo coding have rarely been considered, with the usual approach being to rely on methods developed for motion compensated video coding. In this paper we study the problem of optimal blockwise bit allocation for stereo image coding. We show that this is a dependent bit allocation problem [25], since we predict the target image based on the quantized reference image and thus choices of quantizer for the reference frame result in different residual energy levels in the difference frame, i.e. $F_1(Q_1)$ is used to predict F_2. As will be shown, if bit allocation is performed independently for the reference and difference frames the overall performance can be suboptimal.

A similar situation arises in video coding as choices of quantization for a reference frame affect the frames that are motion predicted from it [25]. However, in the case of video

[1] This constraint implies that the focal rays of the two cameras are parallel and perpendicular to the stereo baseline.

coding it is difficult to take into account blockwise dependencies because each block in the predicted frame depends on up to four blocks in the reference, and, conversely, blocks in the reference frame affect several blocks in the target image. This 2D temporal dependency has led to much of the work concentrating on analyses of framewise dependency, i.e. where a single quantizer is allocated per frame [25], [26]. Also note that schemes such as [27], [28] have addressed the quantizer allocation within a frame, where quantizer choices are dependent in that lossless DPCM is used to encode the quantization indices. However in these cases the assumption was that each frame was coded independently, without taking into account the effect of a particular allocation on future frames.

In Figure 1 the encoding performance can be controlled by the choices of DV and quantizers (Q_1, Q_2). The simplest approach to select these parameters would be as follows. First, F_1 is independently compressed up to a desired quality level. Then, the DV field is estimated by computing the best match in the reference frame for each block in the target frame. Disparity compensation is performed and the resulting disparity compensated difference frame (DCD), i.e. the difference between F_2 and $F_1(Q_1)$ displaced according to the DV, is then encoded. At the decoder, the reference image is decoded and then the target image is reconstructed by adding the disparity compensated image and the decoded DCD. This simple approach completely decouples all the encoding steps and therefore there is no guarantee that the allocation of bits to the various components in any way efficient.

The main novelty of our work is the introduction of *an optimal blockwise dependent allocation scheme* for stereo image coding. We emphasize that the related problem of blockwise dependent bit allocation in video coding *has not been solved exactly* and thus our results may also provide some ideas for approximate solutions for that case. Given that the epipolar constraint is met, the dependency between frames is strictly one dimensional and therefore optimal allocation can be performed[2]. Note that we will assume that the DV field is fixed and thus do not address RD optimized DE. This problem could be approached with simple modifications of the techniques developed for RD optimized Motion Estimation [22], [20], [23]. Also we will assume that the DE is performed "open loop", i.e. based on

[2]Note that if the parallel axis constraint is not strictly met, the disparity is not exactly 1D, but it is *predominantly* 1D, i.e. while there may exist some vertical disparity this is confined to plus/minus a few pixels and thus the dependency comes from mostly from blocks located in the corresponding row in the reference frame.

the original image F_1 rather than the quantized version $F_1(Q_1)$. We will thus concentrate only the quantizer allocation to F_1 and the residue image, and not the DE itself.

Due to the simple $1D$ dependency, we can represent all possible allocations for blocks in the same row in reference and target images by constructing a *trellis*. The costs to the branches and nodes of the trellis correspond, respectively, to the target and reference blocks. We demonstrate how the optimal set of quantizers can be determined using the Viterbi algorithm and in addition we introduce novel methods, which approximate the optimal solution with, limited loss in performance but much faster operation.

Our experimental results demonstrate the proposed scheme provides higher PSNR, about 1-2dB compared to DC with framewise quantization and 0.2-0.5dB compared to DC with independent blockwise quantization. The proposed schemes can be used with arbitrary search window sizes, regardless of the block size used in DE. This blockwise dependent bit allocation can be a benchmark for faster allocation 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. The proposed scheme also 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.

This paper is organized as follows. In Section II we formulate the problem of bit allocation and we describe how to find optimal blockwise quantizer assignments using the Viterbi algorithm. We also discuss how to reduce the complexity of the allocation algorithm. Experimental results are provided in Section III. Finally, we discuss the results and give directions for future work in Section IV.

II. Blockwise Dependent Bit Allocation

A. Definitions and Notations

F_1 and F_2 are, respectively, the reference and target images in a stereo pair (refer to Fig. 1). Assume that an image is segmented into N square blocks. Then, the segmented image can be represented as a set of N blocks $F_l = \{B_{lm}, 0 \leq m \leq N-1, l \in (1,2)\}$, where

B and m represent a block and its index, and $l = 1, 2$ is the image index. Similarly, a blockwise quantizer allocation can be represented as $Q_l = \{q_{lm}, 0 \leq m \leq N-1, l \in (1,2)\}$. The overall rate and distortion are the sum of rates and distortions of the individual blocks, $R_l = \sum_{m=0}^{N-1} r(q_{lm})$, and $D_l = \sum_{m=0}^{N-1} d(q_{lm})$. In the following sections, to simplify the notation, we will sometimes use B_m and B'_m instead of B_{1m} and B_{2m} and q_m and p_m instead of q_{1m} and q_{2m}. Finally we will assume that a blockwise disparity field (V) has been computed, which indicates the correspondence between blocks in the target and reference image. The disparity field is defined as $V = \{v_m, 0 \leq m \leq N-1\}$, where the index m corresponds to a block in the target image. Note that in this work we only consider block based disparity estimation and that we assume that the disparity has only a horizontal component.

We use simple objective measures such as mean squared error (MSE) and peak signal to noise ratio (PSNR). The evaluation of reconstructed stereo images has to take into account properties of human visual perception, i.e. the preservation of 3D perception. However, the subjective evaluation of the quality is still an open problem and is not very reliable and repeatable yet. Therefore, we measure distortions of F_1 and F_2 using MSE, i.e. $D_1 = (F_1 - F_1(Q_1))^2$ and $D_2 = (F_2 - \hat{F}_2(Q_1, Q_2, V))^2$, where $F(Q)$ denotes the decoded image, when quantizer Q is used. The decoded target image, $\hat{F}_2(Q_1, Q_2, V)$, can be reconstructed by adding the compensated target image and the decoded DCD, i.e. $\hat{F}_2(Q_1, Q_2, V) = F_1(Q_1, V) + E(Q_2)$, where $E = F_2 - F_1(Q_1, V)$, i.e. the difference between the target image and the compensated image from the reconstructed reference image with DV.

B. Problem Formulation

We assume that, as is the case in current standards such as MPEG-2 or H.263x, a different quantizer (or quantization scale) can be assigned to each block (chosen from a finite set of available quantization choices.) Using DPCM to encode the quantizer selection would introduce additional *spatial* dependency between consecutive (or neighboring) blocks [27], [28]. For simplicity, however, we assume the quantizer indices are encoded with a constant number of overhead bits per block. Note that such a $1D$ dependencies due to the quantization indices could also be incorporated easily into our scheme.

Let (R_{budget}, R_V) be the given bit budget and the bits that were used for the DV field, respectively. For a given DV field, V, and remaining bit budget, $R_{budget} - R_V$, the optimal dependent bit allocation problem can be formulated as follows.

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

The relative importance of D_1 and D_2 can be controlled using the weighting constant α which allows us to support two different views of the depth perception process: *fusion theory* and *suppression theory* [29], [9]. Fusion theory claims that both images in a stereo pair contribute equally in 3D perception while suppression theory indicates that the highest quality image (or region) dominates the perception. According to suppression theory, one of the images in the stereo pair can be highly compressed as long as the other image retains the details of the scene. We set α equal to one during our experiments. We would have an independent bit allocation problem in the particular case where $D_2(Q_1, Q_2) = D_2(Q_2)$ and $R_2(Q_1, Q_2) = R_2(Q_2)$.

In general, this constrained optimization problem can be transformed into an unconstrained problem using the Lagrange multiplier method [30], [31], [32] and introducing a Lagrangian cost

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

where the Lagrange multiplier λ is a nonnegative constant. In a practical lossy data compression scheme, only a finite number of operational RD (ORD) 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 λ.

Figure 2 demonstrates the implications of operating in dependent bit allocation framework [25]. Note that, for a given λ and three ORD points, Q_{1b} is the RD optimal quantizer for the reference image because its Lagrangian cost $J_1(Q_{1b})$ is the lowest. However if the overall Lagrangian cost for the two images is taken into account things may change. For example, Q_{1a} may turn out to be the best choice for the reference image, if the total Lagrangian cost $J_1(Q_{1a}) + J_2(Q_{2b}|Q_{1a})$ is smaller than $J_1(Q_{1b}) + J_2(Q_{2b}|Q_{1b})$.

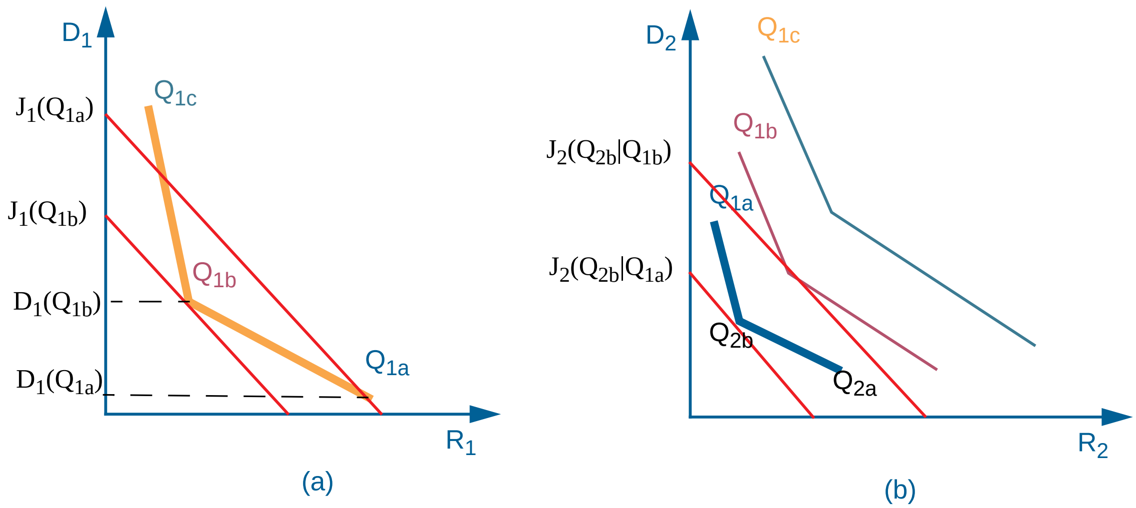


Fig. 2. Operational RD plots in a typical dependent bit allocation scenario: (a) reference image and (b) target image. Independent bit allocation: for given λ, the quantizer Q_{1b} is optimal because the Lagrangian cost $J_1(Q_{1b})$ is smaller than for the others. Dependent bit allocation: if stereo pairs are considered together, there is a chance for the quantizer Q_{1a} to be optimal, because the total Lagrangian cost $J_1(Q_{1a}) + J_2(Q_{2b}|Q_{1a})$ can be smaller than $J_1(Q_{1b}) + J_2(Q_{2b}|Q_{1b})$.

C. Optimal Blockwise Dependent Quantization

The distortion and the bit rate of a block in the DCD frame depend on up to two blocks in the reference image along the DV. Thus, the Lagrangian cost of (1) can be expressed in terms of the blockwise quantizer assignments as,

$$J(\lambda) = \sum_{m=0}^{N-1} \{d(q_m) + \lambda r(q_m)\} + \sum_{n=0}^{N-1} \{d(p_n, q^{\eta_1}(v_n)) + \lambda r(p_n)\} \tag{2}$$

where q^{η_1} is a vector which contains the quantizer indices of the blocks in the reference which are used to predict the current block in the target image. Note that η_1 denotes (at most) two consecutive blocks in the reference image. Figure 3 shows an example of the dependencies reflected in (2): given the disparity vector v_1, the selection of a quantizer for B_1' in the DCD frame will be affected by the selection of quantizers for B_2 and B_3 in the reference image. Thus a block in the DCD frame depends only on the quantizers, p_n and (q_m, q_{m+1}), *i.e.*, $d(p_n, q^{\eta_1}) = d(p_1, q_2, q_3)$ in Figure 3. In general, the index m can be denoted as $m = n + \lfloor \frac{v_n}{|B|} \rfloor$, where $\lfloor \rfloor$ and $|B|$ represent the floor function and the width of the block, respectively.

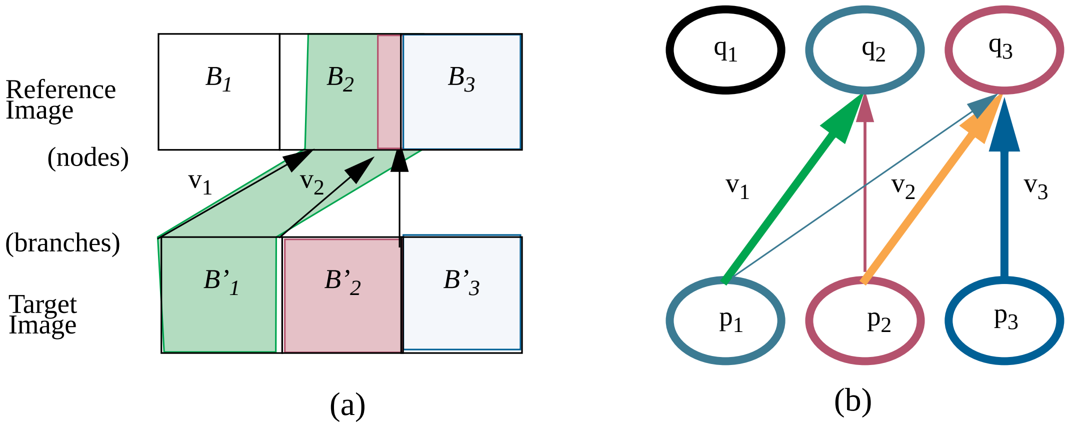


Fig. 3. Binocular dependency between corresponding blocks along the disparity vector. At most two consecutive blocks in the reference image are related to a block in the target image. For example, a block B'_1 in the target image is compensated from (at most) two consecutive blocks, B_2 and B_3, in the reference image along the disparity vector v_1. Therefore, the distortion of the block in the DCD frame is a function of p_1, q_2 and q_3.

Note that, under the assumption of predominantly horizontal disparity, blockwise dependent quantization can be performed independently in each row of blocks (ROB), without affecting the overall optimality. Since a ROB in the target image depends only on the ROB in the position of the reference image we can optimize each pair of ROBs (one from the reference and one from the target) independently: by using the same λ for each pair of ROBs we guarantee overall optimality. By comparison, in the case of video coding an optimal blockwise dependent bit allocation would require that the whole image be considered, because the temporal dependency links blocks in arbitrary directions in the $2D$ plane (whatever direction is indicated by the motion vector).

D. Solution using the Viterbi Algorithm

We can take advantage of the fact that dependencies are limited to at most two blocks to streamline the optimization approach. We construct a trellis to represent all possible quantization assignments for all blocks in a ROB in the reference frame and the corresponding ROB in the DCD. Refer to Fig. 4 for the trellis corresponding to the example in Fig 3. Each stage of the trellis corresponds to a block in the reference frame, with each node corresponding to a different quantizer allocation for that block. Thus the cost of each node is the Lagrangian cost of the block coded with the given quantizer. Then branches linking two nodes correspond to blocks in the DCD that depend on the two blocks in the reference frame represented by the corresponding stages. Note that more than one block in the DCD frame can be assigned to a given branch (this will depend on how large the disparity search region is and our algorithm accommodates any possible disparity range.) When several DCD blocks depend on the reference blocks represented by the two stages we simply add the corresponding Lagrangian costs to the branches linking those two stages. For example in Fig 3 two blocks in the DCD are assigned to a a branch, i.e. B_1' and B_2' both depend on B_2 and B_3 and thus the two Lagrangian costs corresponding to B_1' and B_2' would be added to each branch linking stages 2 and 3 in the trellis.

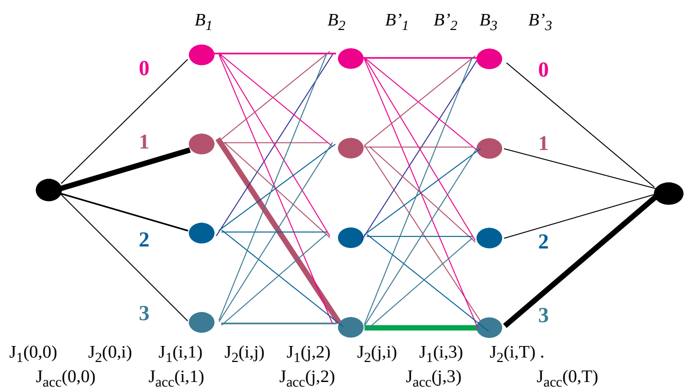


Fig. 4. Trellis structure for blockwise dependent bit allocation. Each node in the trellis corresponds to a quantizer choice for a block in the reference image and has a corresponding Lagrangian cost. The quantizer indices are monotonically increasing ordered from finest to coarsest. Each branch corresponds to a quantization assignment to all the blocks in the DCD frame that depend on the corresponding blocks in the reference frame. Branches can represent more than one quantizer because two consecutive blocks in the reference image can have dependency with more than one block in the target image. A branch linking stages i and $i+1$ has a Lagrangian cost corresponding to the optimal quantizers for the DCD which depend on blocks i and $i+1$ from the reference image. The darker path denotes selected quantizers using the Viterbi algorithm.

Once the trellis has been constructed we can use deterministic dynamic programming techniques, i.e. the Viterbi algorithm (VA) [33], to find the path with the smallest overall Lagrangian cost. This will be the optimal cost for the given λ. We now define our method more formally. Let k be the index of the stage and (i, j) denote indices of nodes in two consecutive stages, k and $k+1$, respectively. Note that a trellis is built for each ROB in the in the image, so when we refer to k-th block it should be clear that this is *within the particular ROB*. Let ROB_1 and ROB_2 be the reference image and DCD image ROBs respectively. We define,

Stage: the kth stage in the trellis corresponds to the kth block in ROB_1. Therefore, the number of stages, K, is equal to the number blocks in the ROB.

Node: each node in the kth stage corresponds to a possible quantizer choice for the kth block of ROB_1. The choices are ordered from top to bottom in order of finest to coarsest. Therefore, the number of state nodes per stage is $L = |q|$, i.e. the number of available quantizers for the reference image. Each node has a corresponding Lagrangian cost, $J_1(i,k)$ in (3), which depends only on the rate and the distortion of the kth block when quantizer i is used.

$$J_1(i,k) = d(q_k^i) + \lambda r(q_k^i) \tag{3}$$

Branch: A branch, joining nodes q_k^i and q_{k+1}^j, corresponds to the *optimal* vector of quantizers, p_n^{ij}, for the (possibly more than one) blocks in ROB_2 which depend on blocks k and $k+1$ in ROB_1. The subscript n denotes all the blocks in ROB_2 which depend on k and $k+1$ in ROB_1. Each branch has a total Lagrangian cost

$$J_2(i,j) = \sum_{n \in \eta_2(k,k+1)} \{d(p_n^{ij}|q_k^i, q_{k+1}^j) + \lambda r(p_n^{ij}|q_k^i, q_{k+1}^j)\} \tag{4}$$

which adds up the Lagrangian costs corresponding to each of the blocks n.

Path: A path is 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 ROB_1 (nodes) and ROB_2 (branches). The cost of a path is the accumulated cost of branches and nodes along the path.

Trellis: The 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 a given ROB in the stereo pair.

The optimal 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. Note that the sum of (3) and (4) over a path is equal to a Lagrangian cost of selected quantizers in (2). Therefore, a path in a trellis corresponds to the quantizer choices assigned to ROB_1 and ROB_2. The Viterbi algorithm can be used in searching the minimum cost path through the trellis. With a given λ, the optimal set of dependent quantizers, (Q_1, Q_2), can be found by applying repeatedly the following procedure to the corresponding ROBs.

Step 0: Initialization: let K and L be the number of stages and nodes per stage, respectively. Add an initial node B_0 and a final node B_T where $T = K + 1$. Select a λ and set $k = 0$ & $J_{acc}(0, 0) = 0$.

Step 1: At stage k, branches are added to the end of each node i (of all surviving paths) and Lagrangian costs, J_1 and J_2, are assigned to the node and the branch, respectively.

Step 2: At a stage $(k + 1)$, for each node j, an accumulated transition cost from node i, $J_{tr}(i, j)$, is calculated by summing the accumulated cost, $J_{acc}(i, k)$, and the transition cost, $J_2(i, j)$. Of all arriving branches (at most L), the one with the lowest accumulated-transition-cost is chosen. The resulting cost is assigned to the accumulated cost, $J_{acc}(j, k + 1)$ and the remaining branches are pruned.

$$\begin{aligned} J_{tr}(i,j) &= J_{acc}(i,k) + J_2(i,j) \\ J_{acc}(j,k+1) &= \min\{J_{tr}(i,j)\}_{i=0}^{L-1} \\ J_{acc}(j,k+1) &= J_{acc}(j,k+1) + J_1(j,k+1) \end{aligned} \tag{5}$$

Step 3 if $k < K$, then $k = k + 1$, go to *step 1* and repeat.

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

In the proposed framework, the quantization choices for the kth block in the reference image and corresponding blocks in the DCD frame do not affect the choices for the future blocks. Thus, based on the Bellman's optimality, the Viterbi algorithm provides a globally optimal solution because pruning suboptimal paths at a given node does not eliminate paths that could potentially be globally optimal [34], [35]. In other words, if the minimum cost path from stage 1 to stage k passes through a node at stage $(k - 1)$, then its subpath from 1 to $(k - 1)$ is also the optimal path from 1 to $(k - 1)$.

E. Selection of λ and a Heuristic Fast Algorithm

Note that the bit rate is a function of λ. Given that the cost function has the form, $J = D + \lambda R$, increasing λ and then finding the optimal point is equivalent to finding the ORD point that first "hits" the line of absolute slope λ (see Figure 2). Thus, increasing λ corresponds to achieving optimal points with higher distortion and lower bit rate.

For a fixed λ, using the Viterbi algorithm, we can obtain the best possible quantizer selection that minimizes the Lagrangian cost defined in (2). However, to find optimal quantizer with a given bit budget, we may need to iteratively change λ until we find λ^* such that $R(\lambda^*) - (R_{budget} - R_V) \leq \epsilon$, for $\epsilon \geq 0$. The desired λ^* can be selected using a fast bisection search algorithm, which can be found in [32], [36].

Let (L, K) be the numbers of nodes per stage and stages per trellis. For each node or branch, L comparisons have to be performed. Thus, the required total number of comparisons is $O(KL^3)$, because the total number of nodes and branches per trellis are $L \times K$ and $L^2 \times K$, respectively. The main complexity of the proposed scheme is in the RD-point generation because each comparison in a node requires that the corresponding RD values be known and these are different for each branch. The complexity can be reduced by approximating the RD values instead of calculating real RD values [26]. We propose an alternative method which reduces the search space in the trellis but this method could be also combined with modeling approaches to further speed up the search.

In general, computing RD values for blocks in the DCD is more complicated. Therefore, we propose a heuristic fast algorithm, which restricts search space to the paths selected by the reference image. First, we only calculate ORD points for the reference image (i.e. we compute the rate and distortion for each quantization choice and each block in the reference image). Then, we apply the Viterbi algorithm *with the branch costs set to zero* for two different values of the Lagrange multiplier λ_1 and λ_2. Each λ will provide an optimal path (a set of nodes). We then restrict ourselves to only consider those paths that lie *in between* the paths selected using λ_1 and λ_2. Finally we use the algorithm outlined above except that we apply the VA on the pruned trellis so that only a subset of the branches representing DCD blocks need to be grown. This reduces the computational complexity significantly. For example, the complexity of the trellis in Figure 4 can be reduced as shown in Figure 5. The proposed fast search algorithm is as follows.

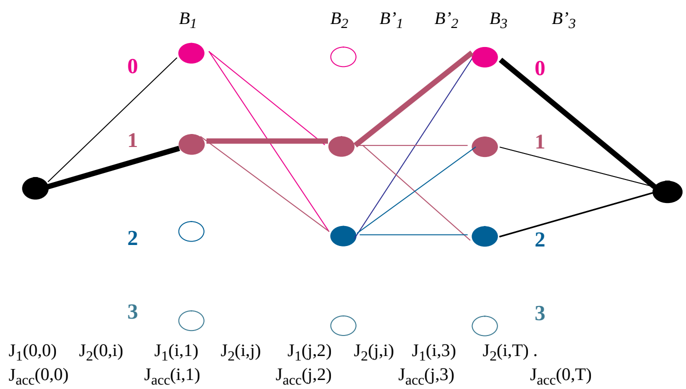


Fig. 5. A heuristic fast search. The trellis in Figure 4 can be restricted using the proposed fast search algorithm. The underlying assumption is that the reference image is more important than the difference. The search space is reduced to the nodes selected by the blockwise quantization for the reference image.

Step 0: Select two λ's, *e.g.*, (λ_1, λ_2), for the reference image.

Step 1: for each λ_i, find an optimal path in the trellis only for the reference image, using only the node costs, and setting all branch costs to zero. Then, keep the nodes between the selected paths. Note that as λ goes to 0, the minimum cost path is the one corresponding to the finest quantizers.

Step 2: calculate corresponding RD-points of the DCD frame and assign to the branches of the pruned trellis.

Step 3: find the shortest path satisfying the given bit budget constraint using Viterbi algorithm on the pruned trellis.

III. Experimental Results

In our experiments we use the images shown in Figures 6 (a) and (c)[3]. The target image is segmented into blocks of size 8×8 pixels and then disparity estimation is performed using fixed size block matching (FSBM) between the target image and the reference image

[3]The test images as well as decoded images corresponding to the results presented in this paper are available in `http://escalus.usc.edu/ wwoo/Stereo`. The original images where obtained from
Room: `http://www-dbv.cs.uni-bonn.de/ ft/stereo.html` and
Fruit: `http://www.ius.cs.cmu.edu/idb/html/stereo/index.html`

within a search window of size 1×16. For this particular selection of DE block size and search region size two consecutive blocks in the reference image will affect *at most* two consecutive blocks in the target image, as in Figure 4.

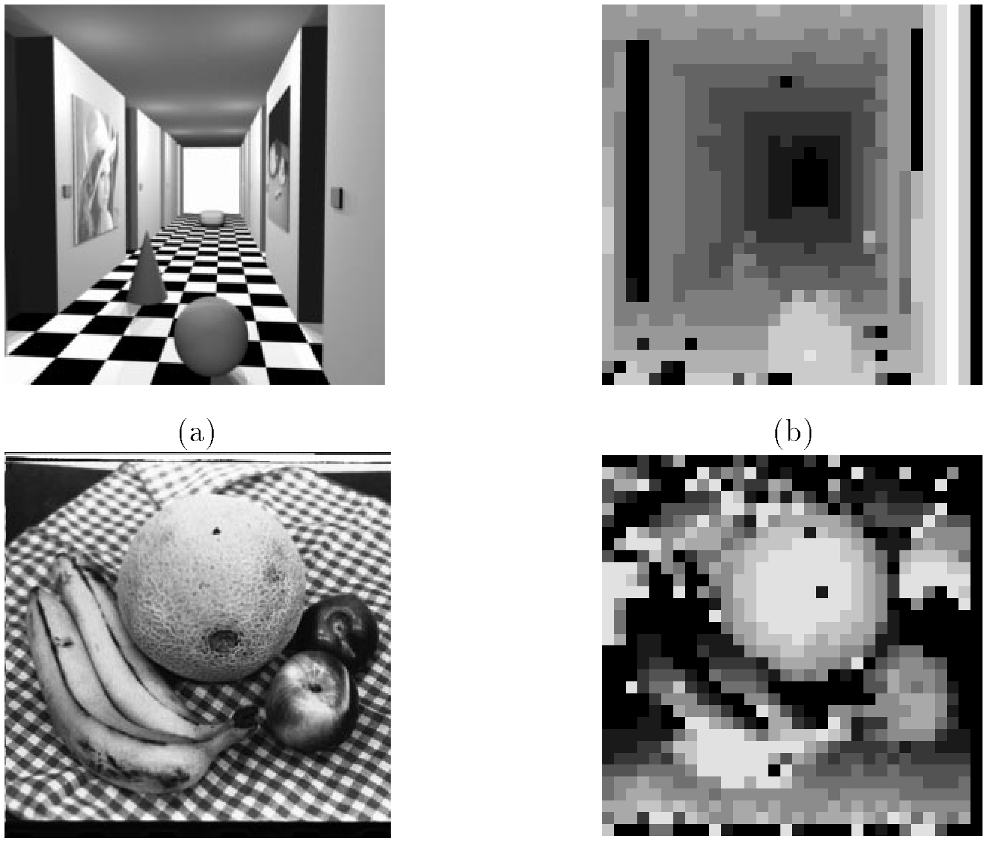
The resulting DV fields are shown in Figure 6 (b) and (d). The DV field is losslessly encoded using DPCM with a causal median predictor to exploit the spatial redundancy among neighboring DVs. The reference image and the DCD frame are encoded using a JPEG-like coder, with the only modification with respect to baseline JPEG [37] being that we allow each block to have a different quantization scale (QS). Note that a given quantization table in JPEG only determines the relative coarseness of quantization step for each coefficient within a block. Consequently, the change of QS per block allows the encoder to assign different levels of quantization coarseness to each block. For each block one of among eight different QS can be chosen from the set $QS = \{90, 80, \cdots, 20\}$, where increasing values indicate finer quantization. In our calculation of rate, we assume a constant overhead is used for each block and thus overhead is not incorporated in our optimization. Similarly our total rate computation includes only the rate for each block and not any applicable headers for the compressed file.

The performance is also assessed in terms of quality, as measured by the peak-signal-to-noise-ratio (PSNR) in dB for each image. The mean PSNR is used to evaluate the overall performance of the stereo pair,

$$PSNR_{mean} = 10 \times \log_{10}\{\frac{255^2}{(D_1 + D_2)/2}\} \qquad (6)$$

where D_1 and D_2 are the MSE of the reconstructed reference and target images, respectively.

Figure 7 compares the RD performance we achieve with four different algorithms: (i) JPEG without DC (ii) JPEG with DC (iii) JPEG with DC and independent blockwise quantization and (iv) the proposed JPEG with DC and dependent blockwise quantization. In algorithms (i) and (ii) a constant quantization scale is used for all blocks in each image. In algorithms (iii) and (iv), we fix λ for the two images and then find the optimal quantization scale for each block. The RD points we plot are obtained for $\lambda = \{0, 0.1, 0.5, 1, 2, 100\}$. As shown in Figure 7 (a), the RD performances for the reference image are similar with or without a dependent bit allocation. However, the dependent bit allocation results in



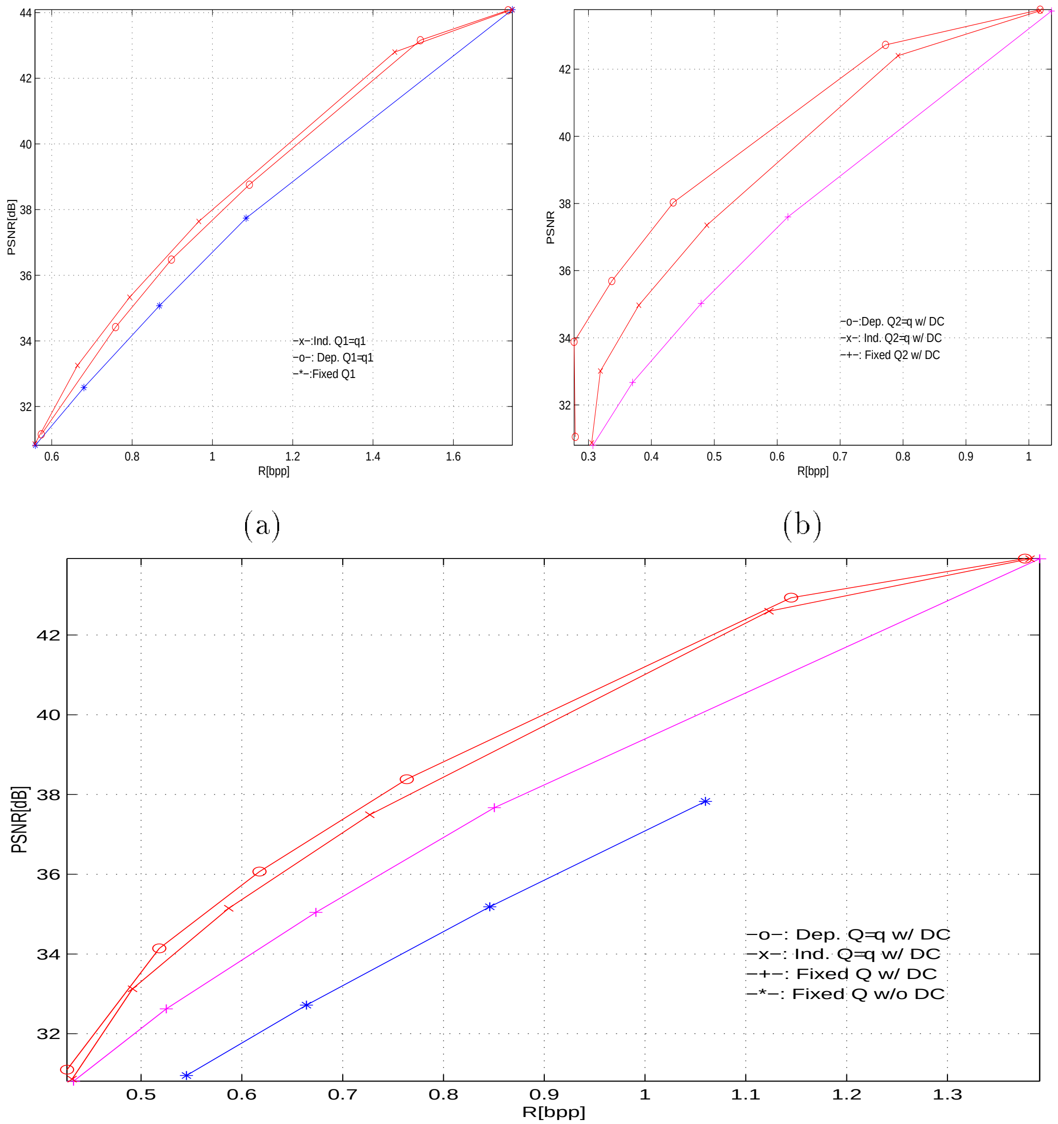
(a) (b) (c) (d)

Fig. 6. Test images and DE results with 8×8 block. (a) the target image (l_room.256) (b) the DV field with FSBM (c) the target image (l_fruit.256) (d) the DV field with FSBM.

significant RD performance gains for the target image as shown in Figure 7 (b). Figure 7 (c) compares the RD performance in terms of the overall bit rate and the mean PSNR (and again each of the points in the plot corresponds to a particular λ.) Figure 8 shows the RD obtained for another stereo pair.

According to the experimental results, the proposed blockwise dependent bit allocation method resulted in 1-2 dB improvement in average PSNR at the given bit rates compared to a fixed quantization without DC (JPEG) and 0.5 dB improvement compared to the independent blockwise quantization with DC.

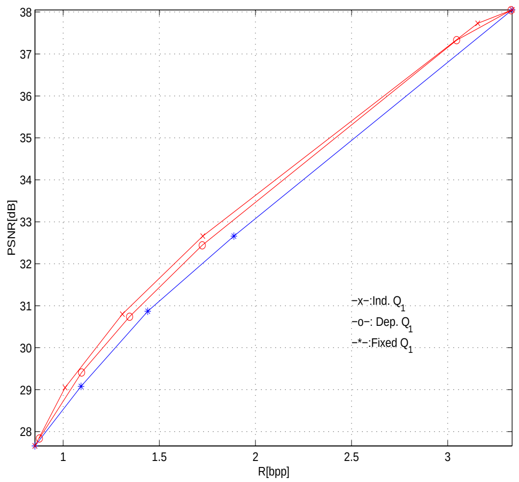
Figure 9 shows mean ORD curves for the reference image (points marked with x) and the dependent DCD frame (points marked with o), respectively. As shown, the monotonicity property is satisfied for the blockwise quantization, i.e. $J(QS_2|QS_1^*) \leq J(QS_2|QS_1)$, for



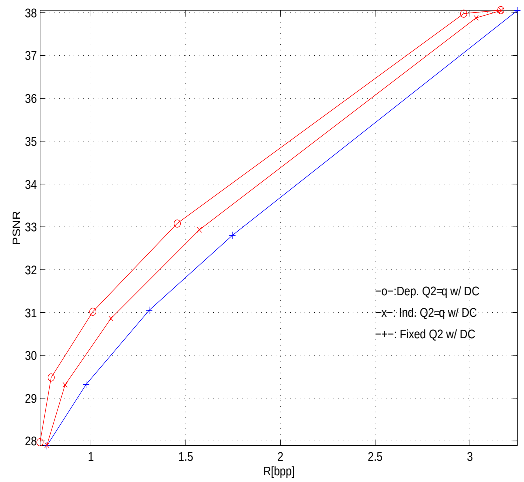
(a) (b)

(c)

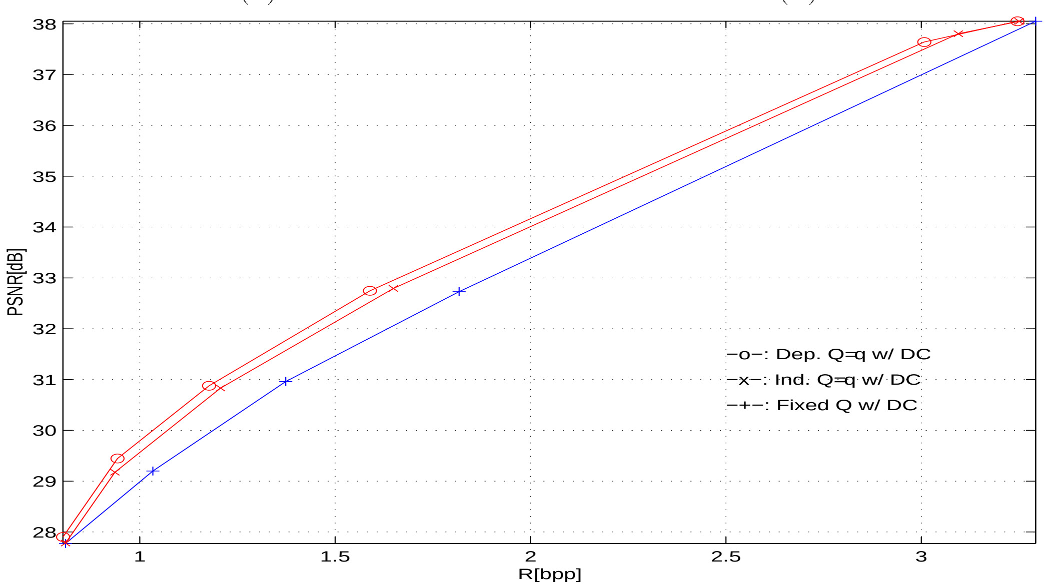
Fig. 7. RD performance comparison (Image: room.256, Block size=8 $\times$ 8, $SW = 16$, $|Q| = 8$, $QS = \{90, 80, \cdots, 20\}$ and $\lambda = \{0, 0.1, 0.5, 1, 2, 100\}$). The points marked with $*$ correspond to the results of JPEG without DC and those marked with $+$ to JPEG with DC. In both cases a fixed QS is used for the whole image. The points marked with x and o correspond to blockwise independent and dependent QS selections, respectively. Each point is generated with one different λ. (a) The RD performance for the reference image is similar for both types of blockwise allocation. (b) Better RD performance for the target image can be achieved using the dependent bit allocation approach. (c) The overall performance also improves when taking dependencies into account.



(a)



(b)



(c)

Fig. 8. RD performance comparison. (fruit.256, Block size=8×8, $SW = 10$, $|Q| = 8$, $QS = \{90, 80, \cdots, 20\}$ and $\lambda = \{0, 0.1, 0.5, 1, 2, 100\}$). The points marked with + correspond to the results of JPEG with DC. The points marked with x and o correspond to blockwise independent and dependent QS selections, respectively. RD characteristics for (a) the reference image, (b) the DCD frame, and (c) the two images.

$QS_1^* \geq QS_1$. Thus if the quality of the reference frame improves, so does, for the same quantization scale QS_2, that of the DCD, i.e. if $\lambda = 0$, $d(QS_2|QS_1^*) \leq d(QS_2|QS_1)$, for $QS_1^* \geq QS_1$ [25]. Thus the finer quantization ($QS_1 = 90$) leads to more efficient coding for the DCD frame in the RD sense so that the corresponding mean ORD curve is closer to the origin, In addition, the plot shows that, in both cases, the distortion, $d(QS_2|QS_1)$, are monotonically increasing by changing quantization scales from the finest to coarsest, i.e. from 90 to 30.

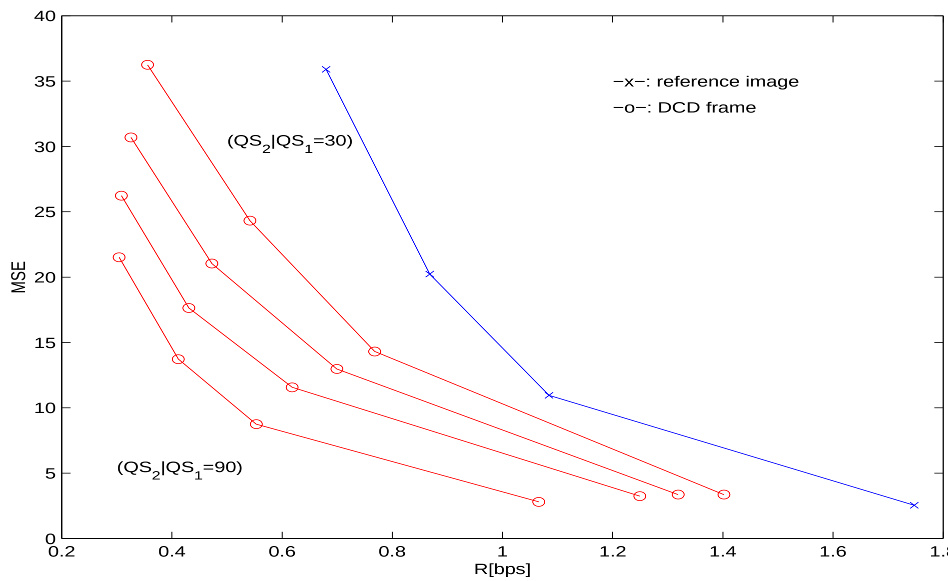


Fig. 9. Mean ORD plots for the block in the reference image and the DCD frame. (room.256, $QS = \{90, 70, 50, 30\}$) QS_2 is changed for the DCD frame with a given QS_1. As shown, the monotonicity property is satisfied, i.e. $J(QS_2|QS_1^*) \leq J(QS_2|QS_1)$, for $QS_1^* \geq QS_1$. In particular, if $\lambda = 0$, $d(QS_2|QS_1^*) \leq d(QS_2|QS_1)$, for $QS_1^* \geq QS_1$.

This observed monotonicity property explains the good performance of the fast algorithm described in the previous section. Figure 10 shows the RD performance of the proposed fast algorithm. As explained, the blockwise quantization scheme is applied to the reference image alone with two λ's, $\{\lambda_1, \lambda_2\} = \{0, 0.5\}$ in our example. Then, the search space is restricted to the nodes between two paths selected by the λ's. As a result, only 61% of original nodes remain, which corresponds to 37.2% of the original number of branches. Finally, the set of dependent quantization assignments is determined using the pruned trellis. The proposed scheme significantly reduces the encoding complexity, since only those ORD points corresponding to the remaining points in the trellis need to be computed. The overall RD performance remains practically unchanged in this case. Note however that we need to make a good choice for the λ range, based on the expected quality level for the overall image. Thus in the example we show good performance at high rates, whereas the low rate points cannot be achieved (since the corresponding nodes have been pruned out).

IV. Discussion

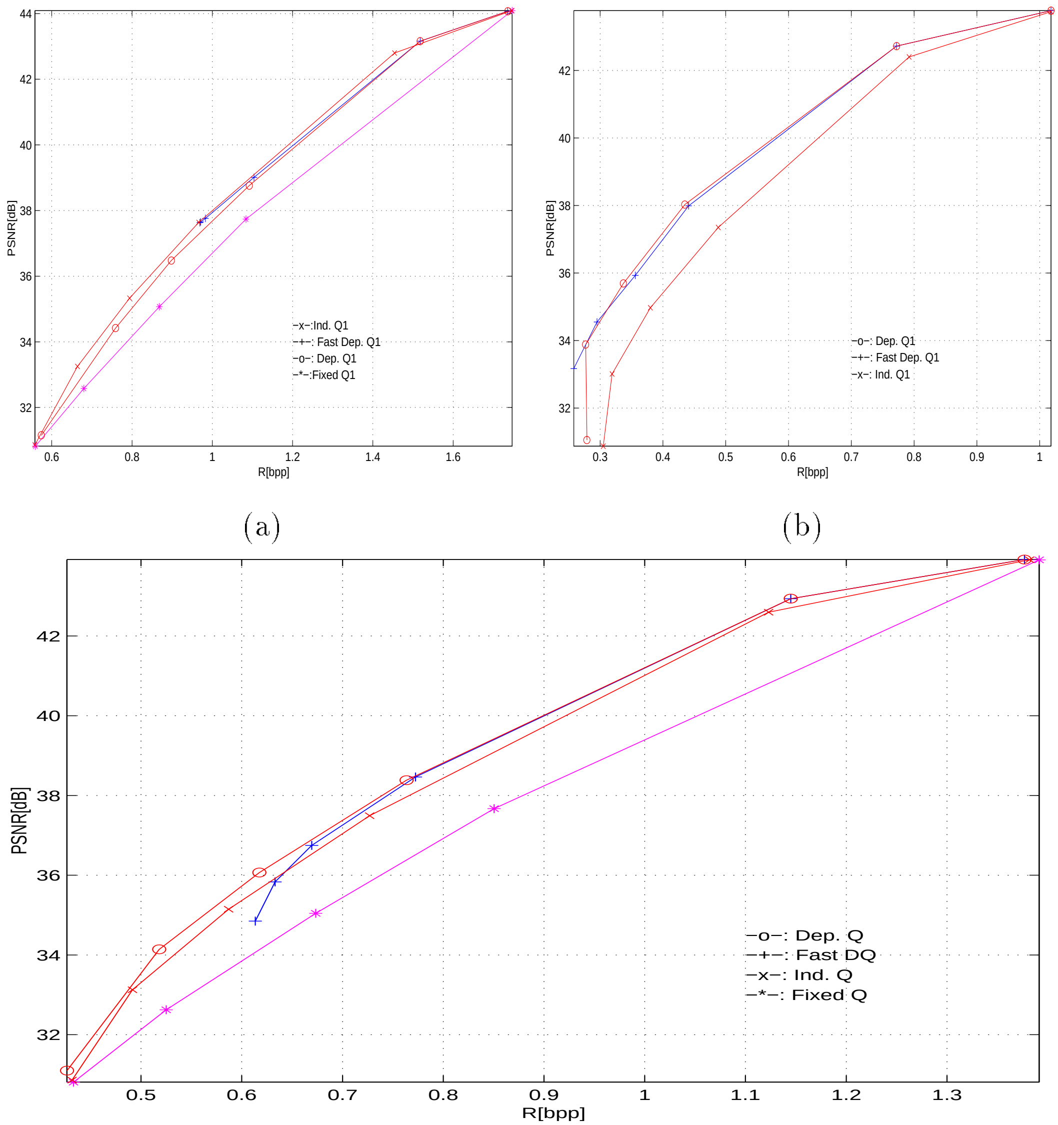
We have proposed an optimal dependent bit allocation scheme for stereo image coding. We have concentrated on quantization issues and assumed that the disparity estimation was performed open-loop. The proposed dynamic programming algorithm leads to an efficient bit allocation between the reference image and the DCD frame. According to our experimental results, the proposed scheme provides significant PSNR gains, for example about 1-2dB compared to constant quantization and 0.2-0.5dB compared to the independent blockwise bit allocation with DC. In addition, we have shown a method to reduce the computational complexity and the encoding delay of the Viterbi algorithm by exploiting the monotonicity property. Adopting reasonable RD models can further reduce the computational complexity of the proposed scheme [26]. This framework has been developed for a JPEG-like codec but it can be directly extended to an MPEG-like codec without loss of generality. Further research, however, is required to achieve a more complete allocation algorithm which includes also the DV. The extension to video coding, in which both temporal and binocular dependencies have to be taken into account is another area of future work. Finally, further study of our algorithm may lead to a better understanding of the similar issues in blockwise dependent allocation for video coding, where an optimal solution cannot be achieved due to the 2D nature of the dependencies.

Acknowledgments

The authors would like to thank the CVPR Group at the University of Bonn, Germany, for providing the "room" test images.

References

[1] W. Woo and A. Ortega, "Dependent quantization for stereo image coding," in *SPIE Proc. on EI-VCIP 98*,



(a) (b)

(c)

Fig. 10. RD performance comparison of fast algorithm. (room.256, Block size=8×8, $SW = 10$, $|Q| = 8$, $QS = \{90, 80, \cdots, 20\}$, $\lambda_1 = [0, 0.5]$, and $\lambda = \{0, 0.1, 0.5, 1, 2, 100\}$). The points marked with + denote the results of the proposed fast algorithm which only uses 61% of the original nodes, which corresponds to 37.2% of the original comparisons. The x-mark-line corresponds to the results of the blockwise independent QS selection and the o-mark-line to those of the blockwise dependent QS selection. RD characteristics for (a) the reference image, (b) the DCD frame, and (c) the two images combined.

Jan. 1998, vol. 3309.

[2] T. Motoki, H. Isono, and I. Yuyama, "Present status of three-dimentional television research," *Proceedings of the IEEE*, vol. 83, no. 7, pp. 1009–1021, July 1995.

[3] B.G. Haskell, A. Puri, and A.N. Netravali, *Digital Video: Introduction to MPEG-2*, The MIT Press, 1986.

[4] A. Puri, R. V. Kollarits, and B. G. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *J. on Signal Processing: Image Comm.*, , no. 10, pp. 201–234, 1997.

[5] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," in *Proc. on ICASSP 86*, Oct. 1986, pp. 521–524.

[6] M. G. Perkins, "Data compression of stereopairs," *IEEE Tr. on Comm.*, vol. 40, pp. 684–696, Apr. 1992.

[7] W. Woo and A. Ortega, "A comparison of techniques for disparity estimation," *IEEE Tr. on CSVT*, 1998, In preparation.

[8] I. Dinstein, G. Guy, and J. Rabany, "On the compression of stereo images: Preliminary results," *Signal Processing*, vol. 17, no. 4, pp. 373–381, Aug. 1989.

[9] I. Dinstein, M.G. Kim, A. Henik, and J. Tzelgov, "Compression of stereo images using subsampling transform coding," *Optical Engineering*, vol. 30, no. 9, pp. 1359–1364, Sept. 1991.

[10] T. Ozkan and E. Salari, "Coding of stereoscopic images," *SPIE Image and Video Processing*, vol. 1903, pp. 228–235, 1993.

[11] H. Jeong, W. Woo, C. Kim, and J. Kim, "A unification theory for early vision," in *Proc. on First Korea-Japan Joint Conf. on the Computer*, Oct. 1991, pp. 298–309.

[12] W. Woo and A. Ortega, "Stereo image compression based on the disparity compensation using the MRF model," in *SPIE Proc. on VCIP 96*, Mar. 1996, vol. 2727, pp. 28–41.

[13] M. Accame, F. G. De Natal, and D. D. Giusto, "Hierarchical block matching for disparity estimation in stereo sequence," in *Proc. IEEE ICIP'95*, May 1993, pp. 200–207.

[14] S. Sethuraman, M.W. Siegel, and A.G. Jordan, "A multiresolution framework for stereoscopic image sequence compression," in *IEEE Proc. ICIP-94*, Oct. 1994, vol. 2, pp. 361–365.

[15] S. Sethuraman, M.W. Siegel, and A.G. Jordan, "A multiresolution region based segmentation scheme for stereoscopic image sequence compression," in *SPIE Proc. EI'95 Digital Video Compression: Algorithms and Technologies*, Feb. 1995, vol. 2419, pp. 265–274.

[16] S. Malassiotis and M.G. Strinzis, "Object-based coding of stereo image sequences using 3-D model," *IEEE Tr. on CSVT*, vol. 7, no. 6, pp. 892–905, Dec. 1997.

[17] W. Woo and A. Ortega, "Stereo image compression based on the disparity field segmentation," in *SPIE Proc. on EI-VCIP 97*, Feb. 1997, vol. 3024, pp. 391–402.

[18] W.C. Chung, F. Kossentini, and M.J.T. Smith, "Rate-Distortion constrained statistical motion estimation for video coding," in *In Proc. on IEEE ICIP*, Oct. 1995, pp. 184–187.

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

[20] M.C. Chen and A.N. Willson Jr., "Rate-Distortion optimal motion estimation algorithms for motion-compensated transform video coding," *IEEE Tr. on CSVT*, vol. 8, no. 2, pp. 147–158, Apr. 1998.

[21] J. Ribas-Corbera and D. Neuhoff, "Optimal bit allocations for lossless video coders: Motion vectors vs. difference frames," in *In Proc. on IEEE ICIP*, Oct. 1995, pp. 180–183.

[22] G. M. Schuster and A. K. Katsaggelos, "A theory for the optimal bit allocation between displacement vector field and displaced frame difference," *IEEE Jr. on Sel. Areas in Comm.*, vol. 15, no. 9, pp. 13–26, Dec. 1997.

[23] D. Tzovaras and M.G. Strintzis, "Motion and disparity field estimation using rate-distortion optimization," *IEEE Tr. on CSVT*, vol. 8, no. 2, pp. 171–180, Apr. 1998.

[24] M.S. Moellenhoff and M.W. Maier, "Transform coding of stereo image residulals," *IEEE Tr. on IP (In Review)*, 1998.

[25] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Tr. on IP*, vol. 3, no. 5, pp. 533–545, Sept. 1994.

[26] L.J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Tr. on CSVT*, Aug. 1998, To appear.

[27] T. Wiegand, M. Lightstone, D. Mukherjee, T.G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit-rate video coding and the emerging H.263 standard," *csvt*, vol. 6, no. 2, pp. 182–190, April 1996.

[28] A. Ortega and K. Ramchandran, "Forward-adaptive quantization with optimal overhead cost for image and video coding with applications to MPEG video coders," in *Proc. of SPIE, Digital Video Compression: Algorithms & Technologies 95*, San Jose, CA, Feb. 1995.

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

[30] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Res.*, vol. 11, pp. 399–417, 1963.

[31] G.L. Nemhauser and L.A. Wolsey, *Integer and Combinatorial Optimization*, Wiley, 1988.

[32] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Tr. on ASSP*, vol. 36, pp. 1445–1453, Sept. 1988.

[33] G.D. Forney, "The Viterbi algorithm," *The Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[34] R.E. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[35] D.P. Bertsekas, *Dynamic Programming:Deterministic and Stochastic Models*, Prentice Hall, 1987.

[36] S.-W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recoding of HDTV," *IEEE Tr. on CSVT*, vol. 1, no. 1, pp. 100–112, Mar. 1991.

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

List of Figures