# MINIMUM DISTORTION SAMPLING OF MULTIPLE IMAGE SEQUENCES BASED ON JOINT TEMPORAL ANALYSIS

*Anthony Vetro and Huifang Sun*

Mitsubishi Electric Research Labs
201 Broadway, Cambridge MA 02139 USA

## ABSTRACT

This paper introduces the problem of sampling frames from multiple image sequences with the goal to minimize distortion of the sampled sequences subject to a total memory constraint. A joint temporal analysis over a window of all input image sequences is performed to determine the optimal set of frames to be recorded. Compared to uniform sampling, simulation results indicate that significant reductions in distortion could be obtained with the proposed variable sampling approach.

## 1. INTRODUCTION

Recording of surveillance video is a memory intensive operation. Typical surveillance systems include digital recorders that are capable of storing compressed images or video from up to 16 different cameras. Most systems contain a hard-disk drive (HDD) of 120GB or 240GB in size. Given this fixed memory, and the desire to record as much content as possible, a lower temporal rate of the video is typically recorded. This recording of the input image sequences at a lower temporal rate is often referred to as time-lapsed recording. In current recorder systems, the lower temporal rate is achieved by simply uniformly sampling the input frames.

The major drawback to uniformly sampling the input images sequences is that frames from inactive inputs are recorded the same as active inputs. In this way, the memory is not being efficiently utilized since inactive inputs are unnecessarily being sampled, and information contained in active sequences is lost due to under-sampling. Clearly, a variable sampling of the input image sequences would overcome these drawbacks.

There are two types of recorder systems that could be considered. The first accepts uncompressed video as input and performs the encoding at a lower temporal rate within the recorder. The second type of system accepts compressed image sequences (which are intra-coded) to reduce network traffic and simply samples the frames to be recorded. This paper focuses on the second type of recorder system. JPEG encoding is assumed, but other intra-coding techniques also apply.

An overview of the variable-rate recording system proposed in this paper is illustrated in Figure 1. The system is shown with several camera inputs in which the frames are intra-coded using JPEG at a full temporal rate. These streams are transmitted over a network and are routed to a cluster of displays that render the image sequences at the full frame rate. The input sequences are also routed to a network video recorder, which performs a joint temporal analysis and samples the frames in a non-uniform manner and records the sampled frames to memory.

Although there exist techniques in the literature for encoding or transcoding video at a variable temporal resolution, e.g., [1],[2], none of these works consider the maximum number of frames to be recorded as a constraint. Rather, the emphasis is usually on satisfying a bit-rate budget, where the spatial quality of frames may also be considered. Also, most existing techniques focus on the encoding or transcoding of a single video. There do exist techniques that address multiple videos, e.g., [3],[4]. But again, the emphasis is more on rate allocation.

The rest of the paper is organized as follows. In the next section, the problem being addressed is stated formally. In Section 3, the joint temporal analysis is described. Simulation results are provided in Section 4, and concluding remarks in Section 5.

## 2. PROBLEM FORMULATION

Let $M$ denote the number of input image sequences, and $D_i$ the distortion of the sampled sequence $i$. The problem is stated as follows:

$$\min \sum_{i=0}^{M-1} D_i \quad \text{subject to } N_r \le N_m \tag{1}$$

where $N_r$ denotes the actual number of recorded frames and $N_m$ denotes the maximum number of recorded frames. Both quantities are aggregate totals over all input image sequences. To achieve a minimum distortion, the constraint in eqn. (1) should be satisfied with equality.

In order to solve the above problem, we must consider the objective function and constraints over a moving window of frames $[f_0, f_{T-1}]$, where $f_0$ is the first frame in the window, and the window size is $T$. Since the frames in our system are simply being sampled without change to

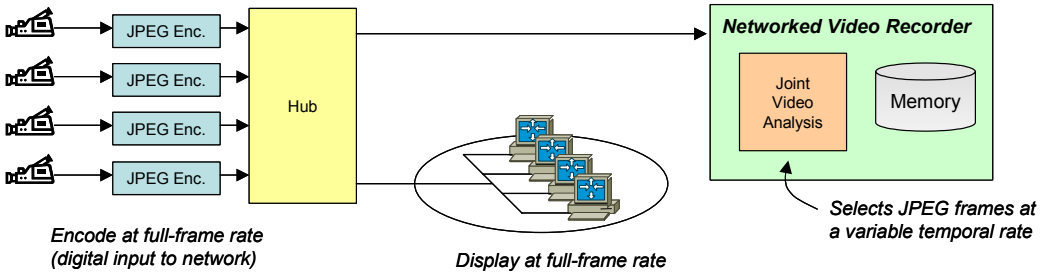



**Figure 1.** Networked video recorder system with variable sampling of input frames.

the spatial quality, the distortion is fully represented by the temporal error incurred by skipped frames in each sequence, $d_{i,s}$, where $s$ is the index of a skipped frame in the set $S_i$ of sequence $i$. In this way, the distortion of each sequence in a specified frame interval is given by:

$$D_i[f_0, f_{T-1}] = \sum_{s \in S_i} d_{i,s}[f_0, f_{T-1}] \qquad (2)$$

The above problem is unique in two ways. First, to the best of our knowledge, explicit constraints on the maximum number of frames to be recorded have not been considered. Second, since we sample the intra-coded input sequences, only temporal distortion of the sampled image sequences needs to be accounted for.

The maximum number of frames to be recorded in the each frame interval is expressed as:

$$N_m = \left\lfloor M \cdot T \cdot \overline{f}_s / F \right\rceil \qquad (3)$$

where $F$ is the original frame rate, $\overline{f}_s$ is the average sampling rate, and $[\cdot]$ denotes a rounding operation.

## 3. JOINT TEMPORAL ANALYSIS

The joint temporal analysis determines the frames to be sampled in each sequence to satisfy eqn. (1). An example of the input and output of this analysis is illustrated in Figure 2, where the input sequence of frames is at full frame-rate and the output is sampled according to the results of the analysis.

In our current system, we impose the constraint that the first frame of each sequence in the window be sampled. This allows us to perform the analysis within a given window independent of the results in the previous stage. With this constraint, a minimum sampling rate for each input sequence is imposed, which may not be optimal, but greatly simplifies the solution to the problem.

Given the above, the number of remaining frames to be sampled under the constraint becomes $r = N_m - M$. Let the number of possible frames to be sampled be denoted by, $n$, then there exist $\binom{n}{r}$ possible solutions that need to be evaluated, where $n = M \cdot (T-1)$.

Obviously, as the window size increases, the number of possible solutions that need to be evaluated can become very high. In this work, we search all the possible solutions exhaustively using the algorithm in [5]. This algorithm supplies a vector of binary digits that allow us to denote sampled frames with 1's and skipped frames with 0's. For each candidate solution, the index of the skipped frames for each sequence becomes part of the set $S_i$. For each sequence, the distortion for each frame in this set is computed with respect to the last coded frame in that particular solution. The solution that yields minimum overall distortion is selected.

To simplify some of the computation, the distortions of skipped frames are computed based on DC images extracted from the JPEG encoded frames. This not only saves on computation usually required for reconstructing the images, but also reduces the number of image points used to compute the temporal distortion between any two frames. As an estimate of the temporal distortion, we calculate the Sum of Absolute Difference (SAD) between frames; it is noted that other well-known metrics, such as MSE, or even metrics that emphasize perceptual changes between frames, may also be used here. Another advantage of using DC images to compute the distortion is that the pixels are low-pass filtered in some sense, thereby eliminating the influence of noise and other insignificant pixel differences in the original image sequence.

As mentioned earlier, the temporal distortion of a skipped frame is computed based its difference compared to the last sampled frame in the sequence. Since the distortions between two frames are used repeatedly in different combinations of possible solutions, it is beneficial to compute these distortion values ahead of time and store them in an indexed array. The total number of values that need to be computed and stored is $M \cdot \binom{T}{2}$.

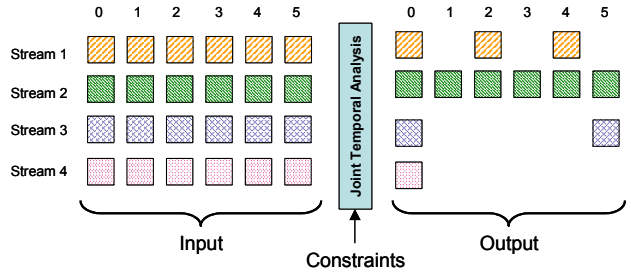



**Figure 2.** Input-output of joint temporal analysis.

## 4. SIMULATION RESULTS

To demonstrate the reduction of distortion in the sampled sequences, we consider two sets of experimental data. The first set consists of 3 sequences of JPEG frames (100 frames each) that are taken from the same camera input. This first set is used to confirm the results of the algorithm in which all input sequences have similar image properties, but with different movements in each. The second set consists of 4 sequences (300 frames each) that are taken from different camera inputs. These sequences have different image properties, as well as different movements with the sequence.

Both sets of data are recorded with the uniform sampling and the proposed variable sampling approach with average sampling rates of $\overline{f}_s$=15, 10 and 5. For the variable sampling method, a window size of T=4, 6 and 12 were used for the different average sampling rates, respectively. The average Mean-Squared Error (MSE) is computed for each sampled input sequence under the various simulation conditions, where a zero-order hold of the last sampled frame is employed to reconstruct a skipped frame. The results for the first data set are shown in Figure 3, while the results for the second data set are shown in Figure 4.

From both sets of plots and the summary of total MSE in Table 1, it is clear that the overall distortion is significantly reduced with the variable sampling approach. The totals indicate that higher gains are achieved with higher average sampling rates. This is likely due to the fact that there is more flexibility in choosing the frames to be sampled from each sequence. Also worth noting is that the results are consistent across both data sets.

Looking more closely at the plots in Figure 3 and 4, we observe that the distortions of certain sequences are reduced to a much greater degree than others. This difference is accounted for in the characteristics of the sequences over time. For instance, both the Walk3 sequence in data set #1 and the Lab sequence in data set #2 contain little to no movement. As a result, the potential to reduce distortion is much less than for sequences with significant activity. On the other hand, the Walk2 sequence in data set #1 and the Walkway sequence in data set #2 are continuously active. In these cases, the variable sampling algorithm will allocate more frames to these sequences to have the most impact on reducing the overall distortion.

**Table 1**. Total MSE of fixed and variable sampling.

| Data Set | Sampling Algorithm | Average Sampling Rate | | |
|---|---|---|---|---|
| | | 15 | 10 | 5 |
| #1 | Fixed | 83.3 | 154.1 | 340.4 |
| | Variable | 29.3 | 70.2 | 202.4 |
| | % Reduction | 64.8% | 54.4% | 40.5% |
| #2 | Fixed | 102.8 | 190.1 | 374.9 |
| | Variable | 44.3 | 96.2 | 238.0 |
| | % Reduction | 56.8% | 49.4% | 36.5% |

## 5. CONCLUDING REMARKS

This paper introduced a new analysis problem that variably samples frames from multiple sequences of intra-coded frames. The objective of this work was to reduce the overall temporal distortion subject to a memory constraint. In the analysis, the temporal distortion was computed based on DC images, and a window-based approach to iterate through the various solutions has been proposed. Compared to uniform sampling, the overall distortion was significantly reduced.

In terms of future work, we plan to investigate the impact of using larger window sizes for the temporal analysis. We expect that further reductions in the overall distortion can be achieved. However, this will come at the cost of analyzing an increased number of possible solutions. Therefore, we also plan to study approaches that would minimize the computational complexity, while still maintaining a similar reduction in overall distortion.

## REFERENCES

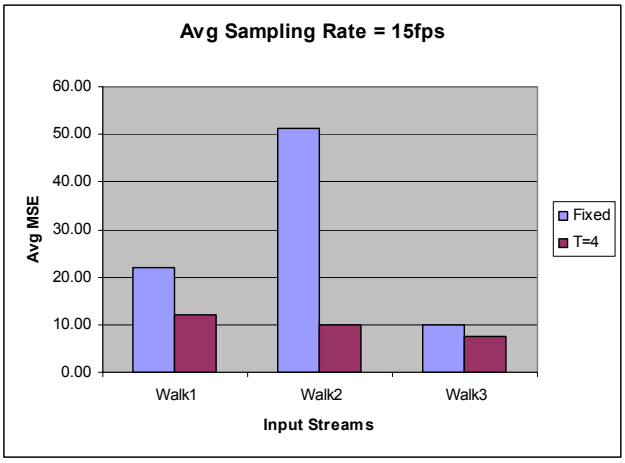
[1] F.C. Martins, W. Ding, E. Feig, "Joint control of spatial quantization and temporal sampling for very low-bit-rate video," *Proc. IEEE Int'l Conf. Acoustics, Speech, Signal Proc.*, Atlanta, GA, May 1996.

[2] A. Vetro, Y. Wang, and H. Sun, "Rate-distortion optimized video coding considering frameskip," *Proc. IEEE Int'l Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001.

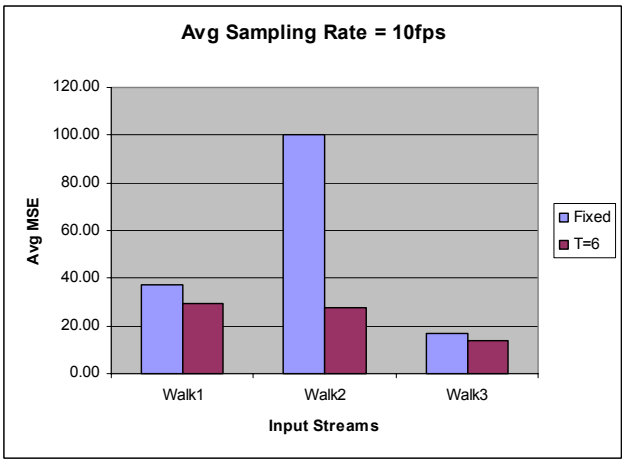
[3] L. Wang and A. Vincent, "Bit allocation for joint coding of multiple video programs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 949-959, Sept 1999.

[4] C.W. Hung and D.W. Lin, "Towards jointly optimal rate allocation for multiple videos with possibly different frame rates," *Proc. IEEE Int'l Symp. Circuits Syst.*, Geneva, Switzerland, May 2001.

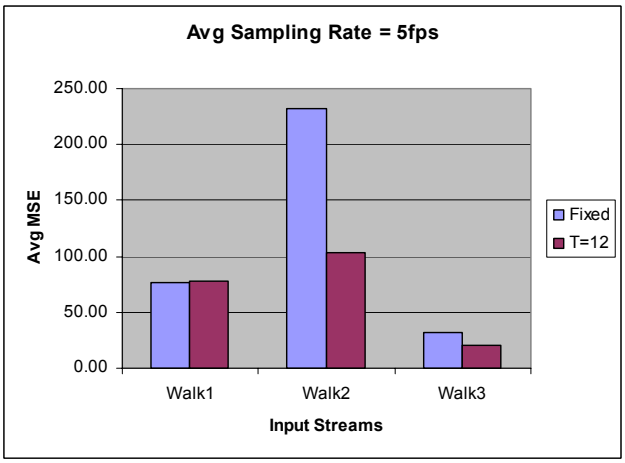
[5] P.J. Chase, "Algorithm 382: Combinations of M out of N Objects [G6]," *Communications of the ACM*, vol. 13, no. 6 pp. 368, June 1970.




(a)

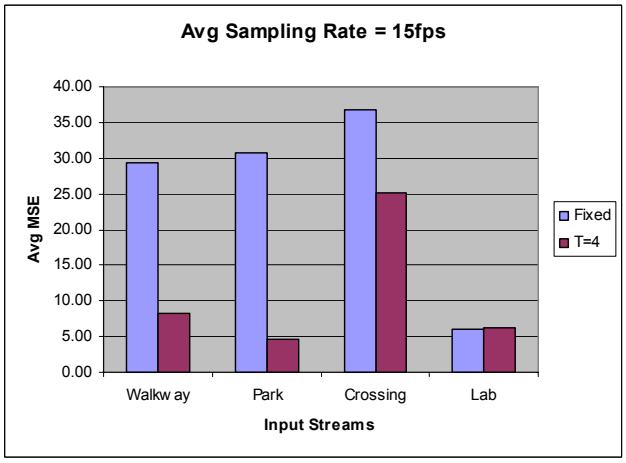



(b)

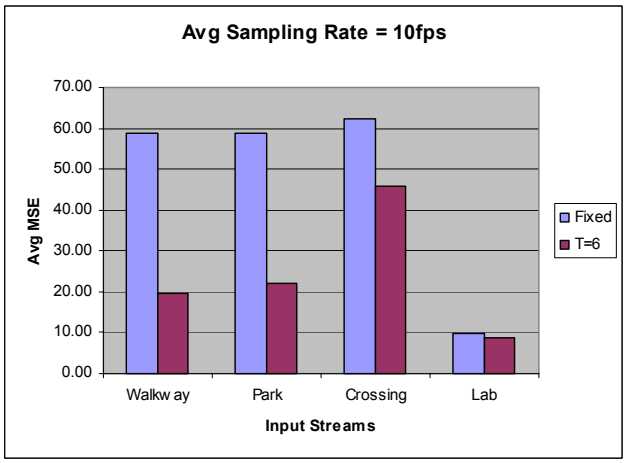



(c)

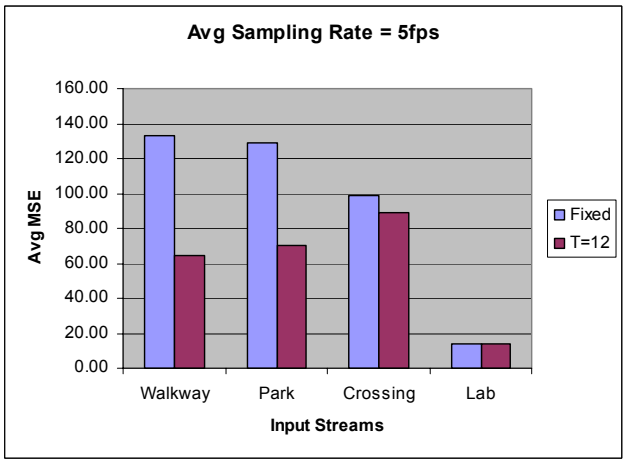
**Figure 3**. Simulation results comparing average MSE per input sequence (M=3) for uniform sampling and proposed variable sampling. (a) $\overline{f}_s$=15, T=4; (b) $\overline{f}_s$=10, T=6; (c) $\overline{f}_s$=5, T=12.




(a)




(b)




(c)

**Figure 4**. Simulation results comparing average MSE per input sequence (M=4) for uniform sampling and proposed variable sampling. (a) $\overline{f}_s$=15, T=4; (b) $\overline{f}_s$=10, T=6; (c) $\overline{f}_s$=5, T=12.