Requirements for motion-estimation search range in MPEG-2 coded video

by C. A. Gonzales
H. Yeo
C. J. Kuo

The motion-estimation search range required for interframe encoding with the MPEG-2 video compression standard depends on a number of factors, including video content, video resolution, elapsed time between reference and predicted pictures, and, just as significantly, pragmatic considerations in implementing a cost-effective solution. In this paper we present a set of experimental results that provide a probabilistic characterization of the size of motion vectors for different types of video, from well-known standard test sequences to fast-paced sports sequences to action movie clips. We study the impact of search range on compression efficiency and video quality. Finally, and on the basis of these results, we conclude with recommendations for target search ranges suitable for high-quality compression of standard and high-definition video.

1. Introduction

The most effective video compression standards [1–4] use the motion-compensated picture difference (MCPD) technique to achieve high degrees of compression at acceptable levels of picture quality. For the purposes of this paper, an MPEG-2 MCPD is generated by three steps:

1. Segmenting a target picture into a grid of macroblocks of 16 × 16 pixels.
2. Predicting the pixel values in each target macroblock by estimating the translational displacement (motion) between macroblocks in the target picture and macroblocks in one or two reference pictures.
3. Subtracting the target macroblocks from their predicted values to generate an MCPD picture.

Two types of MCPD pictures exist in MPEG-2: P- and B-pictures. P-pictures use only one reference picture, which is temporally located before the target picture. B-pictures use two references: one before and one after the target picture. A third type of picture in MPEG-2, the so-called I-picture, is not motion-compensated. Recent reviews of MPEG-2 include [5] and [6].

When generating an MPEG** compressed stream, it is common to define an m-parameter to indicate the distance in picture periods between target P-pictures and their corresponding reference pictures. An $m = 1$

0018-8646/99/$5.00 © 1999 IBM

sequence is made up of P- and I-pictures only, and each P-picture in the sequence is motion-compensated on the basis of the previous P- or I-picture. In an $m = 3$ sequence, two B-pictures are typically sandwiched between pairs of P-pictures. Because in MPEG only P- and I-pictures can be used as reference pictures, the largest distance between target and reference pictures occurs when P-pictures are encoded. It turns out that the larger the distance between target and reference pictures, the larger the displacements found in step 2 above. In the remainder of this paper, we attempt to characterize the statistics for the maximum size of these displacements; thus, we concern ourselves primarily with the prediction of P-pictures at $m = 3$.[1]

In the full-search method for motion estimation, the prediction step for a P-picture is performed by comparing each target macroblock against all possible candidate "matching" macroblocks in a search window centered around the macroblock in the same spatial location of the reference picture. A displacement motion vector of zero ($MV = 0$) indicates that the best-matching macroblock is located at the center of the search window. By definition, half of the horizontal and vertical sizes of this search window are the horizontal and vertical "search ranges" of the motion-estimation process. The displacement that corresponds to the best macroblock match defines the prediction macroblock motion vector whose value is transmitted as part of the MPEG-2 data. For interlaced video, in which alternating lines of pixels in a macroblock correspond to two separate video fields,[2] it is possible to improve the accuracy of the matching by separately predicting each field, at the cost of having to transmit two MV values instead of one. These options in MPEG-2 are known as adaptive frame/field motion compensation.

Puri et al. [7] suggested that a suitable search range in the MPEG-2 Main Profile at Main Level (MP@ML) is $[15 + 16(m - 1)]$. This empirical formulation suggests that for $m = 3$, all we require is a search range of ± 47 in both horizontal and vertical directions. As we will see, however, this range is insufficient for robust encoding in MP@ML. What then is the necessary motion-estimation search range required for effective P-picture encoding? The answer to this question depends on a number of factors, including video content, target video resolution, m-value, and pragmatic considerations of technology and cost limitations. Very little experimental data has been reported that helps a designer choose a search range under the constraints of limited chip size, or, equivalently, limited computation. What are the effects on bit rate and peak signal-to-noise ratio (PSNR) of MPEG-2 coded video with a constrained search range? What is the search range required for 95% or 99% macroblock coverage? What is the impact of a constrained search range on subjective video quality? These issues are all addressed in this paper.

Because an enormous number of computations are required for a full-search motion-estimation algorithm in MPEG-2 MP@ML, most hardware and software implementations of motion estimation are based on hierarchical techniques [8]. However, hierarchical methods also result in suboptimal motion-vector estimates. Furthermore, for a given computational capacity in a hierarchical approach, the larger the search range, the less optimal or *accurate* the vector estimates. A hierarchical motion-estimation design with computation or chip size limitations must thus balance the desire for a large search range that covers "all" possible cases, with the accuracy of the vector estimates that are suitable for the "majority" of cases.

Other factors affecting the quality of MPEG-2 encoding are the magnitudes of motion vectors and motion-vector differences in contiguous macroblocks. This is the case, for example, because "0" motion vectors can be very efficiently coded in MPEG-2 P-pictures, and also because MVs are coded using a differential pulse code modulation (DPCM) technique. Thus, simply choosing the MV with the "best match" (i.e., one with a minimum of motion-compensated pixel differences), without considering the absolute and differential size of the resulting MV value, can actually turn out to be a suboptimal choice. In this paper we study the requirements for motion-estimation search range in the context of a practical hierarchical implementation that takes MV sizes and differences into account.

2. A hierarchical search algorithm

The results in this paper were obtained from software simulations. Limitations of time and computation dictated that we also use a hierarchical motion-estimation algorithm for most of the simulation work. However, we believe that our conclusions remain valid for the case of other similarly hierarchical algorithms or even for the full-search algorithm. In what follows we briefly summarize the main features of the hierarchical algorithm used for this work. We label it the HS algorithm (for hierarchical search) as opposed to FS (for full search).

The HS algorithm is a refinement of an algorithm reported in [9]; it is a three-stage hierarchical approach. In the first stage, here referred to as *coarse search*, the dimensions of both the target and reference pictures are reduced by a factor of 4 in the horizontal direction. With HS, we maintain full vertical resolution to preserve the field structure of interlaced video. The FS algorithm is then applied to find the best *coarse* frame and field MVs for the reduced macroblock size of 16×4. As opposed to

[1] In practice, m-values larger than 3 are not used; therefore, they are not considered here.
[2] This paper focuses on frame-structure MPEG-2 encoding. However, the results are general and also apply to field-structure coding.

other hierarchical algorithms, in which the coarse search is performed on reduced images obtained by subsampling, HS decimates pictures by averaging pixels. This approach is more accurate and helps to reduce the effects of picture noise in the MV estimates. In the second stage, the horizontal component of the MVs is refined to one-pixel accuracy by using full-size target and reference pictures. Finally, the third stage further refines the resolution of the MVs to half-pixel resolution. The mean PSNR loss of HS compared with the FS algorithm is about 0.3 dB [10], which is arguably below the threshold of visibility for picture impairment.

Most of our simulations have been performed with a search range of either $\pm 192 \times \pm 168$ (frame-basis) or $\pm 320 \times \pm 280$, whatever was appropriate for the content at hand. The source video was made up of picture sequences of size 720×480. Larger search ranges were certainly not necessary for the video and film that we chose in our simulations, although we intentionally looked for difficult, fast-action content.

The cost functions we used to measure the accuracy of macroblock "matching" are empirical formulations which incorporate previously explained facts, i.e., that *optimum* is not only a function of minimizing the motion-compensated prediction error, but also a function of the magnitude of the motion vectors and the coded motion-vector differences.[3] The specific cost functions we used in this study are

$$\begin{aligned} CF(i,j) = {} & SAE(i,j) \\ & + w_x[|MV_x(i,j)| + |MV_x(i,j) - MV_x(i,j-1)|] \\ & + w_y[|MV_y(i,j)| + |MV_y(i,j) - MV_y(i,j-1)|] \end{aligned}$$

and

$$\begin{aligned} Cf_{\mathrm{t,b}}(i,j) = {} & SAE_{\mathrm{t,b}}(i,j) \\ & + \frac{w_x}{2}[|MV'_x(i,j)| + |MV'_x(i,j) - MV'_x(i,j-1)|] \\ & + \frac{w_y}{2}[|MV'_y(i,j)| + |MV'_y(i,j) - MV'_y(i,j-1)|], \end{aligned}$$

where CF is the cost function for a frame MV and $Cf_{\mathrm{t,b}}$ are the cost functions for the top- and bottom-field motion vectors of a macroblock located at *row*, *column* (i, j). SAE represents the sum of absolute values of the prediction error (in the case of field prediction, the SAE computation is based on half the number of pixels of the frame-prediction computation). MV_x and MV_y represent the horizontal and vertical components of frame motion vectors; MV'_x, MV'_y represent field motion-vector components measured in frame units. After substantial experimentation, we determined that $w_x = w_y = 4$ (see Footnote 3).

Finally, a decision must be made as to whether a macroblock is coded with frame or field motion compensation, or, alternatively, with no motion compensation (intra-macroblock). The criteria for making such a decision are based on an algorithm similar to that in Test Model 4 of MPEG-2 [11].

3. Simulation results

• *Interpretation of experimental results*

The horizontal and vertical components of motion vectors (x and y, respectively) behave like random processes; as such, their statistics can be characterized by their probability density functions, $f(x)$ and $f(y)$, or alternatively by their corresponding probability distribution functions, $F(x)$ and $F(y)$. An approximation of the density functions can be obtained from histograms of motion-vector components representing the frequency of occurrence of vector values for a picture or set of pictures. For the purpose of evaluating search-range requirements, we are more interested in the cumulative distribution functions of motion-vector components. **Figure 1** shows these functions for the well-known set of MPEG test sequences. Note that we evaluate the cumulative distribution functions for the absolute value of motion-vector components; this is because we are interested in investigating only motion-estimation search ranges that are symmetrical around 0.

It is important to understand how these plots were derived in order to interpret them correctly. As previously explained, macroblocks in a P-picture can be coded as intra-, frame-predicted, or field-predicted. In the case of field-predicted macroblocks, there exist two motion vectors per macroblock. In the case of frame-predicted macroblocks, only one motion vector per macroblock is used, whereas no motion vectors are specified for intra-macroblocks. Since we wish to make "motion-vector statistics" correspond to actual "picture area statistics," we use the following rules in deriving f and F from the experimental data:

1. Frame-motion vectors are counted twice in calculating f or F.
2. Intra-macroblocks are counted as two zero-motion vectors in calculating F.

In this manner, when calculating F, two motion vectors are always used for *every* macroblock in a P-picture. Thus, a value of $MV_x = x_0$, corresponding to $F(MV_x) = 0.9$, indicates that 90% of the total pixel area in P-pictures can be "coded" with motion vectors $MV_x \leq x_0$ (note that this 90% includes "unpredicted" intra-macroblocks). We refer

[3] C. Gonzales, J. Kouloheris, W. Lam, H. Yeo, and C. J. Kuo, "A Family of Cost Functions for Motion Estimation in MPEG-2," work in preparation.

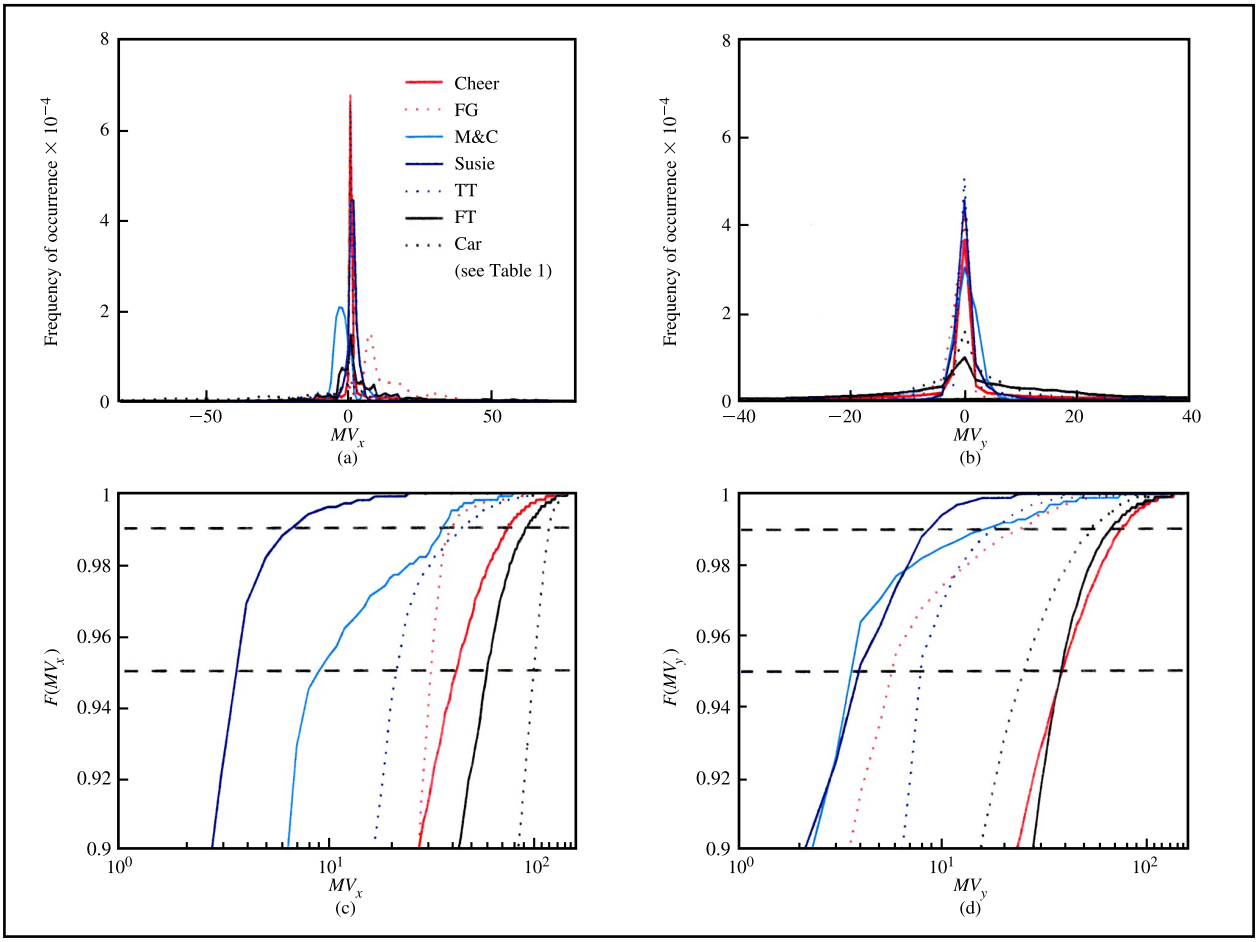

Figure 1

MPEG test sequence motion vector statistics: (a) and (b) respectively show the histogram of the horizontal and vertical components; (c) and (d) are the corresponding estimated distribution functions (95% and 99% statistics are indicated with dotted lines).

to x_0 as the 90% probability search range for the x-component of motion vectors.

- *Short-term versus long-term statistics*

To experimentally describe the f and F statistics, we must collect macroblock motion-vector measurements for one or more pictures. The elapsed time of the observations is very important. As one might expect, the statistics are heavily dependent on video content, and their behavior tends to appear stationary only from one scene change to the next. It should be clear, for example, that measuring the long-term statistics of motion vectors for the length of a two-hour movie will tell us very little about the short-term statistics of each individual scene in the movie. In this paper we are interested in both long-term and short-term behavior of motion-vector statistics. Long-term statistics provide us with an indication of the value of the required motion-estimation search range for effective video compression, where by *effective* we mean compression with optimum video quality "most of the time." Short-term statistics provide us with an indication of the value of the required motion-estimation search range for robust video compression, where by *robust* we mean compression with close-to-optimum video quality[4] "all the time." Clearly, robust encoding is a desirable objective; however, one must deal with practical considerations such as the tradeoff between accuracy and search range when limited computational power is available.

We report on the results of simulation studies for several video sequences. The sequences we tested comprise three different groups: 1) the MPEG-2 set of

[4] In this paper we measure video quality by either peak signal-to-noise ratio (PSNR) or a subjective evaluation of picture quality.

test sequences, with which many practitioners of the MPEG-2 standard are familiar; 2) a set of sports-related test sequences which were deliberately chosen to stress the requirements for large search ranges; 3) several minutes of 24-frame-per-second film material, representative of typical action movie content.

• *Statistics for standard MPEG sequences*

Table 1 lists the MPEG test sequences we used in our experiments, including our attempt to describe their content. The term *simple motion* means few objects moving at slow speeds; *complex motion* means one or more objects moving at moderate to high speeds; *zoom* and *camera panning* are self-descriptive. We measured motion-vector statistics for sixty pictures in each test sequence at $m = 1$ and $m = 3$.

While we are interested primarily in $m = 3$, it is useful to observe how search-range requirements scale with m. For most sequences, motion vectors do not scale linearly with m, as one might erroneously assume. In fact, **Table 2** shows that only in the case of simple motion and camera panning is the scaling approximately linear. For all other cases, the motion-vector search range required for $m = 3$ is typically less than the linear rule would predict.

The picture-by-picture statistics for these sequences are shown in **Figure 2**. In this figure we show the average motion-vector component, the 95% probability search range, and the 99% probability search range, for the horizontal and vertical components at $m = 3$. Also shown are the percentage of intra-macroblocks (note that, as one would expect, the number of intra-macroblocks correlates well with the amount of motion). Each of these sequences corresponds to a single video scene, and each of them shows an approximately statistically stationary behavior. Their overall statistics have already been presented in Figure 1. We observe in Table 2 that for even a 99% probability search range, $MV_x \leq 120$ and $MV_y \leq 72$ in all cases. To achieve 99% coverage for MV_x, only Carousel requires a search range of ±120. This means that when we limit the horizontal search range to ±120, fewer than 1% of the macroblocks in this sequence may become "unpredictable" and may have to be coded as intra-macroblocks, thus adding to the roughly 6% intra-macroblocks required by these sequences to start with. This small increment has an imperceptible impact on coding efficiency or video quality, as we later see. To achieve 95% coverage, we see from Figure 1 and Table 2 that $MV_x \leq 100$ and $MV_y \leq 40$. Once again, only Carousel requires $MV_x = 100$; if we limited the search range to this value, the number of intra-macroblocks for this sequence could potentially double to about 10% (see Figure 2). As we later see, objective and subjective measurements of video quality suggest that, even at 95% search range, video quality degradation appears to be below the threshold of human perception.

Table 1 Description of MPEG-2 test sequences.

Motion sequence	*Simple motion*	*Complex motion*	*Zoom (in/out)*	*Camera panning*
Cheerleaders		✓		
Flower Garden	✓			✓
Mobile and Cal.	✓		✓	✓
Susie	✓			
Table Tennis	✓		✓	
Football		✓		✓
Carousel		✓		✓

• *Statistics of sports sequences*

Action sports stress the requirements for motion-estimation search range. For this reason we chose to simulate a set of six different sports clips of interlaced 60-Hz video, as shown in **Table 3**. We captured 30 seconds for each clip, and sampled two pictures every 0.5 second. These two pictures were separated by three picture periods such that the simulation results correspond to $m = 3$. The contents of each sequence are described in Table 3. The term *close-up shot* indicates that the height of the person or persons that the camera is tracking is greater than one half of the picture height; otherwise we

Table 2 Experimental results for $m = 1$ and $m = 3$.

	$F(MV_x) = 0.95$		$F(MV_y) = 0.95$		$F(MV_x) = 0.99$		$F(MV_y) = 0.99$	
	$m = 1$	$m = 3$	$m = 1$	$m = 3$	$m = 1$	$m = 3$	$m = 1$	$m = 3$
Cheerleaders	16	42	14	40	38	75	38	72
Flower Garden	11	32	2	6	15	40	8	20
Mobile and Cal.	3	9	2	4	28	36	4	16
Susie	2	4	1	4	4	7	3	10
Table Tennis	8	22	6	8	19	44	14	18
Football	39	60	16	40	52	92	32	68
Carousel	38	100	10	26	55	120	40	54

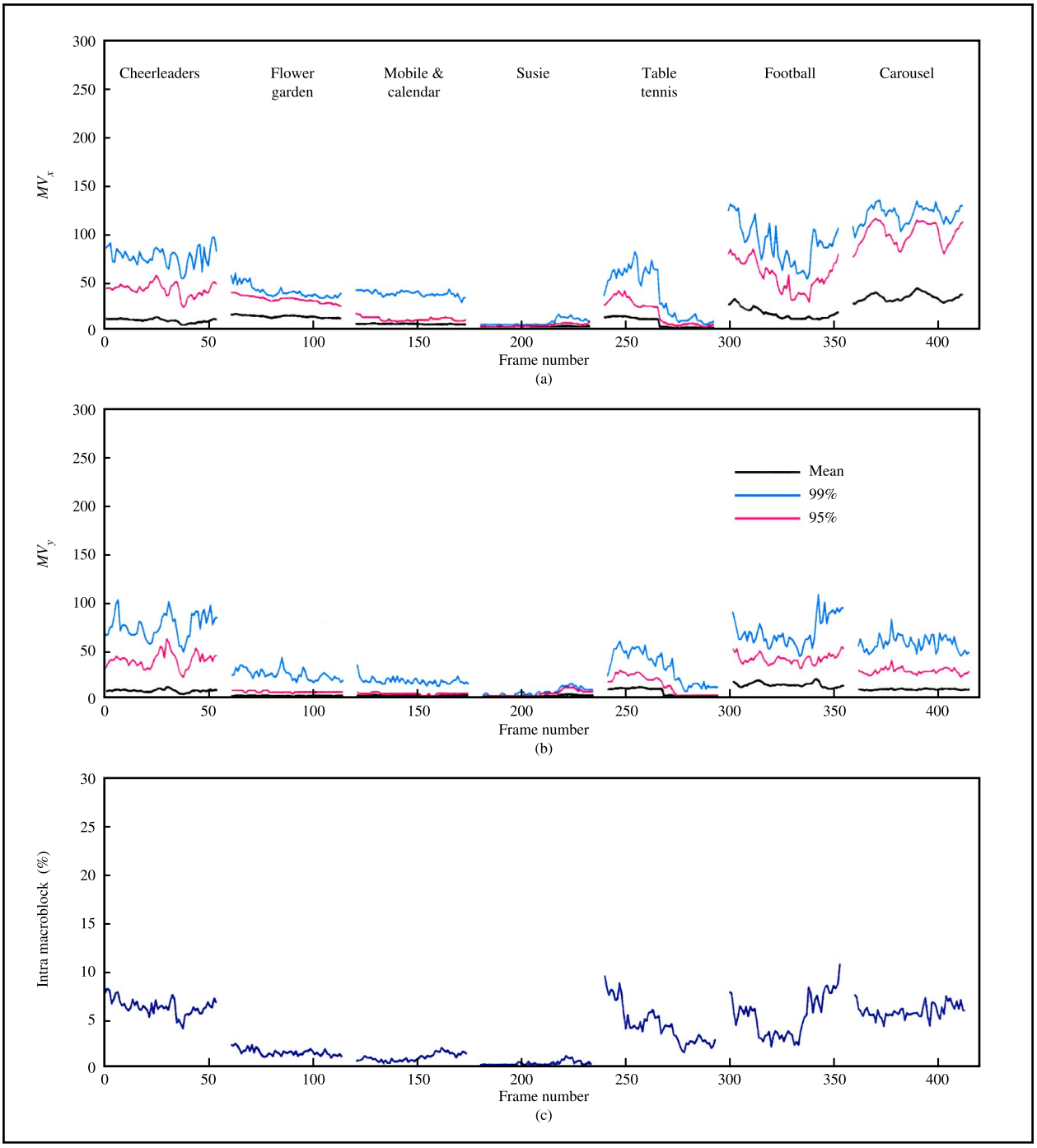

Figure 2

MPEG test sequence motion vector statistics for $m = 3$: (a) Horizontal; (b) vertical; (c) intra macroblock (%).

label it a *long shot*. In close-up shots, tracking a moving object can result in extremely fast panning of the more distant background.

Figure 3 shows the picture statistics for all of these sequences. The two basketball sequences appear to be the most demanding in terms of horizontal and vertical

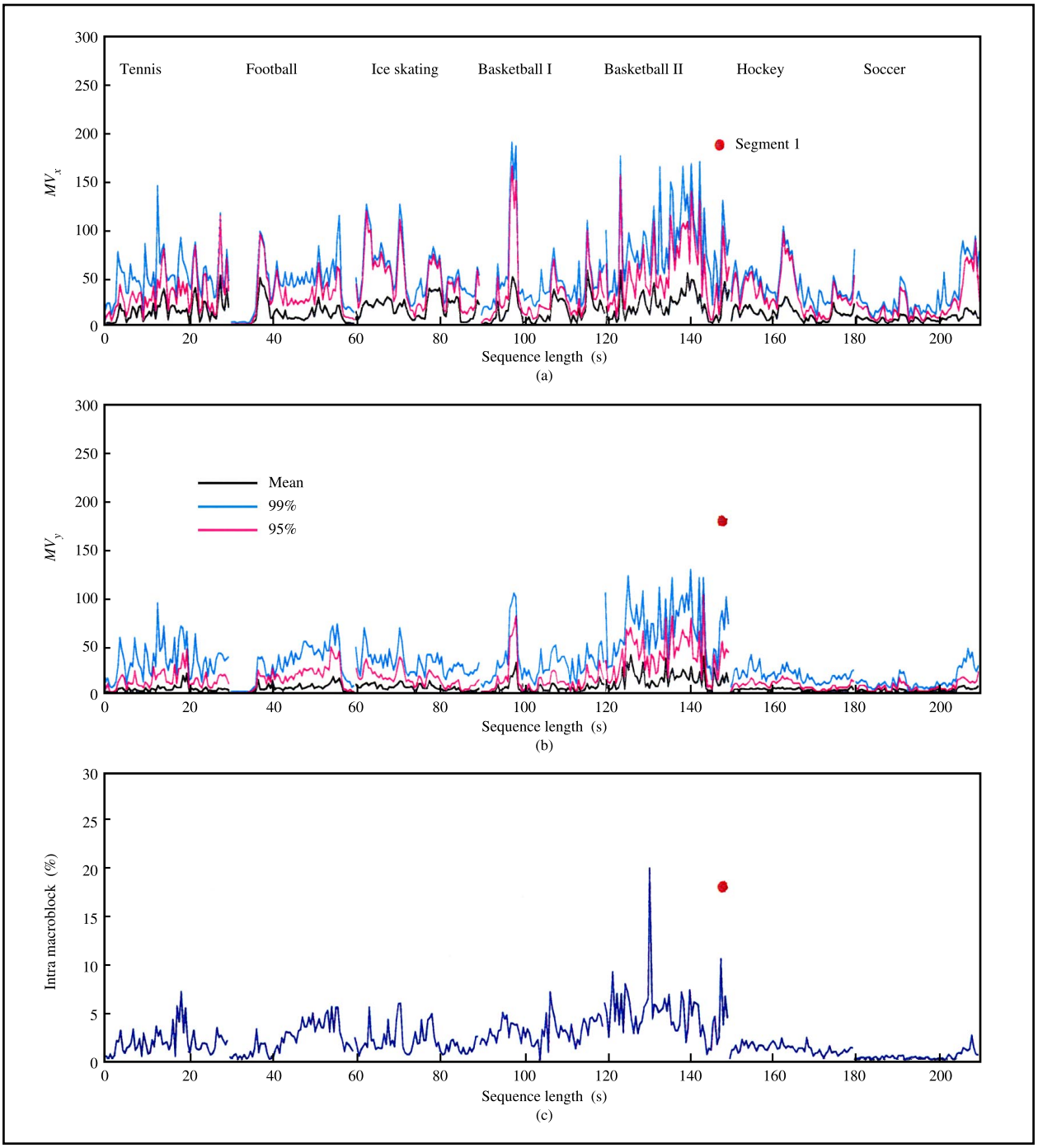

Figure 3

Sports sequence motion vector statistics for $m = 3$ (each sequence is 30 seconds long): (a) Horizontal; (b) vertical; (c) intra macroblock (%).

motion. Also identified in this figure is a portion of video with large motion vectors which is indicated as Segment 1. The short-term statistics of this segment are studied in more detail in Section 4. To compare "Basketball I" and "Basketball II," the close-up sequence requires a larger overall search. This is seen more easily in **Table 4** and

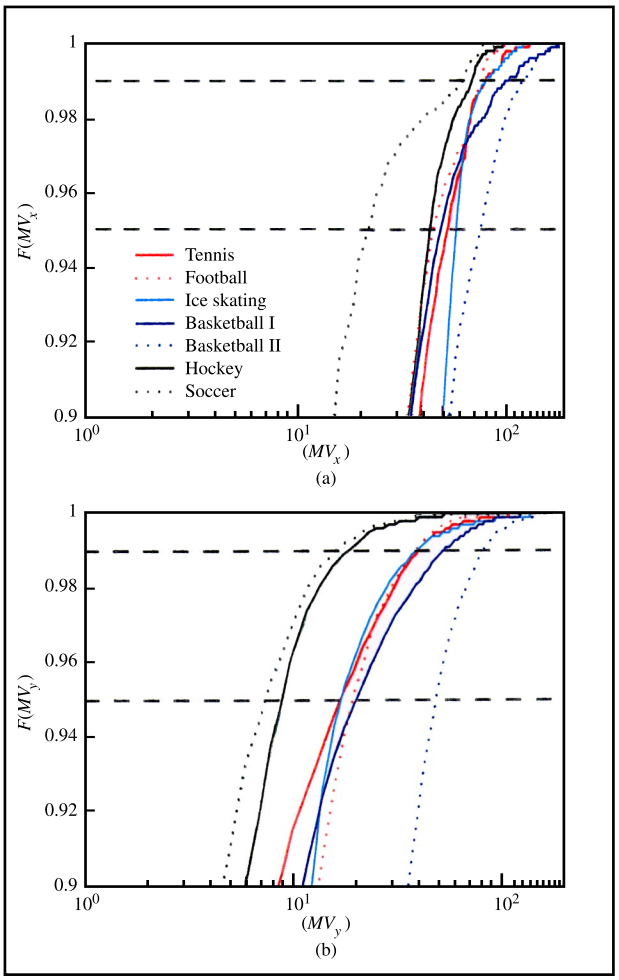

Figure 4

Long-term statistics for sports sequence. The 95% and 99% statistics are indicated with dotted lines.

Figure 4, where long-term (30-second) statistics for each sequence are shown. The overall 99% probability search range requirement is $MV_x \simeq 123$ and $MV_y \simeq 84$, whereas the 95% probability search range is $MV_x \simeq 77$ and $MV_y \simeq 50$. Both of these ranges are defined by the statistics of "Basketball II."

Table 3 Description of sports sequence contents.

Motion sequence	*Close-up shot*	*Long shot*	*Zoom (in/out)*
Tennis	✓		✓
Football II	✓	✓	✓
Ice Skating	✓		
Basketball I		✓	
Basketball II	✓		
Hockey	✓	✓	
Soccer		✓	

- *Statistics of movie sequences*

To complete our experiments, we simulated film at 24 pictures per second. This content was provided to us through the courtesy of a movie studio, as an example of typical action material. Four different movie clips with different time lengths, labeled Movie 1 through Movie 4, were used. We sampled pairs of *progressive* pictures corresponding to $m = 3$ (after telecine inversion) every half second. We intentionally avoided those cases in which the two pictures in a pair belonged to different scenes (these cases should not be handled with MCPD techniques). The clips contain a variety of material ranging from low motion, e.g., people talking, to extreme motion, e.g., close-up of horseback riding scene.

Picture statistics for these clips are shown in **Figures 5–7**. Long-term averages are shown in **Table 5** and **Figure 8**. The largest averages correspond to Movie 2, which has a 99% macroblock coverage with a search range of $MV_x \simeq 103$ and $MV_y \simeq 61$. In contrast, the corresponding 95% range is $MV_x \simeq 46$ and $MV_y \simeq 35$.

In Figures 5–7 we have also identified a number of segments that significantly exceeded the 99% search range for Movie 2; they are labeled Segment 2, Segment 3, and Segment 4. Segment 2 was chosen because of its large vertical motion. In the next section of this paper, we analyze the short-term statistics of these segments more closely.

Table 4 Experimental results for $m = 3$. The 95% and 99% probability search ranges are compared.

	$F(MV_x) = 0.95$	$F(MV_y) = 0.95$	$F(MV_x) = 0.99$	$F(MV_y) = 0.99$
Tennis	54	18	82	42
Football	46	20	78	40
Ice Skating	59	18	82	40
Basketball I	50	22	102	54
Basketball II	77	50	123	84
Hockey	44	10	71	20
Soccer	22	8	64	18

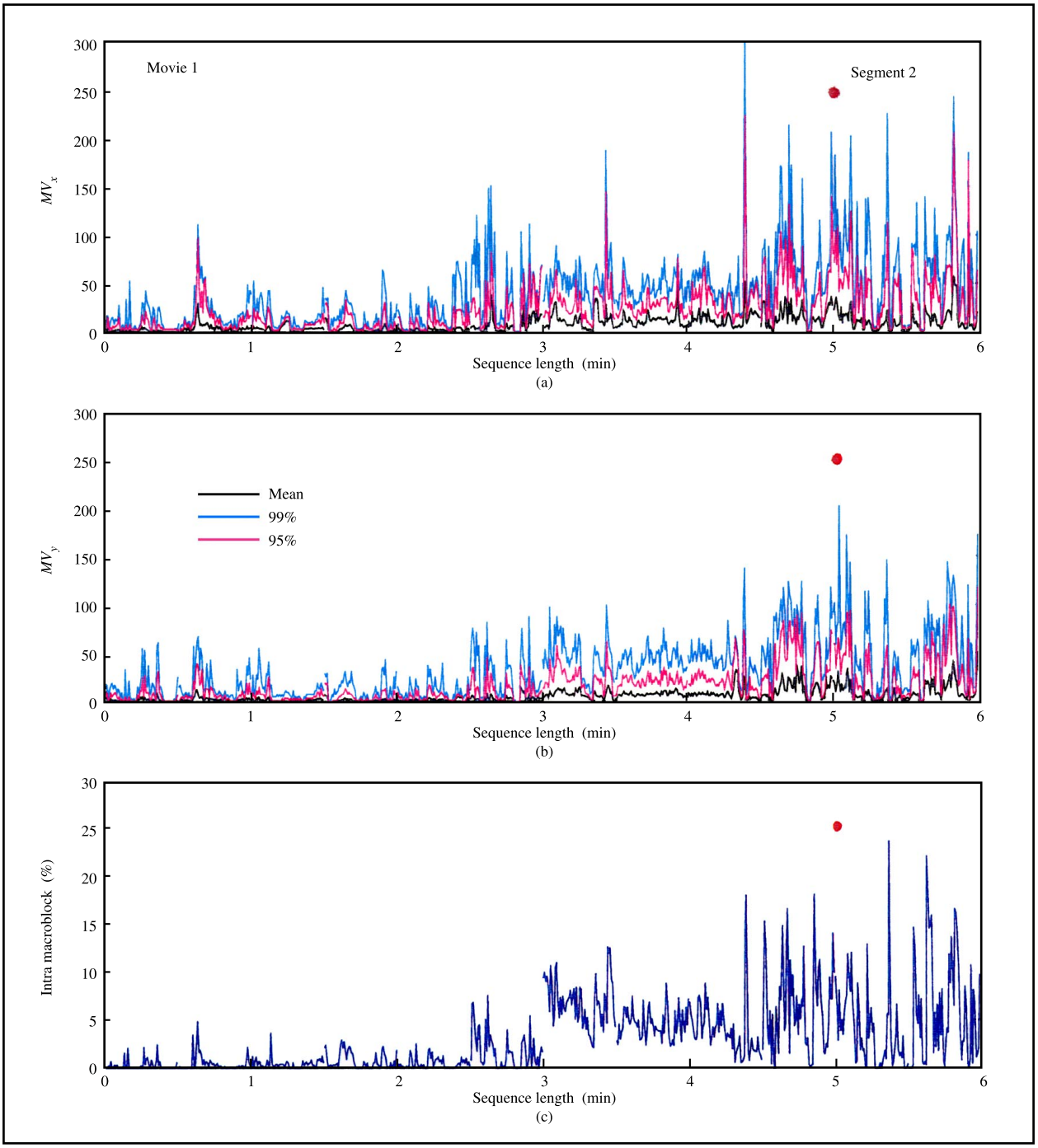

Figure 5

Movie 1 motion vector statistics for $m = 3$: (a) Horizontal; (b) vertical; (c) intra macroblock (%).

4. Picture quality and coding efficiency with a constrained search range

The experiments presented in Section 3 showed that the most demanding content in terms of motion-estimation search range was that of sports video. On the basis of those measurements, we suggest that a search range somewhere between the 95% and the 99% statistics of our sports video clips is sufficient to guarantee close-to-

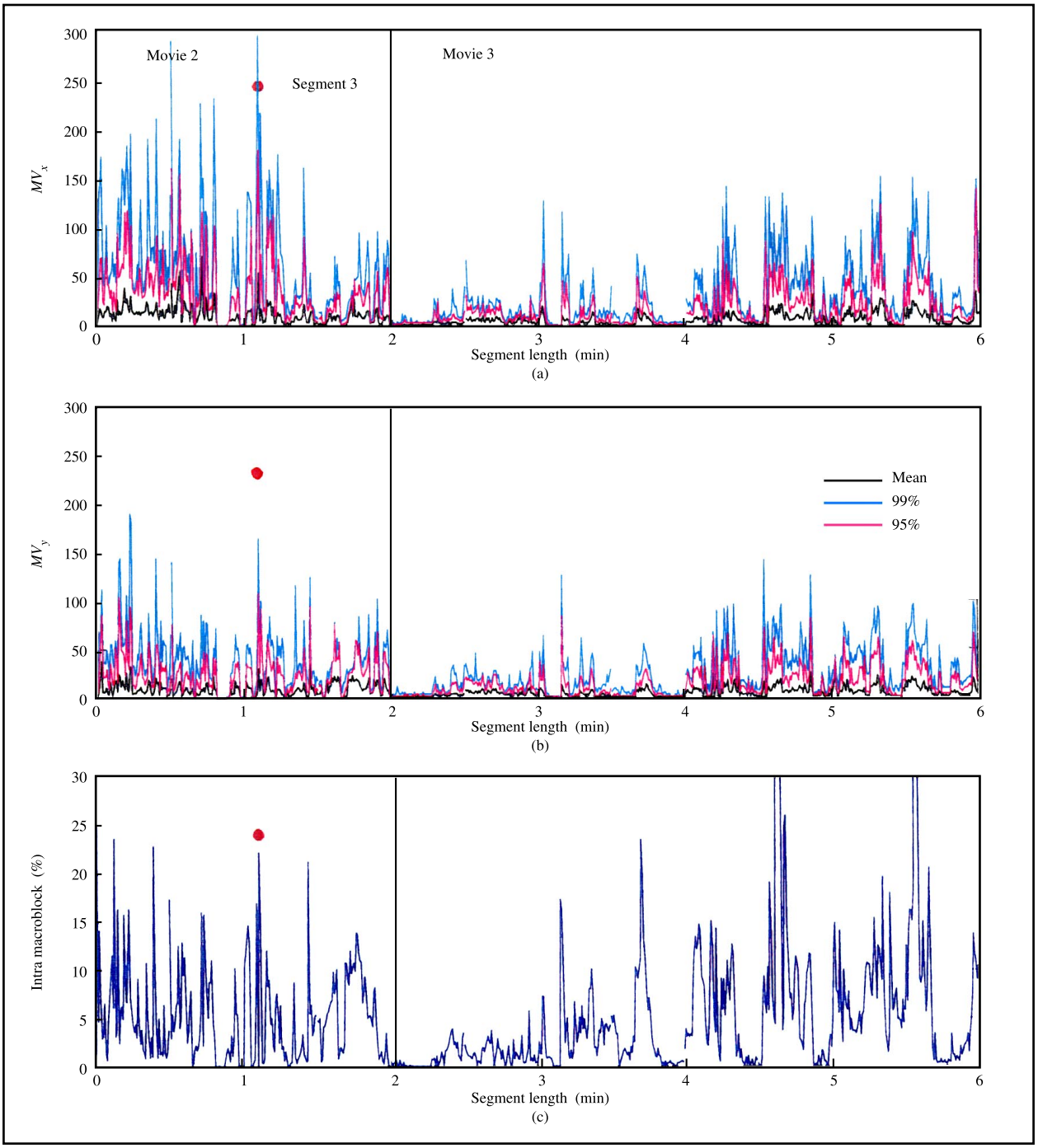

Figure 6

Movie 2 and Movie 3 motion vector statistics for $m = 3$: (a) Horizontal; (b) vertical; (c) intra macroblock (%).

optimum coding results for even critical applications. More concretely, we recommend a search range in the interval

$$80 \le MV_x \le 120, \qquad 50 \le MV_y \le 85$$

for all applications of MPEG-2 encoding of CCIR 601 video resolution. We observe, however, that over short periods of time, short-term statistics can exceed this recommended

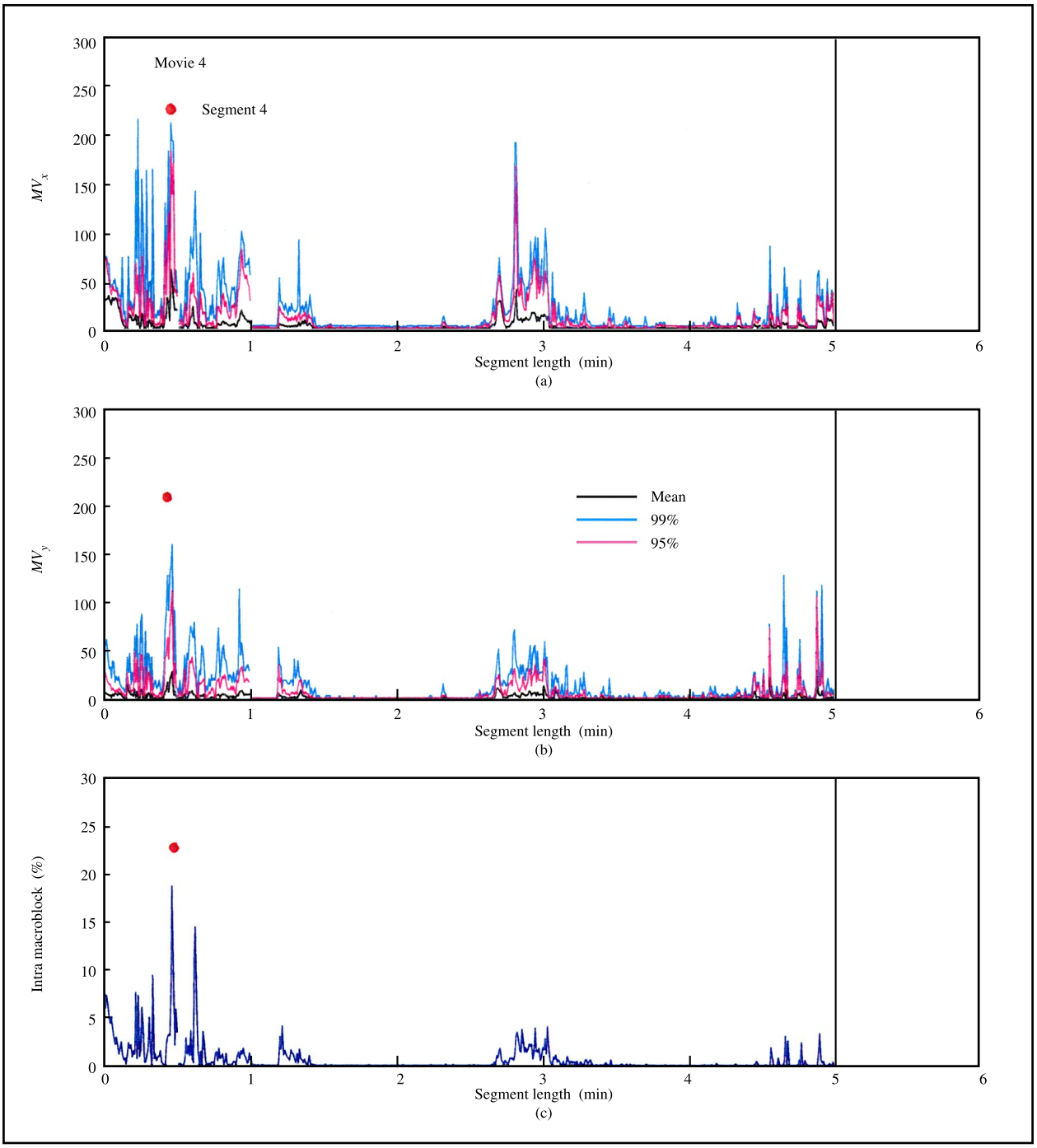

Figure 7

Movie 4 motion vector statistics for ($m = 3$): (a) Horizontal; (b) vertical; (c) intra macroblock (%).

search range significantly. Several examples were singled out in Section 3 and labeled as Segments 1–4. Using these segments and other video samples, we now wish to study the impact on picture quality (or, alternately, coding efficiency) of MPEG compression with a constrained search range. We study this impact with subjective and objective measures.

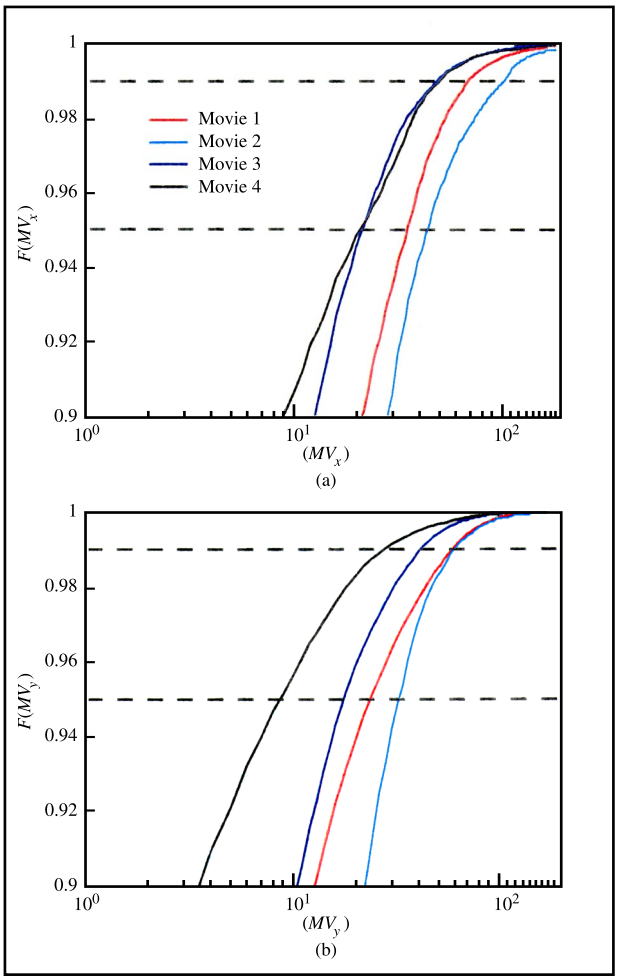

Figure 8

Long-term statistics for movie clips; 95% and 99% statistics are indicated with dashed lines.

Factors affecting picture quality in MPEG-2 coding are numerous, and their interaction is complex. Some of the factors that relate to the presence of motion include 1) motion-estimation search range; 2) human perception in the presence of fast motion; 3) picture complexity (motion tends to produce blurred images of low complexity); 4) number of unpredicted (intra-) macroblocks; and 5) data rate. Predicting the effect of a constrained search range in the presence of fast-moving images is difficult because of the complex interactions of all of these factors. For example, does it matter if a few extra macroblocks cannot be motion-compensated when most of a picture is blurred and of low complexity? What is the impact on image quality of adding a few more unpredicted macroblocks to an already significant percentage of intra-macroblocks? Even if all of these elements matter quantitatively, do they matter subjectively? That is, can the eye perceive the impact on quality in the presence of fast motion? At what data rate can the human eye perceive these effects? In practice, these complex interactions cannot be predicted; they can only be measured by experiments. That is the goal of this section.

- *Objective measurements of video quality and compression efficiency*

In this section we measure the impact on PSNR of compressing video at constant bit rates as a function of motion-estimation search range. We also present an alternative view: the impact on compressed bit rate when coding with a constrained search range at constant PSNR. While we recognize that PSNR is not a perfect measure of video quality, relative values of PSNR do correlate with subjective evaluation of quality. This is particularly true for PSNR values below 38 dB; above 38 dB, video tends to be of "high" quality, and changes in PSNR are very difficult to observe.

We focus on a couple of examples from the MPEG-2 test sequences, as well as on the short segments that we identified in Section 3 as cases that fall outside the search range we recommend. From the MPEG test set we choose two sequences, "Carousel" and "Mobile and Calendar," as examples of high motion and low motion, respectively. (We have experimented with the remainder of the sequences and find that these two are most representative of these two extremes.) **Figures 9** and **10** respectively show the results of our simulations for the MPEG-2 test sequences and the selected video segments. The sequences in Figure 9 are more difficult to compress than those in Figure 10: We observe that at 4 Mbps the sequences in Figure 9 result in a PSNR well below 38 dB, whereas those in Figure 10 result in a PSNR well above 38 dB.

Table 5 Experimental results for $m = 3$. The 95% and 99% probability horizontal search ranges are compared.

	$F(MV_x) = 0.95$	$F(MV_y) = 0.95$	$F(MV_x) = 0.99$	$F(MV_y) = 0.99$
Movie 1	37	25	72	61
Movie 2	46	35	103	61
Movie 3	23	19	51	43
Movie 4	22	11	53	29

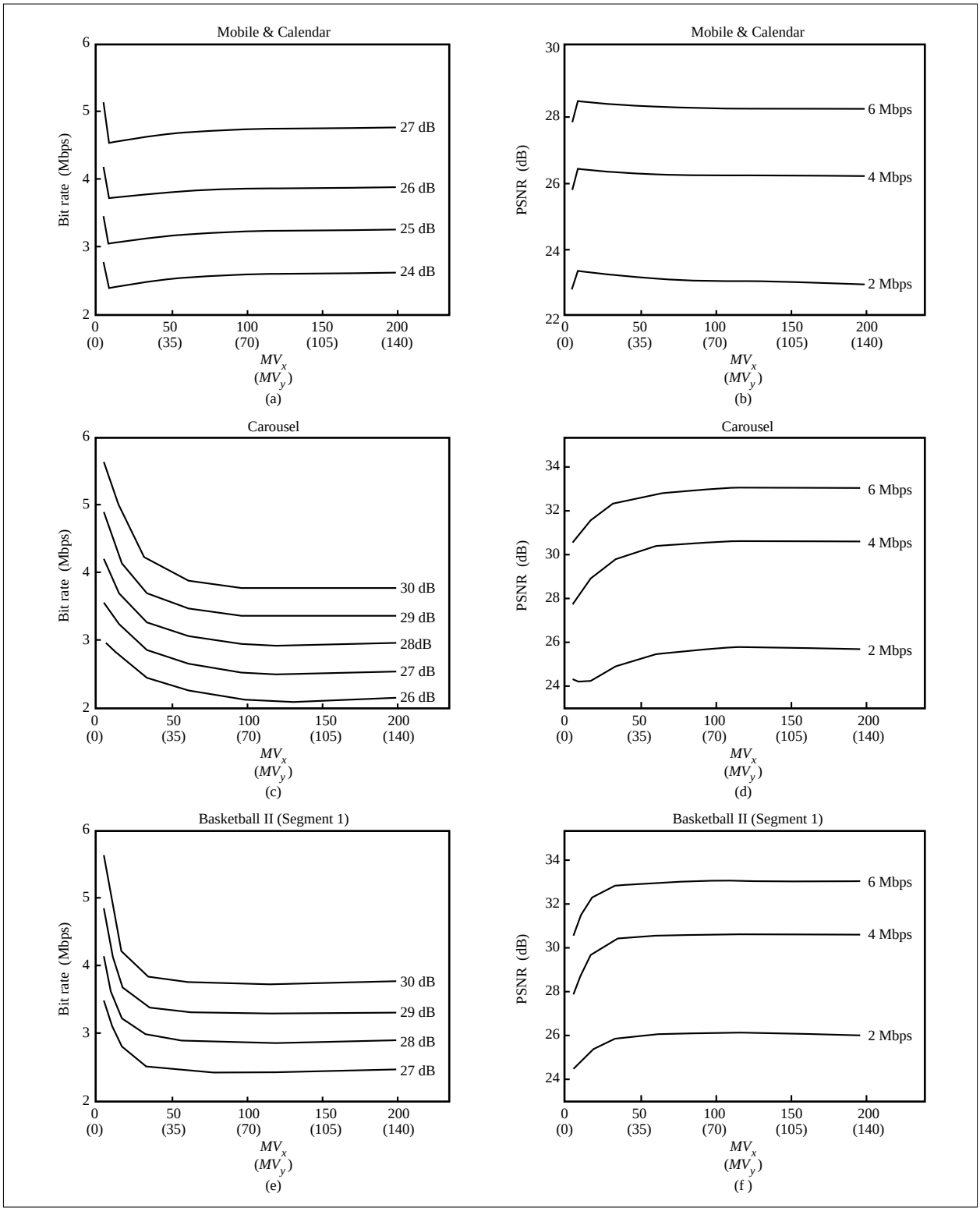

Figure 9

Two views of video quality (PSNR) as a function of search range (MV_x, MV_y): Search range is changed maintaining an approximately constant horizontal-to-vertical component ratio. In this figure Mobile & Calendar, Carousel, and Basketball are shown.

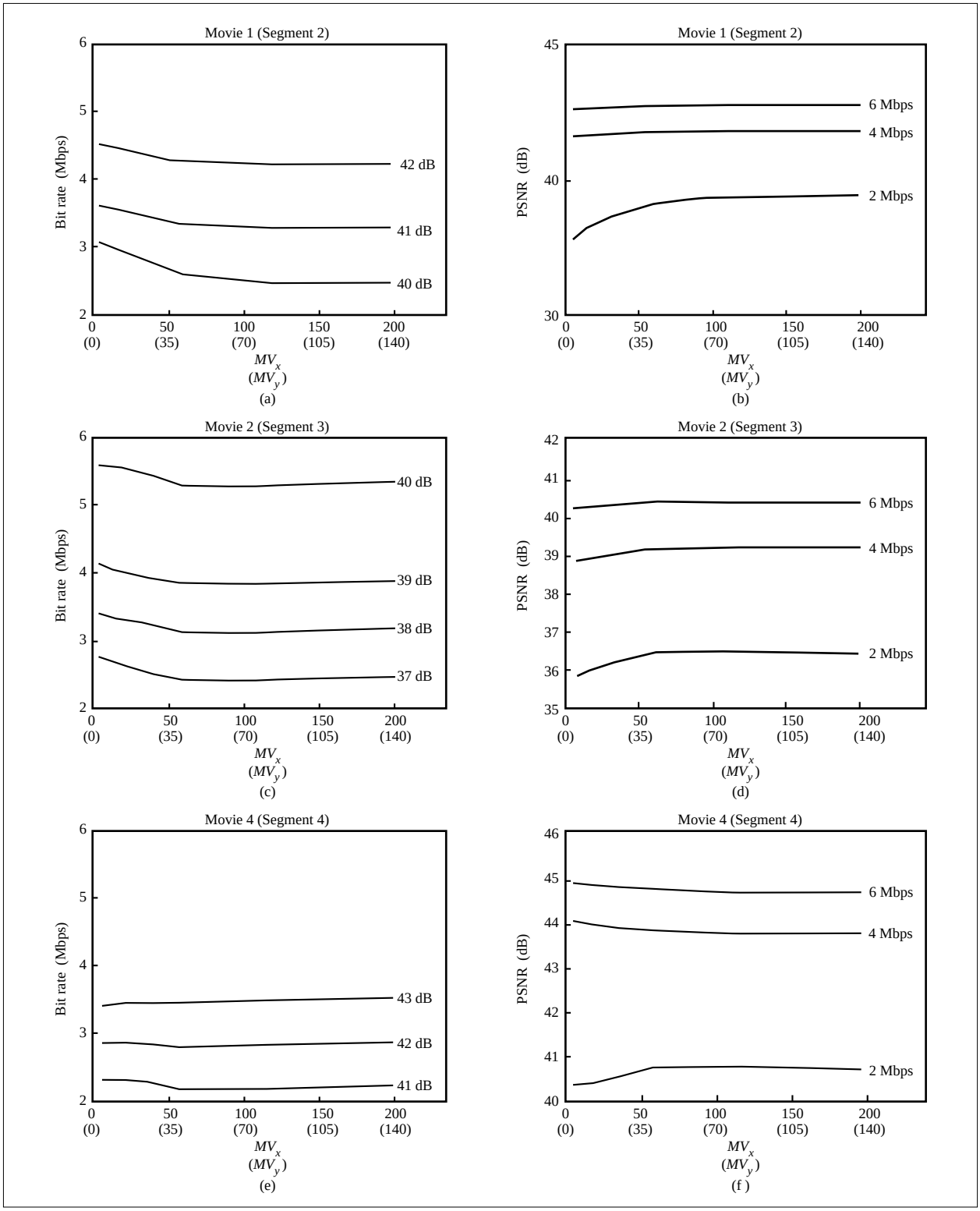

Figure 10

Two views of video quality (PSNR) as a function of search range (MV_x, MV_y): Search range is changed maintaining an approximately constant horizontal-to-vertical component ratio. In this figure segments 2–4 from the movie sequences are shown.

The results in Figures 9 and 10 are based on actual MPEG-2 encoding experiments at search ranges of 4 × 4, 8 × 6, 16 × 12, 32 × 22, 60 × 42, 100 × 70, 128 × 90, and 200 × 140. For all of these, the ratio of horizontal to vertical search range is approximately constant and proportional to 100/70.

It is interesting to note that movie segments 2–4 are the "easiest" to compress, despite having the largest apparent requirement for search range. Even at 2 Mbps, these segments result in a PSNR greater than 36 dB, compared to 30 dB or less for the sequences in Figure 9. Figure 10 also shows that constraining the search range to 100 × 70 has an insignificant impact on PSNR or data rate.

Of the sequences in Figure 9, "Segment 1" (Basketball II) shows the most sensitivity to search range, particularly at low data rates. The loss of PSNR, due to our recommendation for constrained search range, is limited to less than 0.3 dB. In fact, at a 100 × 70 search range, the loss is 0.1 dB or less. One interesting artifact is seen in Figure 9(a): It appears that for "Mobile and Calendar" the coding efficiency or video quality, as measured by PSNR, actually *decreases* with increased search range! We have noticed this effect in almost all other sequences when the search range becomes larger than the 99% statistics of the sequence. Careful examination of the data shows that this effect can be explained by a combination of our choice for motion-estimation cost function (Section 2) and the frame/field/intra-coding decision algorithm. What happens is that with a larger search range, larger motion vectors are being selected, and more bits are being used to code the motion-vector differences. These additional bits, however, are not being sufficiently compensated for by the corresponding savings in coding smaller macroblock residual errors. We can think of other cost and other intra/inter-macroblock decision functions that could avoid this paradoxical behavior. However, such functions are generally not of practical value. One example is an algorithm in which we optimize the global picture PSNR at a given target bit rate by trying out various combinations of MPEG-2 coding and motion-compensation modalities for individual macroblocks. Such an algorithm, however, is simply impractical. As we pointed out in Sections 1 and 2, in this paper we are interested in practical results; thus, we believe that this paradoxical effect will always be present with practical implementations of MPEG-2 encoding and motion compensation. Fortunately the effect is small and, as far as we can tell from visual experiments, below the threshold of visibility.

- *Subjective measurements of video quality*

We performed limited subjective evaluation experiments using fourteen observers. Included among these fourteen were five with experience in MPEG video coding. Our experiments followed closely the setup and procedures used in the MPEG committee [12]. A subjective rating from 0 to 5 was assigned in conformity with the following quality scale: Bad (0–1), Poor (1–2), Fair (2–3), Good (3–4), and Excellent (4–5).

The experiment comprised a number of sessions. For any one session we tested multiple data rates and multiple search ranges for one video segment and one observer. A session consisted of presenting the observer with multiple pairs of sequences, in which each pair consisted of one reference sequence and one test sequence in random order. Pairs of sequences were presented twice: the first time for previewing, followed immediately by a second time during which observers registered a quality score. All reference sequences were MPEG coded with a search range of 200 × 140, and at the same bit rate as the corresponding test sequence. Three data rates (2, 4, and 6 Mbps) and three search ranges (30 × 22, 60 × 42, and 100 × 70) were chosen for the test sequences; thus, a total of nine pairs were possible for each video segment. In addition, we chose to repeat each pair (but in reverse order of presentation) so as to test the consistency of the observer's evaluations. Thus, for each session, eighteen pairs were presented to observers in a random order. In addition, and in order to evaluate the effects of coding impairment at the test data rates (without including limitations of search range), we also paired the uncompressed originals against the reference sequences.

We asked the assessors to judge the relative and absolute quality of the sequences in each pair by marking a score sheet during the scoring portion of the presentation (we refer again to [12] for more details). To help the assessors calibrate their evaluation of quality, they were first shown the original sequence without degradation and then the most degraded sequence, prior to beginning the actual rating experiment. After this, observers proceeded with the formal testing, in which, as mentioned above, they did not know the order of presentation, or, for that matter, the specific coding parameters for which we were testing.

We calculated the average statistics for all observers. The results for three representative video segments are shown in **Figure 11**; these are the same as "Carousel," "Segment 1," and "Segment 2" in Figures 9 and 10. The *Diff* data shown in Figure 11(a) were obtained by subtracting the reference sequence score from the test sequence score for each pair and for each observer. This figure uses asterisks to show the average values of *Diff* for all observers; also shown are the one-sigma intervals (68% confidence intervals) for these statistics. Positive values of *Diff* mean that, on the average, observers rated the test sequences as being of higher quality than the reference sequences. On the other hand, negative values indicate that observers were able to perceive impairments in the

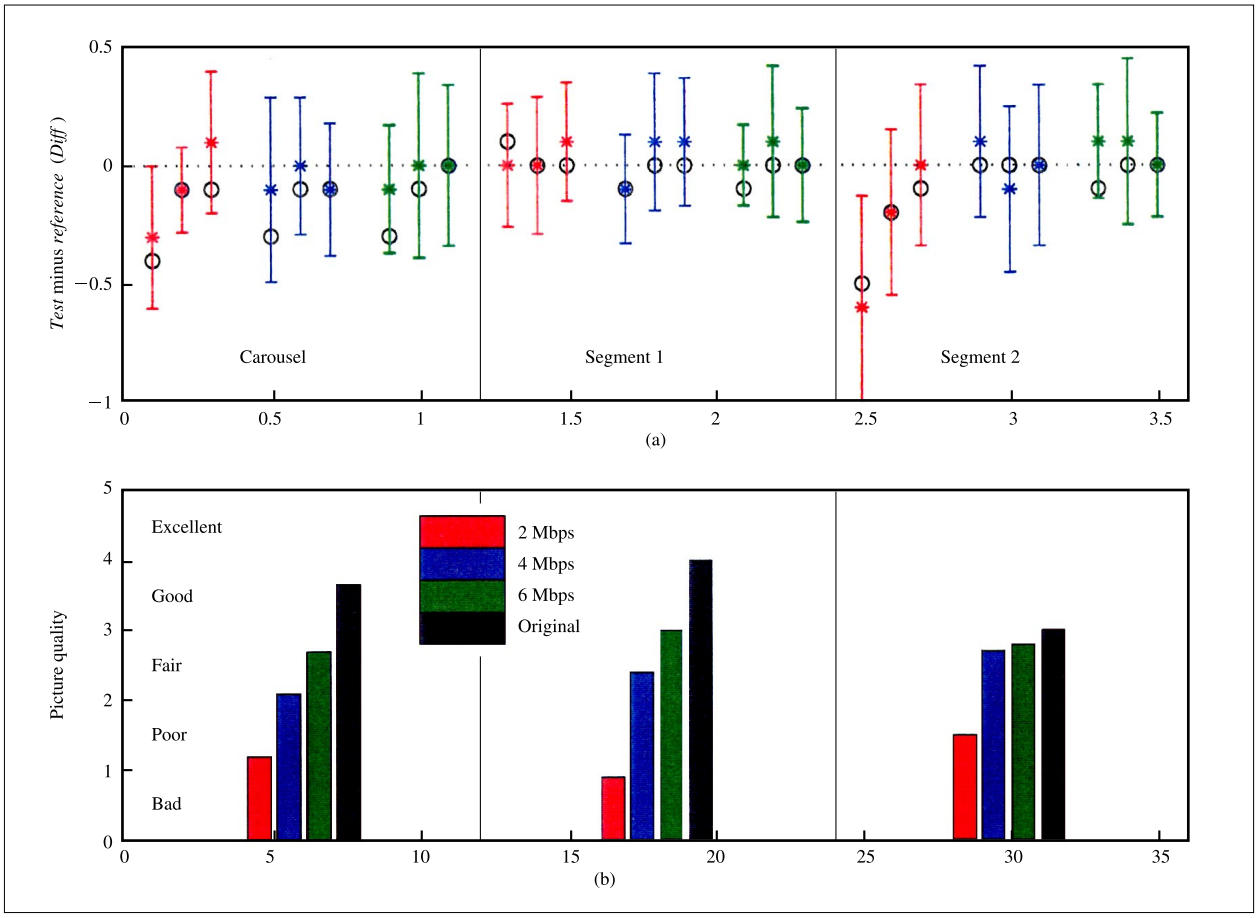

Figure 11

Results of subjective quality experiments: (a) Average of *test minus reference* scores for fourteen observers (*Diff*). For each compressed data rate, three *Diff* scores are shown. They correspond to three search ranges in the following order: 30 × 22, 60 × 42, and 100 × 70. Overall *Diff* averages are shown with asterisks; the average of four selected experts are also shown with circles. (b) Average of absolute subjective quality score.

sequences of limited search range, as compared to their corresponding references. We note that differences in *Diff* values of the order of 0.1 are statistically insignificant. In fact, we believe that *Diff* values in the interval $-0.1 \leq Diff \leq 0.1$ correspond to cases in which test and reference sequences were indistinguishable.

For each video segment, we also analyzed separately the results for a subgroup of four "expert" observers from the pool of fourteen observers. "Experts" were selected from the four largest scores resulting from an algorithm in which, for each test-reference difference (or *Diff* score), we counted the number of negative *Diff* scores and then subtracted the number of positive *Diff* scores (zero differences were excluded from this count). In other words, this algorithm rewards observers who favored the reference sequence over the test sequence, and penalizes those who favored test over reference. The results for these "expert" groups are shown with black circles in Figure 11(a).

The results from subjective experiments tend to confirm some of the trends and observations of the objective PSNR measurements. However, they also demonstrate that measuring differences in PSNR is much easier than "seeing" the effects of those differences in compressed MPEG video. In fact, the quality of sequences with extreme and complex motion, such as "Segment 1" (Basketball), appears to be *unaffected* by limitations in search range! Only absolute data rate appears to have an impact on subjective video quality. Clearly, the human eye is not capable of discerning small improvements in PSNR (due to increased search range,) when the video has much disorganized motion. This conclusion appears to apply to both expert and nonexpert observers.

Humans appear to be a little more sensitive to differences in PSNR when the motion is more organized and predictable. Such is the case for "Carousel," in which statistically significant differences are observed in Figure 11, particularly for low data rates and expert observers. In fact, the effects of limited search range, when observable, are more significant at the lower bit rates. It is important to note, however, that at the 2-Mbps data rate, none of the three sequences we tested are of acceptable quality. As shown in Figure 11(b), all of the sequences were rated as "poor" when coded at this rate. Thus, the fact that observers could see that search-range-limited sequences were somewhat worse than "poor" reference sequences is of doubtful value. Regardless of data rate, however, we could not find one example for which expert assessors could tell the difference between a 100×70 and a 200×140 search range. We thus believe that this search range is a conservative value we can use for robust MPEG-2 encoding of CCIR 601 video.

5. Conclusions

Although our experimental data is strictly relevant only to our particular experimental setup and to the video sequences we tested, we believe that our conclusions have a much wider validity. We have conducted our experiments by using practical algorithms and intentionally looking for a range of demanding sequences that we believe stress the requirements for motion-estimation search range.

On the basis of the experimental data we conclude that for CCIR 601, 4:3 aspect ratio, video, and film, a search range of around 100×70 is sufficient for *robust* motion estimation. In fact, this may well be a conservative search range to use. Even when this search range is clearly exceeded by the 99% statistics of a video segment, we find that objective differences in video quality are nonexistent or insignificant, while subjective differences in video quality are simply not observed. Furthermore, our experiments show that further constraining the search range will have an impact primarily on compression at very low data rates, where the quality of the video is poor, regardless of search range.

When these results are extrapolated for video with the same number of samples per picture, but with an aspect ratio of 16:9 instead of 4:3, the 100×70 search range becomes a more symmetrical 75×70. Further extrapolating from the CCIR 601 16:9 aspect ratio to the ATSC 16:9, 1920×1080 high-definition format, the required search range for robust MPEG-2 encoding becomes 200×158. The latter result takes into account the different pixel shapes between CCIR 601 720×480 and the ATSC 1920×1080 picture format.

Acknowledgments

The authors appreciate the help and encouragement from their colleagues Jack Kouloheris and Wai Man Lam.

**Trademark or registered trademark of Moving Picture Experts Group.

References

1. "Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media up to 1.5 Mbits/s: Video," *ISO/IEC 11172-2*, August 1993.
2. "Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Video," *ISO/IEC 13818-2*, November 1995.
3. "Video Codec for Audiovisual Services at p × 64 kbit/s," *ITU-T Recommendation H.261*, March 1993.
4. "Video Coding for Low Bit Rate Communication," *ITU-T Recommendation H.263*, May 1996.
5. J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall, *MPEG Video Compression Standard*, Chapman & Hall, New York, 1996.
6. B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman & Hall, New York, 1997.
7. A. Puri, R. Aravind, and B. Haskell, "Adaptive Frame/Field Motion Compensated Video Coding," *Signal Process.: Image Commun.* **5,** 39–58 (1993).
8. F. Dufaux and F. Moscheni, "Motion Estimation Techniques for Digital TV: A Review and a New Contribution," *Proc. IEEE* **83,** 858–876 (1995).
9. E. Linzer, "Hierarchical Motion Estimation for MPEG2," *Research Report RC-20364*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, March 1996.
10. E. Linzer and A. Ngai, "Chip Model Encoder," *Research Report RC-20363*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, March 1996.
11. "Coded Representation of Pictures and Audio Information, Test Model 4," *ISO/IEC/JTC1/SC29/WG11*, February 1993.
12. T. Hidaka and K. Ozawa, "Subjective Assessment of Redundancy-Reduced Moving Images for Interactive Application: Test Methodology and Report," *Signal Process.: Image Commun.* **2,** 201–219 (1990).

Received June 10, 1998; accepted for publication June 2, 1999

Cesar A. Gonzales *IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598 (butron@us.ibm.com).* Dr. Gonzales is an IBM Fellow, Senior Manager for Multimedia Technologies at the IBM Thomas J. Watson Research Center, and manager of the development organization of the IBM Digital Video Products Group. He is an expert in image and video processing and compression; his experience spans the development of algorithms, chip and system architectures, and multimedia applications. He is a co-inventor of various still-frame and motion video compression techniques that IBM contributed to the JPEG and MPEG international standards. He is the author of a number of patents and publications related to these techniques. While at IBM, he has received multiple Outstanding Achievement awards, including a Corporate-level award for his contributions to IBM's MPEG products. In June 1998 Dr. Gonzales was named an IBM Fellow. He was also elected to serve in IBM's Academy of Technology, and has twice been named Master Inventor for his contributions to IBM's patent portfolio. Prior to joining IBM, Dr. Gonzales worked on radar signal processing and physics of the ionosphere at the Arecibo Observatory in Puerto Rico. Dr. Gonzales is a Senior Member of the IEEE. He has served as an Associate Editor of the *IEEE Transactions for Circuits and Systems for Video Technology*. He has also served as the head of the U.S. delegation in the MPEG committee of the International Standards Organization. Dr. Gonzales received his B.S. and engineering degrees from the National University of Engineering (UNI) in Peru and his Ph.D. from Cornell University, all in electrical engineering.

Hangu Yeo *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (hangu@us.ibm.com*. Dr. Yeo is a Postdoctoral Fellow in the Multimedia Technologies Department at the IBM Thomas J. Watson Research Center. He received his B.S. degree in electronic engineering from Yonsei University, Seoul, Korea, his M.S. degree in electrical engineering from Columbia University, and his Ph.D. degree in electrical and computer engineering from the University of Wisconsin, Madison. From 1990 to 1992, he worked as a computer assistant at the Ministry of National Defense in Seoul, Korea. His current research interests include video/image compression algorithms, design of VLSI architectures for video and image processing, and advanced set-top-box application development.

Chung J. Kuo *Graduate Institute of Communication Engineering, National Chung Cheng University, Chaiyi, Taiwan 62107 (kuo@ee.ccu.edu.tw)*. Dr. Kuo received the B.S. and M.S. degrees in power mechanical engineering from National Tsing Hua University, Taiwan, in 1982 and 1984, respectively, and the Ph.D. degree in electrical engineering from Michigan State University (MSU) in 1990. He joined the Electrical Engineering Department of National Chung Cheng University (NCCU) in 1990 as an associate professor, becoming a full professor in 1996. He is now the chairman of the Graduate Institute of Communication Engineering of NCCU. Dr. Kuo was a visiting scientist at the Opto-Electronics & System Laboratory, Industrial Technology Research Institute, in 1991 and at the IBM Thomas J. Watson Research Center from 1997 to 1998. He has been a consultant to several international and local companies, and is also an adjunct professor at National Cheng Kung University. Dr. Kuo has interests in image/video signal processing, VLSI signal processing, and optical information processing/computing. He is the Director of the Signal and Media (SAM) Laboratories at NCCU. Dr. Kuo received the Outstanding Research Award from National Chung Cheng University in 1998, the Overseas Research Fellowship from the National Science Council (NSC) in 1997, the Outstanding Research Award from the College of Engineering, NCCU, in 1997, the Medal of Honor from NCCU in 1995, the Research Award from NSC every year since 1991, the Best Engineering Paper Award from Taiwan's Computer Society in 1991, an Electrical Engineering Fellowship from MSU in 1989, and the Outstanding Academic Achievement Award from MSU in 1987. He was a guest editor for two special sections of *Optical Engineering and 3D Holographic Imaging* (to be published by John Wiley and Sons) and has been an invited speaker and program committee chairman or member for several international/local conferences. He also serves as an Associate Editor of *IEEE Signal Processing Magazine* and as President of the Taiwan chapter of SPIE. Dr. Kuo is a member of Phi Kappa Phi, Phi Beta Delta, IEEE, OSA, and SPIE; he is listed in *Who's Who in the World*.