
Peer reviewed version

Link to published version (if available): 10.1109/ICIP.2000.899608

Link to publication record in Explore Bristol Research
PDF-document
STATISTICAL FEATURE EXTRACTION FROM COMPRESSED VIDEO SEQUENCES

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull
Image Communications Group, Centre for Communications Research,
University of Bristol, Merchant Ventures Building, Woodland Road, Bristol BS8 1UB,
United Kingdom.
Voice - +44-117-954-5198, Fax - +44-117-954-5206
Email - W.A.C.Fernando@bristol.ac.uk

ABSTRACT
To maximise the benefits from data compression, it would be advantageous to develop algorithms that do not require decompression to extract relevant information for post processing. In this paper, a novel technique for extracting variance is proposed for MPEG-2 compressed video using Parseval’s theorem. Results show that the estimated variance closely matches with the actual variance. Furthermore, this technique is applied to identify scene changes in compressed domain.

1. INTRODUCTION
A major feature required in a visual information system is an efficient indexing scheme to enable fast access to the stored data. A common and natural idea is to index the video sequences first into video shots by identifying both sudden and gradual scene changes and then to extract features.

The large channel bandwidth and memory requirements for the transmission and storage of image and video necessitate the use of compression techniques [1,2]. Hence, the visual data in future multimedia databases is expected to be stored in the compressed form. In almost all multimedia systems, there is a need to integrate compression and processing of images and video. To maximise the benefits from data compression without incurring extra computation and storage, it would be advantageous to develop processing algorithms that do not require decompression of the entire compressed data. Therefore a powerful scene change detection, which can operate both in compressed and uncompressed domain, is required to allow for a complete characterisation of the video sequences.

Statistical features have been used to identify both sudden and gradual scene changes in uncompressed video sequences [3-7]. Most of these algorithms, which are proposed for uncompressed video, can be applied for compressed video, if we can extract the relevant statistical features. Yeo [7] proposed DC-sequences to extract mean from each image in compressed video. However, variance calculation in compressed domain is difficult.

In this paper we present a novel scheme based on Parseval’s theorem to extract variance from compressed video. A reduced image is formed with extracted statistical features and defined this image as the statistical image. These statistical images can then be used for higher level processing such as object tracking and scene change detection in video sequences. Rest of the paper is organised as follows: Section 2 presents how the variance is evaluated from compressed domain features. Experimental results are given in section 3. Finally, section 4 presents the conclusions and the future work.

2. ESTIMATION OF VARIANCE IN COMPRESSED DOMAIN
In general, a macroblock (MB) of size $N_1 \times N_2$ is considered. Let, $S(k,i,j)$ and $F(k,i,j)$ be the uncompressed image and compressed image respectively. Here, $k$ is the MB number ($k = 1:N$, $N$ is the total number of MBs in an image) Let, $\mu_{k,s}$ and $\sigma_{k,s}^2$ be mean and the variance of the $k^{th}$ MB in uncompressed image respectively. Then, mean and variance of the $k^{th}$ MB can be evaluated as shown in Equation (1) and Equation (2).

$$\mu_{k,s} = \left(\frac{1}{N_1N_2}\right) \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} S(k,i,j)$$

$$\sigma_{k,s}^2 = \left(\frac{1}{N_1N_2}\right) \left[ \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} S(k,i,j)^2 \right] - \mu_{k,s}^2$$

Two-dimensional DCT of a $N_1 \times N_2$ MB can be defined as follows,

$$F(k,p,l) = \frac{4}{N_1N_2} C(p) C(l) \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} S(k,i,j) \cos\left(\frac{\pi i p}{N_1}\right) \cos\left(\frac{\pi j l}{N_2}\right)$$

where, $C(p), C(l) = \frac{1}{\sqrt{2}}$ for $p, l = 0 \quad \text{and} \quad 1 \quad \text{otherwise}$

Using Parseval’s theorem, power in uncompressed domain signal is related to compressed domain signal as shown in Equation (4) when the quantisation noise is negligible.

$$\sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} S(k,i,j)^2 = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} F(k,i,j)^2$$
Mean of the $k^{th}$ MB can be calculated from Equation (3) as follows.

$$\mu_{k} = \frac{1}{N_{1}N_{2}} \sum_{k=0}^{N_{1}-1} \sum_{j=0}^{N_{2}-1} F(k,0,0)$$

Therefore Equations (2), (4) and (5) gives,

$$\sigma_{k}^{2} = \frac{1}{N_{1}N_{2}} \left[ \sum_{k=0}^{N_{1}-1} \sum_{j=0}^{N_{2}-1} F(k,i,j)^{2} - \mu_{k}^{2} \right]$$

Equation (6) shows that $\sigma_{k}^{2}$ can be calculated by taking the average power of AC-coefficients in each MB. Therefore, mean and variance of each MB can be evaluated in DCT compressed domain using equation (5) and (6) respectively. It can also be shown that mean ($\mu_{k}$) and variance ($\sigma_{k}^{2}$) of the whole image, which contains $N$ number of MBs, can be calculated from Equation (7) and (8) as given below.

$$\mu_{k} = \frac{1}{N} \sum_{k=0}^{N_{1}-1} \sum_{j=0}^{N_{2}-1} F(k,0,0)$$

$$\sigma_{k}^{2} = \frac{1}{NN_{1}N_{2}} \left[ \sum_{k=0}^{N_{1}-1} \sum_{j=0}^{N_{2}-1} F(k,i,j)^{2} - \frac{1}{N} \sum_{k=0}^{N_{1}-1} \sum_{j=0}^{N_{2}-1} F(k,0,0)^{2} \right]$$

MPEG-2 defines three types of pictures: intra (I)-Pictures, Predictive pictures (P-Pictures) and Bi-directional pictures (B-Pictures). The extraction of statistical features from each MB in intra (I) frames is explained before. However, feature extraction for inter-coded frames (P and B) is not trivial since DCT coefficients for these MBs are not readily available. Therefore, DCT coefficients of each inter-coded MBs need to be evaluated.

### 2.1 DCT Coefficients estimation

**Figure 1**: Graphical representation of four neighbouring MBs and motion vectors

<table>
<thead>
<tr>
<th>Sub-block</th>
<th>Position</th>
<th>$S_{11}$</th>
<th>$S_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MB_{1}$</td>
<td>lower right</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$MB_{2}$</td>
<td>lower left</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$MB_{3}$</td>
<td>upper right</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$MB_{4}$</td>
<td>upper left</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: $S_{11}$ and $S_{12}$ matrices ($I$ is an identity matrix of $h_{i}$ or $w_{j}$)

When DCT coefficients are extracted both mean and variance can be evaluated with Equations (7) and (8) for inter-coded frames.

Hence S-sequences can be defined for both intra and motion compensated frames from Equation (11) and (12).

$$S_{\mu,
u,n} = \frac{1}{k_{1}k_{2}\sqrt{N_{1}N_{2}}} \sum_{k=0}^{k_{1}} \sum_{l=0}^{k_{2}} F(k,l,n)$$

$$S_{\mu,
u,n} = \frac{1}{k_{1}k_{2}\sqrt{N_{1}N_{2}}} \sum_{k=0}^{k_{1}} \sum_{l=0}^{k_{2}} F(k,l,n)$$

(9)

(10)

(11)

(12)
where, \( k_1 \) and \( k_2 \) are defined through \( M' = k_1 N_i \) and \( N' = k_2 N_j \), respectively. ( \( N_i = N_j = 8 \) for MPEG-2).

These S-sequences can be used for higher level processing in image or video.

### 3. RESULTS

In this section, some results are presented with the proposed scheme. We used several test sequences and compare the estimated features against the actual values. Two QCIF test sequences: akiyo and foreman are considered here. First sequence doesn’t contain any significant variations and second sequence exhibit large scene variations. The compressed version of these sequences is considered and both mean and standard deviation (variance) is extracted using the proposed scheme.

**Figure 2:** Comparison between the actual and the estimated mean (akiyo)

**Figure 3:** Comparison between the actual and the estimated mean (foreman)

**Figure 4:** Comparison between the actual and the estimated standard deviation (akiyo)

**Figure 5:** Comparison between the actual and the estimated standard deviation (foreman)

**Figure 6:** Comparison between the actual variance and the estimated variance for the fade-in sequence

**Figure 7:** Comparison between the actual mean and the estimated mean for the fade-in sequence
Figure 2 and Figure 3 present the actual and extracted mean for akiyo and forman video sequences. Actual and extracted standard deviation are illustrated in Figure 4 and Figure 5 for the same sequences. These illustrations clearly show that extracted statistical features closely follows the actual values. Finally, this variance extraction technique is applied to identify fading and dissolving in MPEG-2 compressed domain with one of our algorithm proposed for uncompressed video [3]. In this algorithm [3], ratio between the second derivative of the variance curve and the first derivative of the mean curve is considered. The differentiation of this ratio is checked to identify fading and dissolving with a threshold T_{th}. Figure 6 shows the behaviour of actual and extracted variance for a fade-in sequence. These clearly illustrate that extracted variance closely follows the actual variance and hence this can be used to identify fading/dissolving in compressed domain. Similar observations can also be seen in Figure 7 for mean.

Table 2 presents some of the results for fading/dissolving scene change detection with the proposed feature extraction technique along with the algorithms discussed in [3]. Results show that detected transition regions in compressed domain are equal or very close to the detected regions in uncompressed domain.

<table>
<thead>
<tr>
<th>Actual transition region</th>
<th>Detected transition region (Uncompressed)</th>
<th>Detected transition region (Compressed)</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>446-496</td>
<td>446-497</td>
<td>446-497</td>
<td>dissolve</td>
</tr>
<tr>
<td>754-778</td>
<td>754-778</td>
<td>754-778</td>
<td>fade-out</td>
</tr>
<tr>
<td>804-868</td>
<td>804-868</td>
<td>804-868</td>
<td>dissolve</td>
</tr>
<tr>
<td>1010-1089</td>
<td>1012-1089</td>
<td>1012-1089</td>
<td>dissolve</td>
</tr>
<tr>
<td>1166-1232</td>
<td>1169-1233</td>
<td>1169-1233</td>
<td>dissolve</td>
</tr>
<tr>
<td>1356-1424</td>
<td>1356-1424</td>
<td>1356-1424</td>
<td>dissolve</td>
</tr>
<tr>
<td>1500-1550</td>
<td>1500-1552</td>
<td>1500-1552</td>
<td>fade-in</td>
</tr>
<tr>
<td>1620-1680</td>
<td>1620-1681</td>
<td>1620-1681</td>
<td>fade-in</td>
</tr>
<tr>
<td>1760-1840</td>
<td>1762-1840</td>
<td>1762-1840</td>
<td>fade-out</td>
</tr>
</tbody>
</table>

Table 2: Dissolve/fade region identification

Furthermore, we used this variance extraction scheme to identify wiping in MPEG-2 compressed video and observed these scene changes were also identified accurately. Therefore, this feature extraction technique can be used effectively in compressed video to identify scene changes.

4. CONCLUSIONS

In this paper we presented a novel scheme to estimate variance of the images in MPEG-2 compressed video directly without full frame decompression. Results show that the estimated variance closely matches the actual variance. Furthermore, the proposed feature extraction technique was used to identify scene changes in compressed video without full frame decompression.

ACKNOWLEDGEMENTS

First author would like to express his gratitude and sincere appreciation to the university of Bristol and CVCP for providing financial support for this work.

REFERENCES