Available online at www.sciencedirect.com
SCIENCE DIRECT®

J. Vis. Commun. Image R. 15 (2004) 132–144

JOURNAL OF VISUAL Communication & IMAGE Representation

www.elsevier.com/locate/jvci

A new cut detection algorithm with constant false-alarm ratio for video segmentation

Tie-Yan Liu,[a] Kwok-Tung Lo,[b,*] Xu-Dong Zhang,[a] and Jian Feng[c]

[a] *Department of Electronic Engineering, Tsinghua University, Beijing, PR China*
[b] *Department of Electronic and Information Engineering, Centre for Multimedia Signal Processing, The Hong Kong Polytechnic University, Hong Kong, Hung Hom, Hong Kong*
[c] *Department of Computer Engineering and Information Technology, City University of Hong Kong, Hong Kong*

Received 9 May 2002; accepted 30 October 2003

Abstract

In this paper, a new cut detection algorithm with constant false-alarm ratio (CFAR) is proposed for video segmentation. In our method, a theoretical threshold determination strategy using the non-parameter based CFAR processing technique is developed to achieve a controllable precision as well as an evaluative recall performance for video cut detection. Simulation results show that this algorithm leads to very good detection performance as compared to the existing techniques.

Keywords: Cut detection; Video retrieval; Constant false-alarm ratio; Shot-boundary detection; Video segmentation

1. Introduction

With the increasing demand for the storage of a variety of visual information, the topic of content-based video retrieval (CBVR) has been actively investigated by a

* Corresponding author. Fax: +852-23628439.
E-mail addresses: t-tyliu@microsoft.com (T.-Y. Liu), enktlo@polyu.edu.hk (K.-T. Lo), zhangxd@tsinghua.edu.cn (X.-D. Zhang), itjfeng@cityu.edu.hk (J. Feng).

1047-3203/$ - see front matter
doi:10.1016/j.jvcir.2003.10.001

number of researchers from different communities in recent few years. In a CBVR system, the temporal feature of a video is first extracted by segmenting the video into successive shots in the temporal domain. Other features such as key objects and key frames are then extracted to abstract the video. Obviously, the video shot boundary detection process is the first but important step in such a system. Only after segmenting a video sequence into shots, can the key frame extraction and index generation steps follow sequentially.

In the literature, people have developed different kinds of shot boundary detection algorithms, such as histogram-based algorithms (Zhang et al., 1993; Sethi, 1995), motion-based algorithms (Zhang et al., 1995; Shahraray, 1995) and contour-based algorithms (Zabih et al., 1995) for cut detection; twin-comparison algorithm (Zhang et al., 1993) and production-model based algorithms (Hampapur et al., 1995) for gradual transition detection. The criteria of *recall* and *precision* as shown below are commonly used in evaluating the performance of these detection algorithms.

$$recall = N_{\text{detect}}/(N_{\text{detect}} + N_{\text{miss-detect}}), \tag{1a}$$

$$precision = N_{\text{detect}}/(N_{\text{detect}} + N_{\text{false-alarm}}), \tag{1b}$$

where N_{detect} is the number of shot transitions detected, $N_{\text{miss-detect}}$ is the number of mis-detections and $N_{\text{false-alarm}}$ represents the number of false alarms.

It is known that a high false-alarm ratio (FAR) in video shot boundary detection results in fragmentation of a video sequence, which will destroy the integrity of video contexts and hence damage to the structure of the video database. Although FAR is quite important, the shot boundary detection algorithms developed so far seldom theoretically guarantee a pre-determined FAR. In this paper, we try to design an algorithm to tackle this problem, or say, to guarantee a controllable detection performance. For simplicity and clarity, in the following discussions, we concentrate only on the control of FAR for cut detection. To control that of gradual transition detection will be a little more difficult but theoretically similar. In order to control the FAR, the concept of constant false-alarm ratio (CFAR) processing in radar signal detection (Minkler and Minkler, 1990) is applied. In fact, both the problem statement and physical variables of video cut detection are analogous to those of radar signal detection. The detailed comparison is shown in Table 1. We can see that both video cut detection and radar signal detection possess the same objective, which is to improve the recall performance as much as possible given a controllable precision.

Table 1
Comparison of radar and video CFAR

	Cut detection	Radar signal detection
Problem statement	Improve the recall performance as much as possible provided that the precision performance is appropriately controlled.	
Physical variables	Cuts	Objects
	Regular frames	Spatial noises
	Camera motion	Clutter
	Background luminance change	Interference

As we know, the radar CFAR algorithms can be classified into two categories, the parameter based and non-parameter based algorithms. The former assumes that the clutter has a certain distribution, such as Gaussian or Rayleigh distribution. However, for video signal, the statistical characteristics are so complicated that such simple approximations will lead to poor performance due to model mismatch. Therefore, the work described in this paper will focus on the non-parameter based CFAR techniques (Carlyle and Thomas, 1964), which do not rely on any prior distributions.

In the rest of the paper, the details of the proposed CFAR algorithm for video cut detection are described in Section 2. This section will expatiate on the concept of the algorithm to show why it leads to a CFAR. Also, the recall performance of the algorithm is deduced. Computer simulations are used to test the proposed algorithm and the results are discussed in Section 3. Finally, some concluding remarks are given in Section 4.

2. CFAR algorithm for video cut detection

In this section, a new video cut detection algorithm that guarantees a controllable FAR is described. The magnitude of successive frame difference is no longer the key criterion in this algorithm. Instead, we detect cuts by evaluating how many previous frame differences are less than the current one. As shown in Fig. 1, the video frames under consideration are classified into three categories: *current frame* (denoted by F_C), *protective frames* (by F_P) and *reference frames* (by F_i, $i = 1 \cdots N$). The *current frame* is the frame to be detected. We have two basic hypotheses on the current frame, H_0: it is a regular frame; H_1: it is a cut. For the *reference frames*, we can assume that all of them are regular frames, if the previous cut detection result is convinced. (In fact, experiments show that just following this assumption without considering whether it is fully satisfied or not, the performance of our algorithm still corresponds to the theoretical deductions well.) When the *current frame* is not a cut, it will have the same statistical characteristic as the references; otherwise, it will be quite distinguished. Considering in some cases, cut may occur in more than one frame due to frame rate conversion, we use two *protective frames* to separate the *current frame* out of the references.

In order to carry out the following CFAR operations, it is necessary to assume that all regular frames are independent and identically distributed. The identical

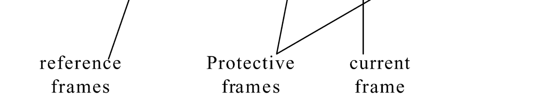

Fig. 1. Video frames in a video sequence.

condition is easily understood, however, the independency is not true in all cases because the luminance fields of successive frames are highly correlated. In order to decorrelate the frames, a differential feature should be defined. In this paper, the *visual discontinuity* (Hanjalic and Zhang, 1999) is selected as the desired feature. For each video frame, the *visual discontinuity* is defined as below.

$$x = \frac{1}{N_{\mathrm{B}}} \sum_{l=1}^{N_{\mathrm{B}}} \left(\left| \overline{Y}_l - \overline{Y}_{\mathrm{match}} \right| + \left| \overline{U}_l - \overline{U}_{\mathrm{match}} \right| + \left| \overline{V}_l - \overline{V}_{\mathrm{match}} \right| \right), \tag{2}$$

where l is the block index, *match* refers to the best matched block in the previous frame during motion estimation. N_{B} is the total number of the blocks. $\overline{Y}$, $\overline{U}$, and $\overline{V}$ are the average luminance and chrominance of an image block, respectively. Here, to be noted, such a definition of *visual discontinuity* is only an example, any measure that can guarantee the independent and identical distribution does not affect the following deductions.

For the current frame, the distribution of the *visual discontinuity* is denoted by $p_1(x_{\mathrm{c}})$, and those of the references are denoted by $p_0(x_i)$, $(i = 1 \ldots N)$, I.I.D. Considering N references frames are used, one may argue that in some cases, the length of the shot containing the current frame may be shorter than N so that we cannot gather enough references for detecting the current frame. Fortunately, with the assumption of independent and identical distribution, there is no need to guarantee that all the N reference frames come from the same shot. In fact, in the experiments we designed a FIFO structure which only buffers regular frames to help provide enough reference frames.

To obtain a constant false-alarm ratio, we proceed as follows. First, to represent the difference between the *current frame* and the references, just as in the Wilcoxon signal detector (Minkler and Minkler, 1990), the rank of the *current frame* is defined as

$$C(x_{\mathrm{c}}, x_i) = \begin{cases} 1, & x_{\mathrm{c}} > x_i, \\ 0, & x_{\mathrm{c}} < x_i. \end{cases} \tag{3}$$

It is obvious that the sum of the ranks $z = \sum_{i=1}^{N} C(x_{\mathrm{c}}, x_i)$ is a random variable distributed over the interval [0, N] and has a distribution function as below.

$$\begin{aligned} P(z = L) &= C_N^L \int_{-\infty}^{+\infty} p_1(x_{\mathrm{c}}) \left[\int_{-\infty}^{x_{\mathrm{c}}} p_0(x_i)\,\mathrm{d}x_i \right]^L \left[\int_{x_{\mathrm{c}}}^{+\infty} p_0(x_i)\,\mathrm{d}x_i \right]^{N-L} \mathrm{d}x_{\mathrm{c}} \\ &= C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \int_{-\infty}^{+\infty} p_1(x_{\mathrm{c}}) R^{N-L+j}\,\mathrm{d}x_{\mathrm{c}}, \end{aligned} \tag{4}$$

where $R = \int_{x_{\mathrm{c}}}^{\infty} p_0(x_i)\,\mathrm{d}x_i$ (see Appendix A).

When H_0 holds, the *current frame* has the same statistical distribution as the references. Thus, it can be proved that (see Appendix B) z is uniformly distributed, i.e.,

$$P(z = L|H_0) = \frac{1}{N+1}. \tag{5}$$

If setting the detection threshold to K, (that is, when z is not less than K, a cut frame is determined; otherwise no cut is detected), we can calculate the FAR P_n as follows.

$$P_n = \sum_{L=K}^{N} P(z = L|H_0) = \sum_{L=K}^{N} \frac{1}{N+1} = \frac{N+1-K}{N+1}. \tag{6}$$

Based on the above deduction, we can conclude that the FAR P_n is independent of the video sequences' distributions. As a result, by choosing N and K carefully as indicated in (7), the FAR and thus the precision performance of the algorithm can be controlled.

$$K = (N+1)(1-P_n). \tag{7}$$

In fact, the idea resident in the proposed algorithm is to replace the Bayesian criterion that to what degree the feature of the *current frame* is larger than those of the references, by a new criterion that the feature of the *current frame* is larger than those of how many *reference frames*. Such a replacement provides some new benefits. The most important one is that now we are able to evaluate and thus control the precision performance of the detection algorithm at the designing stage.

After deriving the false-alarm ratio of the proposed algorithm, we will also evaluate the miss-detection ratio. For this purpose, we have to know the prior distribution of the visual feature used. In this paper, the empirical distribution of the *visual discontinuity* formulated in (Hanjalic and Zhang, 1999) for both cut and regular frames are used.

For regular frames,

$$p_0(x_i) = \frac{4}{3} x_i^4 \mathrm{e}^{-2x_i}. \tag{8}$$

For cut frames,

$$p_1(x_c|H_1) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x_c-\mu)^2/2\sigma^2}. \tag{9}$$

Then, we have

$$R = \int_{x_c}^{+\infty} p_0(x_i)\,\mathrm{d}x_i = \frac{\mathrm{e}^{-2x_c}}{24}(16x_c^4 + 32x_c^3 + 48x_c^2 + 96x_c + 24). \tag{10}$$

In the following deductions, we will simplify the expression of R by only maintaining the first two items:

$$R = \frac{\mathrm{e}^{-2x_c}}{24}(16x_c^4 + 32x_c^3). \tag{11}$$

When N is not a very large number, for $\mu = 42$, $\sigma = 10$ (the parameters used in Hanjalic and Zhang, 1999), it can be proved that the approximation is reasonable. Substitute Eq. (11) into (4), we have (see Appendix C)

$$p(z = L|H_1) = \frac{1}{2\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \times \sum_{i=0}^{t} C_t^i (-a)^i \left(\sqrt{2}\sigma\right)^{t-i+1} \Gamma\left(\frac{t-i+1}{2}, \frac{a^2}{2\sigma^2}\right), \tag{12}$$

where $t = 3(N - L + j) + m$, $a = 2\sigma^2(N - L + j) - \mu$.

Then the miss-detection ratio (MDR) P_m is computed as follows:

$$P_m = \sum_{L=0}^{K-1} p(z = L|H_1). \tag{13}$$

Some typical values of the MDR P_m calculated from Eq. (13) are shown in Table 2. The probability density function formulated in (Hanjalic and Zhang, 1999) is used in Table 2a. Furthermore, in order to verify the formula deduced, we use the true statistics of a video sequence *News1* in Tables 2b and c. As shown, P_m decreases monotonously with increasing P_n. And when P_n is fixed, P_m increases with increasing N approximately.

As shown above, we really achieve a constant false-alarm ratio and a calculable miss-detection ratio. When mapping these two ratios to precision and recall, some new problems arise. From (6), with a reasonable assumption that the average shot length is several hundreds of frames, in order to get a precision of about 90%, we

Table 2
Typical MDR values

Number of reference frames N	Threshold K	Miss detection ratio P_m
(a) FAR $P_n = 10\%$, $\mu = 42$, $\sigma = 10$		
10	9	8.1907e − 005
15	14	9.4208e − 005
20	18	6.8401e − 005
25	23	7.2907e − 005
30	27	1.6143e − 004
(b) FAR $P_n = 10\%$, $\mu = 76.8, \sigma = 29.3$		
10	9	0.0024
15	14	0.0026
20	18	0.0023
25	23	0.0024
30	27	0.0106
P_n	Threshold K	Miss detection ratio P_m
(c) $N = 20$, $\mu = 76.8, \sigma = 29.3$		
5%	19	0.0028
10%	18	0.0023
20%	16	0.0018
30%	14	0.0016
40%	12	0.0015

need a false-alarm ratio of less than 0.1%. That is to say, we should have an N no less than 1000. Such a large number will be a challenge for the simplification from (10) to (11). In order to tackle this problem, we modified the definition of the current frame's rank a little as below.

$$C(x, x_i) = \begin{cases} 1, & x > x_i + \delta, \\ 0, & x < x_i + \delta, \end{cases} \tag{14}$$

where δ is an experimental constant. The main idea here is to reduce the false alarm ratio by introducing a more restricted condition for cut detection, so that we can achieve an acceptable precision performance with a relatively smaller N. The detailed selection of δ will be finally determined through experiments in Section 3.

3. Simulation results

In order to test the performance of the proposed cut detection algorithm, we carried out a series of experiments as follows. In our experiments, Windows 2000 running on a 2-GHz Pentium-IV PC is used for the simulation platform. Three CIF-format video sequences (available at ftp://msplab.ee.tsinghua.edu.cn/pub/VideoSequence/), with nearly 100,000 frames, or 100 min long, are used:

Shakespeare in Love: 11,837 frames and 121 cuts.

The Others: 27,599 frames, 169 cuts and five gradual transitions.

News1: 57,400 frames, 303 cuts and 34 gradual transitions.

The first two sequences are extracted from video CDs, which contain medium motions and the luminance level all over the sequences is relatively low; the third one is a standard MPEG-7 test sequence which contains various local and global motions and many gradual transitions. The ground truths of cut occurrences in *The Others* and *Shakespeare in Love* are labeled by human beings, while that of *News1* is provided with the MPEG-7 CD. The statistical characteristics of these three sequences are calculated and listed in Table 3.

At first, the sensitivities of some experimental parameters to the detection performance of the proposed algorithm are investigated. Fig. 2 shows the detection performance for sequence *News1* with different parameter settings.

From Fig. 2, we firstly see that introducing the CFAR operation can greatly increase the precision performance of cut detection, no matter what N, K and δ are. Especially, for the parameters N and K used (which correspond to a false-alarm ratio

Table 3
Statistical characteristics of *visual discontinuity* of cut frames

Sequence	Mean value μ	Standard deviation, σ
Shakespeare in Love	33.1	20.5
The Others	31.7	16.4
News1	76.8	29.3
Model used in Hanjalic and Zhang (1999)	42	10

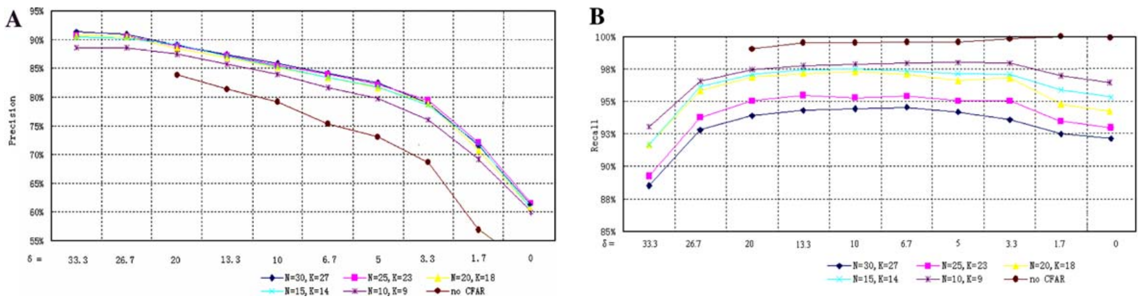

Fig. 2. Detection performance for sequence *News1* with difference parameters. (A) Precision performance. (B) Recall performance.

of 10%), we can get more detailed discussions as below. If following the assumption that each shot contains approximately several hundreds of frames, such a false-alarm ratio will lead to a very poor precision. In fact, when using (3) to compute the rank (see the points corresponding to $\delta = 0$), the experimental results coincide with this analysis. On the other hand, we find that by replacing (3) with (14), the precision performance is much more increased with increasing δ. For example, when $\delta = 26.7$ for N of only 30 or less, we have achieved a precision of about 90%. This is very meaningful for practical implementation, because we cannot always accumulate the information of a very large number of reference frames.

Among all these results shown in Fig. 2, we find that those when $N = 15$ and $K = 14$ are relatively insensitive to the value of δ. And when $\delta = 26.7$ it is the best compromise between precision and recall. As a result, these parameter settings are used for a typical performance to be compared with other cut detection algorithms.

In the second experiment, we compare the performance of our CFAR algorithm (denoted as C-1) with some previously proposed cut detection algorithms, which include the global luminance histogram method (denoted as H-1) (Zhang et al., 1993),

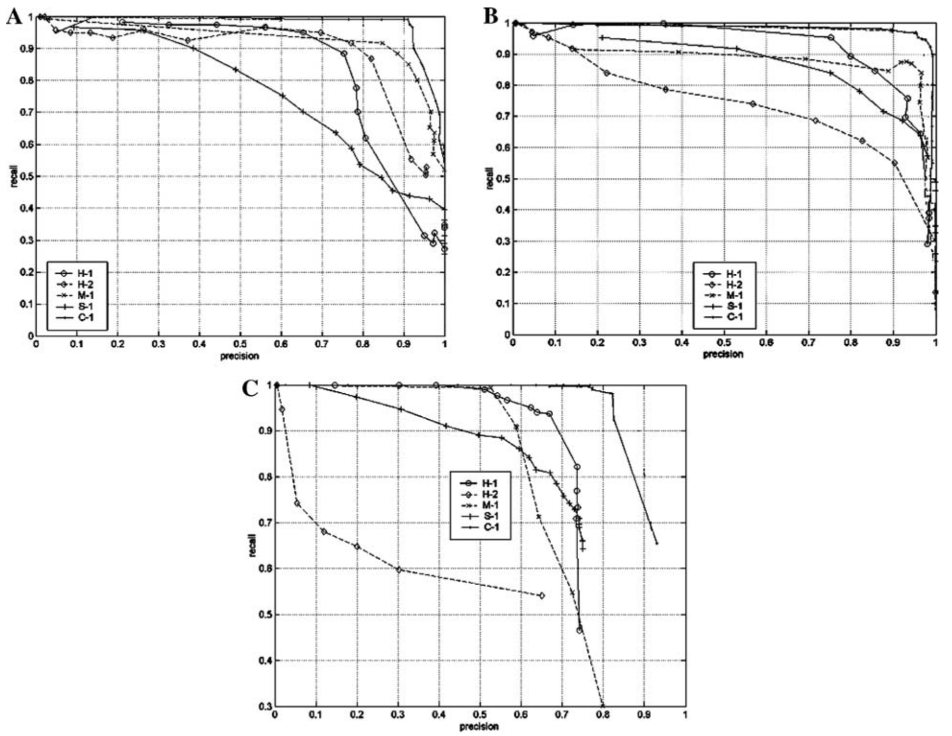

Fig. 3. Comparison of the proposed algorithm with several reference algorithms. (A) *Shakespeare in Love*. (B) *The Others*. (C) *News1*.

the global hue histogram method (Lupatini et al., 1998) (denoted as H-2), the motion compensation based algorithm (Shahraray, 1995) (denoted as M-1), and the central moment based algorithm (Fernando et al., 2000) (denoted as S-1). The operating curves for different test sequences are shown in Fig. 3.

The results listed above show that the proposed algorithm is quite good as compared to the previous works, especially for the precision performance. We can take the sequence *News1* as an example. For this sequence where there are quite a number of gradual transitions, we can see that our method results in much less number of false alarms than others. In other words, our method leads to over 10% higher precision when the recall performance is at the same level as other methods. This is just what we expect at the design stage. In fact, from the results of the other two sequences, we can also come to the similar conclusion.

Finally, we will count on the computational complexity and memory occupation of the proposed algorithm. As seen in Section 2, the *visual discontinuity* is calculated for each frame. For this purpose, block-based motion estimation should be carried out for uncompressed video sequences. That can be accomplished in real time. (For compressed sequences, the motion vectors are already available in the bit streams so that these computations can be saved.) The calculations of the sum of ranks are just some comparisons and additions, which can be neglected as compared to motion estimation. In the proposed algorithm, we make use of the information in the previous N frames to guarantee a CFAR, only one float need allocating to store the value of *visual discontinuity* for each frame. Thus the total memory usage is acceptable.

Through the above experiments and discussions, we can come to the conclusion that the proposed CFAR technique can successfully restrict the precision performance in a given range by selecting a prior threshold, and the recall performance of the algorithm is also very good as compared to the previous detection algorithms.

4. Conclusions

In this paper, a novel video cut detection algorithm with constant false-alarm ratio is proposed. Utilizing the CFAR concept from radar signal detection, the proposed algorithm leads to a very attractive performance in sense of a controllable false alarm ratio and high detection accuracy. Simulation results show that this algorithm is very effective as compared to the previous works.

Acknowledgments

The work of Kwok-Tung Lo was supported by the Centre for Multimedia Signal Processing, The Hong Kong Polytechnic University. The work of Jian Feng was supported by the City University of Hong Kong under Grant A/C 7100250.

Appendix A

$$R = \int_{x_c}^{\infty} p_0(x_i)\,\mathrm{d}x_i \;\therefore\; 1 - R = \int_{-\infty}^{x_c} p_0(x_i)\,\mathrm{d}x_i,$$

$$P(z = L) = C_N^L \int_{-\infty}^{+\infty} p_1(x_c)(1-R)^L R^{N-L}\,\mathrm{d}x_c,$$

$$\because R < 1, (1-R)^L = \sum_{j=0}^{L} (-1)^j C_L^j R^j,$$

$$P(z = L) = C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \int_{-\infty}^{+\infty} p_1(x_c) R^{N-L+j}\,\mathrm{d}x_c.$$

Appendix B

$$\because p_1(x_c|H_0) = p_0(x_c),$$

$$\begin{aligned} P(z = L|H_0) &= C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \int_{-\infty}^{+\infty} R^{N-L+j} p_0(x_c)\,\mathrm{d}x_c \\ &= C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \int_{-\infty}^{+\infty} R^{N-L+j} p_0(x_i)\,\mathrm{d}x_i. \end{aligned}$$

$$\because \mathrm{d}R = -p_0(x_i)\,\mathrm{d}x_i,$$

$$\begin{aligned} P(z = L|H_0) &= C_N^L \sum_{j=0}^{L} (-1)^{j+1} C_L^j \int_{1}^{0} R^{N-L+j}\,\mathrm{d}R \\ &= C_N^L \sum_{j=0}^{L} (-1)^{j+1} C_L^j \frac{1}{N-l+j+1} R^{N-L+j+1} \Big|_1^0 \\ &= C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \frac{1}{N-L+j+1}. \end{aligned}$$

Considering that

$$\sum_{j=0}^{L} \frac{(-1)^j}{a+j} C_L^j = \frac{L!}{a(a+1)\cdots(a+L)},$$

we have

$$P(z = L|H_0) = \frac{1}{N+1}.$$

Appendix C

Let $a = 2\sigma^2(N-L+j) - \mu$,

$$p(z=L|H_1) = C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x_\mathrm{c}-\mu)^2/2\sigma^2} \mathrm{e}^{-2(N-L+j)x_\mathrm{c}} \times (x_\mathrm{c}^4 + 2x_\mathrm{c}^3)^{N-L+j}\,\mathrm{d}x_\mathrm{c},$$

$$= C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \mathrm{e}^{-(x_\mathrm{c}+a)^2/2\sigma^2} \times \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} x_\mathrm{c}^{3(N-L+j)+m}\,\mathrm{d}x_\mathrm{c},$$

$$= \frac{1}{\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \times \int_0^{+\infty} \mathrm{e}^{-(x_\mathrm{c}+a)^2/2\sigma^2} x_\mathrm{c}^{3(N-L+j)+m}\,\mathrm{d}x_\mathrm{c}.$$

Let $t = 3(N-L+j) + m$ and $y = x_\mathrm{c} + a$, then

$$p(z=L|H_1)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \times \int_a^{+\infty} \mathrm{e}^{-y^2/2\sigma^2} (y-a)^t\,\mathrm{d}y,$$

$$= \frac{1}{\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \sum_{i=0}^{t} C_t^i (-a)^i \times \int_a^{+\infty} \mathrm{e}^{-y^2/2\sigma^2} y^{t-i}\,\mathrm{d}y,$$

$$= \frac{1}{2\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \times \sum_{i=0}^{t} C_t^i (-a)^i\,(\sqrt{2}\sigma)^{t-i+1} \int_{a^2/2\sigma^2}^{+\infty} \mathrm{e}^{-y} y^{\left(\frac{t-i+1}{2}-1\right)}\,\mathrm{d}y,$$

$$= \frac{1}{2\sqrt{2\pi}\sigma} C_N^L \sum_{j=0}^{L} (-1)^j C_L^j \left(\frac{2}{3}\right)^{N-L+j} \mathrm{e}^{-(\mu^2-a^2)/2\sigma^2} \sum_{m=0}^{N-L+j} C_{N-L+j}^m 2^{N-L+j-m} \sum_{i=0}^{t} C_t^i (-a)^i \times (\sqrt{2}\sigma)^{t-i+1} \Gamma\left(\frac{t-i+1}{2}, \frac{a^2}{2\sigma^2}\right),$$

where

$$\Gamma\left(\frac{t-i+1}{2},\frac{a^2}{2\sigma^2}\right) = \Gamma\left(\frac{t-i+1}{2}\right) - \gamma\left(\frac{t-i+1}{2},\frac{a^2}{2\sigma^2}\right)$$
$$= \Gamma\left(\frac{t-i+1}{2}\right) - \sum_{h=0}^{+\infty}\frac{(-1)^h\left(\frac{a^2}{2\sigma^2}\right)^{\frac{t-i+1}{2}+h}}{h!\left(\frac{t-i+1}{2}+h\right)},$$

and

$$\Gamma\left(\frac{t-i+1}{2}\right) = \begin{cases}\left(\frac{t-i-1}{2}\right)! & \text{if } t-i+1 \text{ is even,}\\ \frac{(t-i-1)!}{2^{((t-i)/2)}}\sqrt{\pi} & \text{else.}\end{cases}$$

References

Zhang, H.J., Kankanhalli, A., Smoliar, S.W., 1993. Automatic partitioning of full-motion video. Multimedia Systems 1 (1), 10–28.

Sethi, I.K., Patel, N., 1995. A statistical approach to scene change detection. SPIE Proceedings 2420, 329–338.

Zhang, H.J., Kankanhalli, A., Smoliar, S.W., 1995. Video parsing and browsing using compressed data. Multimedia Tools and Applications 1 (1), 89–111.

Shahraray, B., 1995. Scene change detection and content-based sampling of video sequences. In: Proceedings SPIE Conference on Digital Video Compression: Algorithms and Technologies, vol. 2419, pp. 2–13.

Zabih, R., Miller, J., Mai, K., 1995. Feature-based algorithms for detecting and classifying scene breaks. In: Proceedings of the Fourth ACM International Conference On Multimedia, San Francisco, CA, pp. 189–200.

Hampapur, A., Jain, R., Weymouth, T., 1995. Production model based digital video segmentation. Multimedia Tools and Applications 1 (1), 9–46.

Minkler, G., Minkler, J., 1990. The Principles of Automatic Radar Detection in Clutter—CFAR. Magellan Book Company, Baltimove, MD.

Carlyle, J.W., Thomas, J.B., 1964. On non-parametric signal detectors. IEEE Transactions on IT 10 (2), 146–152.

Hanjalic, A., Zhang, H.J., 1999. Optimal shot boundary detection based on robust statistical models. In: Proceedings of the IEEE International Conference on Multimedia Computing and Systems, vol. 2, pp. 710–714.

Lupatini, G., Saraceno, C., Leonardi, R., 1998. Scene break detection: a comparison. In: Proceedings of the Eighth International Workshop on Research Issues In Data Engineering: Continuous-Media Databases and Applications, pp. 34–41.

Fernando, W.A.C., Canagarajah, C.N., Bull, D.R., 2000. A unified approach to scene change detection in uncompressed and compressed video. IEEE Transactions on Consumer Electronics 46 (3), 769–779.