Detecting Changes in Unlabeled Data Streams using Martingale

Shen-Shyang Ho and Harry Wechsler
George Mason University
Department of Computer Science
{sho, wechsler}@cs.gmu.edu

Abstract

The martingale framework for detecting changes in data stream, currently only applicable to labeled data, is extended here to unlabeled data using clustering concept. The one-pass incremental change-detection algorithm (i) does not require a sliding window on the data stream, (ii) does not require monitoring the performance of the clustering algorithm as data points are streaming, and (iii) works well for high-dimensional data streams. To enhance the performance of the martingale change detection method, the multiple martingale test method using multiple views is proposed. Experimental results show (i) the feasibility of the martingale method for detecting changes in unlabeled data streams, and (ii) the multiple-martingale test method compares favorably with alternative methods using the recall and precision measures for the video-shot change detection problem.

1 Introduction

The problem of detecting changes in low-dimensional sequential data has been studied by statisticians for more than fifty years. Methods of change detection first appeared in the 1940s based on Wald's sequential analysis [Wald, 1947], in particular the sequential probability sequential test (SPRT) [Basseville and Nikiforov, 1993], and later, Page introduced the cumulative sum method [Page, 1954]. Recently, the machine learning and the data mining communities become interested in the change detection problem due to the need to discover changes in data, such as customer click streams, high-dimensional multimedia data, and retail chain transactions, generated from online processes that are not stationary [Domingos and Hulten, 2001]. The target concepts change over time. It is, hence, vital to detect the changes in the data generating processes so that timely decisions can be made.

One real-world problem that requires detecting changes is the video segmentation problem which corresponds to video-shot change or video break detection. Many algorithms [Gargi *et al.*, 2000; Lefevre *et al.*, 2003; Zhai and Shah, 2005] have been proposed to perform video-shot change detection. The range of existing methods includes pixel and histogram-based difference methods and motion-based methods (e.g. optical flow). Threshold selection, a critical step for successful change detection, is required by methods using global or local thresholds. For video sequences with clear and distinct shots, a single global threshold would be sufficient. For video sequences that have both abrupt and gradual changes between shots, however, a global threshold cannot be found. To address such concerns [Gargi *et al.*, 2000] argued for the use of local thresholds. The use of local thresholds requires choosing appropriate window size. Alternatively, Zhai and Shah [Zhai and Shah, 2005] proposed that video breaks should be detected using the deviation from some current model.

Recently, [Ho, 2005] proposed a martingale method for change detection for high-dimensional labeled data streams. An adaptive support vector machine for time-varying data streams was proposed based on the martingale method [Ho and Wechsler, 2005a]. The main contributions of the present paper are: (i) the extension of the martingale methodology to unlabeled data stream, and (ii) the multiple-martingale test based on multiple views (features). We show empirically that the martingale method works well for unlabeled data stream and the multiple-martingale test compares favorably with alternative video-shot change detection methods for unlabeled video stream.

The paper is organized as follows. In Section 2, we review the concept of exchangeability and martingale. In Section 3, we introduce the concept of strangeness and p-values, and propose the strangeness measure for unlabeled data using clustering concept. In Section 4, we state the theorem on the martingale method for change detection which is applicable to both labeled and unlabeled data stream. In Section 5, we show the feasibility of the martingale method for change detection on an unlabeled synthetic data stream. In Section 6, we propose the multiple-martingale test for change detection. In Section 7, we apply the multiple-martingale test to video-shot change detection.

2 Exchangeability and Martingale

Let $\{Z_i : 1 \leq i < \infty\}$ be a sequence of random variables. A finite sequence of random variables $Z_1, \cdots, Z_i$ is *exchangeable* if the joint distribution $p(Z_1, \cdots, Z_i)$ is invariant under any permutation of the indices of the random variables. The satisfaction of exchangeability condition indicates that the distribution that the sequence of random variables is drawn from is stationary.

[Vovk *et al.*, 2003] introduced the idea of testing exchangeability online using the martingale. A *martingale* is a sequence of random variables $\{M_i : 0 \leq i < \infty\}$ such that M_n is a measurable function of $Z_1, \cdots, Z_n$ for all $n = 0, 1, \cdots$ (in particular, M_0 is a constant value) and the conditional expectation of M_{n+1} given $M_0, \cdots, M_n$ is equal to M_n, i.e.

$$E(M_{n+1}|M_1, \cdots, M_n) = M_n \tag{1}$$

After each new data point is received, an observer outputs a positive martingale value reflecting the strength of evidence found against the null hypothesis of data exchangeability. The testing of exchangeability is used to detect changes in time-varying labeled data streams [Ho and Wechsler, 2005b].

3 Strangeness and p-values

To apply the martingale method one needs to rank the data points according to their differences. Towards that end, one defines a *strangeness measure* that scores how much a data point is different from the other data points. Consider the set of labeled data points $Z = \{z_1, \cdots, z_{n-1}\}$ and the new labeled data point z_n. Each data point is assigned a strangeness value based on the classifier, such as the support vector machine (SVM) or the nearest neighbor rule, used to classify the data points $Z \bigcup \{z_n\}$ [Vovk *et al.*, 2005].

To define a valid strangeness value for each data point, the simple assumption that *at any time instance the strangeness value of each data point seen so far should be independent of the order these data points are used in the strangeness computation* must be satisfied [Vovk *et al.*, 2005]. For instance, when a k-nearest neighbor rule is used, the strangeness value of a particular data point is the ratio of the sum of the k-nearest data points with similar label ($\mathcal{S}_s$) to the sum of the k-nearest data points with difference label ($\mathcal{S}_d$). Hence, the higher $\mathcal{S}_s$ is, the higher the strangeness value, and vice versa. On the other hand, when $\mathcal{S}_d$ is high, the strangeness value is low. This is clearly a method that constructs valid strangeness values for labeled data points.

For unlabeled data, the strangeness measure is derived using clustering algorithm such as K-mean/median clustering with $K = 1$. Consider the set of unlabeled data points $Z = \{z_1, \cdots, z_{n-1}\}$ and the new unlabeled data point z_n. The strangeness value s_i of z_i for $i = 1, \cdots, n$ is

$$s_i(Z, z_n) = \parallel z_i - C(Z \cup \{z_n\}) \parallel \tag{2}$$

where $C(\cdot)$ is some cluster representation and $\parallel \cdot \parallel$ is some distance measure. The strangeness value for a data point is high when it is further away from the cluster representation, e.g. the cluster center.

Next, a statistic is constructed to rank the strangeness value of the new data point z_n with respect to the strangeness values of all the observed data points. The statistic, called the p-value of z_n, is defined as

$$\text{p-value} = V(Z \cup \{z_n\}, \theta_n) = \frac{\#\{i : s_i > s_n\} + \theta_n \#\{i : s_i = s_n\}}{n} \tag{3}$$

where s_i is the strangeness measure for z_i, $i = 1, 2, \cdots, n$ and θ_n is randomly chosen from $[0, 1]$.

The random number θ_n in (3) ensures that the p-values $p_1, p_2, \cdots$ output by the p-value function V are distributed uniformly in $[0, 1]$, provided that the input examples $z_1, z_2, \cdots$ are generated by an exchangeable probability distribution in the input space [Vovk *et al.*, 2003]. This property of output p-values no longer holds when the exchangeability condition is not satisfied.

4 Change Detection using Martingale

Intuitively, we assume that a sequence of data points with a change consists of concatenating two data segments, S_1 and S_2, such that the data distribution of S_1 and S_2 are P_1 and P_2 respectively and $P_1 \neq P_2$. Switching a data point z_i from S_2 to a position in S_1 will make the data point stand out in S_1. The exchangeability condition is, therefore, violated. Exchangeability is a sufficient condition for a stable data stream. The absence of exchangeability suggests the occurrence of change.

A family of martingales, indexed by $\epsilon \in [0, 1]$, and referred to as the *power martingale*, is defined as

$$M_n^{(\epsilon)} = \prod_{i=1}^{n} \left(\epsilon p_i^{\epsilon - 1}\right) \tag{4}$$

where the p_is are the output p-values from the function V, with the initial martingale $M_0^{(\epsilon)} = 1$. We note that $M_n^{(\epsilon)} = \epsilon p_n^{\epsilon-1} M_{n-1}^{(\epsilon)}$. Hence, it is not necessary to store the previous p-values. In our experiments, we use $\epsilon = 0.92$, which is within the desirable range where the martingale value is more sensitive to a violation of the exchangeability condition [Vovk *et al.*, 2003].

The following theorem is applicable to both labeled and unlabeled data streams as long as the assumption (stated in Section 3) on the strangeness measure is satisfied.

Theorem 1 ([Ho and Wechsler, 2005b]) *Let $\{M_i^{(\epsilon)} : 0 \leq i < \infty\}$ be a martingale sequence of the form (4) constructed using p-values $\{p_i : 1 \leq i < \infty\}$ computed from (3) based on a valid strangeness measure for a given data stream:*

1. *If no change occurs in the given data stream, then*

$$P\left(\max_k M_k^{(\epsilon)} \geq \lambda\right) \leq \frac{1}{\lambda}, \tag{5}$$

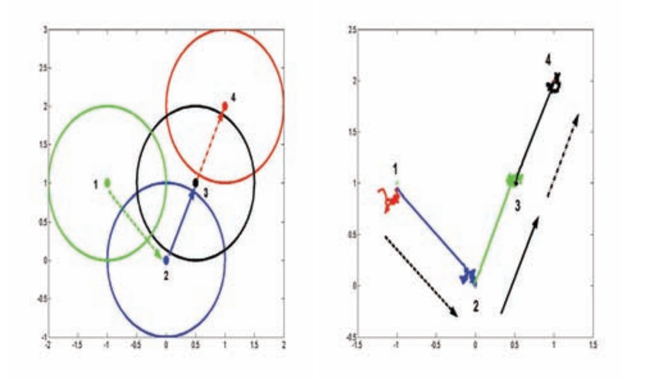
where λ is a positive number.

2. *Let α be the size of the test deciding in favor of the alternative hypothesis "change occurs in the data stream" when the null hypothesis "no change occurs in the data stream" is true and $1 - \beta$ be the power of the test deciding in favor of the alternative hypothesis when it is true, the martingale test according to (5) is an approximation of the sequential probability ratio test (SPRT), with*

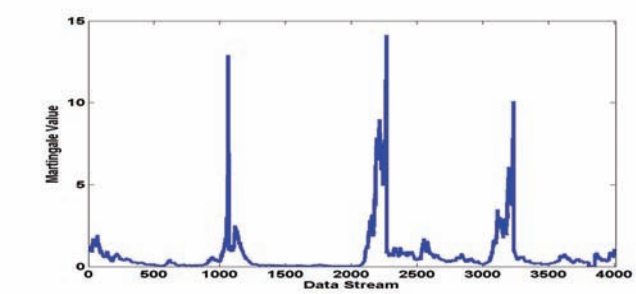
$$\lambda \leq \frac{1 - \beta}{\alpha}, \tag{6}$$

and the mean delay time $E(m)$, i.e. the expected number of data points, m, observed before a change is detected, is approximated from the SPRT as follows:

$$E(m) \approx \frac{(1 - \beta) \log \lambda}{E(\mathcal{L})} \tag{7}$$



(a)



(b)

Figure 1: (a) Left: A sketch of four overlapping 2-dimensional Gaussian distributions and the way the data sequence changes in distribution; Right: The trajectory of the center of the 1-mean cluster as data points are observed one by one; (b) The martingale values of the 10-D synthetic unlabeled data stream with 3 change points.

where

$$\mathcal{L} = \log \epsilon p_i^{\epsilon - 1} \tag{8}$$

A user selects a desirable threshold λ for the martingale test (5) based on (6). To estimate the start of the change, (7) is used .

5 Experimental Result: Synthetic Unlabeled Data Stream

We construct an artificial unlabeled data stream as follows:

1. Randomly generate four sets of 1050 data points S_1, S_2, S_3, and S_4 from four overlapping Gaussian distributions with randomly generated mean with variance 1.
2. Concatenate S_1, S_2, S_3, and S_4 to form a sequence D consisting of 4200 data points such that $D = \{S_1; S_2; S_3; S_4\}$.

Left graph in Figure 1(a) shows a sketch of four overlapping 2-dimensional Gaussian distributions generated by Step 1 of the above procedure. The data distribution of the data sequence D changes with 3 change points.

In the experiment, the first 50 points from D is first used to compute the center of the initial 1-mean cluster with strangeness of each data point computed using (2) with the Euclidean distance. The threshold λ value for the martingale test is set to 10. The data points from the data stream D are observed sequentially. The cluster center is updated. Strangeness for all seen data points are computed. The martingale at each instance is computed using (4) based on the computed p-values from (3). When the martingale value is greater than λ, a change in the mean of the Gaussian distribution is detected. All previously seen data points are removed. The martingale is reset to 1. The next 50 data points are used to construct a new 1-mean cluster. The process repeats till the data points in D are exhausted.

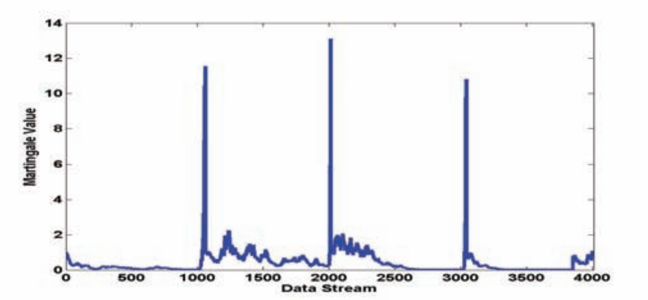


Figure 2: The martingale values of the 10-D synthetic unlabeled data stream with 3 change points.

Right graph in Figure 1(a) shows the trajectory of the center of 1-mean cluster in the experiment. The theoretical basis for signaling a change is (5). Figure 1(b) shows the martingale values. Detection occurs when the martingale value is greater than $\lambda = 10$. The true change points are at data points 1000, 2000 and 3000. The dimension of the Gaussian distribution is then increased to 10. Figure 2 shows that the changes in mean parameter of the high-dimensional Gaussian distribution are detected as well.

We note that in some real-world problems, more than one data feature is good in representing the data. In order to utilize this observation to improve the sensitivity and performance of the martingale test, one can perform a number of martingale tests constructed based on different characteristics or features of the data.

6 Multiple-Martingale Test using Multi-Views

In the multi-view learning problem, an example z is represented by a number of feature subsets. Each feature subset describes a view of the example [Muslea *et al.*, 2002]. The multi-view setting is closely related to co-training [Blum and Mitchell, 1998].

For the multiple-martingale test, we consider the multi-view setting such that each constructed martingale attempts to identify changes with respect to the particular feature subset. Besides the fact that the features are extracted from the original data, the feature subsets should be independent of one another to minimize redundancy.

We note that according to the martingale theory, the data representation does not affect the probability bound (5). On the other hand,

Corollary 1 *When the multiple martingale test with* $\mathcal{M}$-

views is used for change detection, the expected number of data points, m, observed before a change point is detected,

$$E^*_{\mathcal{M}}(m) \leq E(m) \quad (9)$$

Obviously, one can conclude that the number of miss detection using the multiple-martingale test is upper-bounded by the number of miss detection using the original martingale method. Moreover, the number of false alarm using the multiple martingale test is lower-bounded by the number of false alarm using the original martingale method.

Algorithm 1 is the multiple-martingale test with two views. The algorithm can be trivially extended to $\mathcal{M} > 2$.

Algorithm 1 : Multiple-Martingale Test with $\mathcal{M}$-View, $\mathcal{M} = 2$

Initialize: $M_1(0) = M_2(0) = 1; i = 1; T = \{\}$.
Set: λ.

loop
  A new unlabeled example x_i is observed.
  Construct the two views/features from x_i, i.e., $\bar{x}_i = \{f_{i1}, f_{i2}\}$.
  Compute the strangeness measure $\vec{s}_1$ and $\vec{s}_2$ (vectors containing strangeness for seen examples) using (2) from $\{f_{11}, \cdots, f_{i1}\}$ and $\{f_{12}, \cdots, f_{i2}\}$.
  Compute the p-values p_1 and p_2 from $\vec{s}_1$ and $\vec{s}_2$, respectively, using (3).
  Compute $M_1(i)$ and $M_2(i)$ from p_1 and p_2 using (4).
  if $M_1(i) > \lambda$ OR $M_2(i) > \lambda$ **then**
    CHANGE DETECTED
    Set $M_1(i)$ and $M_2(i)$ to 1;
    Re-initialize T to an empty set.
  else
    Add x_i into T.
  end if
  $i := i + 1$;
end loop

7 Video-Shot Change Detection Problem

In Section 7.1, we describe the method of constructing valid strangeness for video stream. In Section 7.2, we analyze the martingale test method and motivate using the multiple-martingale test on detecting changes in video stream. In Section 7.3, we compare the multiple martingale test method with alternative video-shot change detection methods.

7.1 Strangeness Measure for Video Stream

We describe the representation and the strangeness measure for the unlabeled video image in this subsection.

Image Representation Using Color and Edge Histograms

The image representations are based on the color and edge histograms. The color histogram is constructed with 4096 bins. $[(r/16) * 256 + (g/16) * 16 + b/16]$ is used to convert RGB values of a pixel into an index, with integer division where r, g and b are red, green and blue values respectively. The edge histogram consists of 36 bins such that a gradient angle is converted into an index by dividing the orientation angle by 10 and rounding to the nearest integer.

To capture local information from an image frame, the image frame is partitioned such that a histogram is constructed for each area partitioned along either the horizontal or the vertical axis. For example, for a 352×240 image, histograms are constructed on 3 areas of 88×240 by partitioning along the horizontal axis and 3 areas of 352×80 by partitioning along the vertical axis. Using this construction, we have two image representations: one consisting of six color histograms, $R_{cc} = \{H_c^1, H_c^2, H_c^3, H_c^4, H_c^5, H_c^6\}$, and one consisting of six edge histograms, $R_{ee} = \{H_e^1, H_e^2, H_e^3, H_e^4, H_e^5, H_e^6\}$. The two image representations are used as the two views in Algorithm 1.

Strangeness Measure of the Image Representations, R_{cc} and R_{ee}

Consider the set of image representations, i.e. $Z = \{R_1, \cdots, R_{(n-1)}\}$ and a new image representation R_n in the form of either R_{cc} or R_{ee}. First, we define

$$\Omega = \max\left(Z \cup \{R_n\}\right) \quad (10)$$

which contains the maximum value for each bin in the histograms of the image representation.

We note that as the number of observed image frames increases, $||\Omega||$, where $||\cdot||$ is some distance measure, is monotonically increasing. Ω ensures that the cluster center will maintain information from previously observed image frames and not be affected by a small drift in image content when a new image frame is observed. The effect of a small drift in image content is significant when the mean, median or minimum value is used. Ω assumes the role of the cluster center for the images.

We are interested in the difference between the image representation $R_i, i = 1, \cdots, n$ and Ω. Using the Euclidean norm, the strangeness value for R_i is

$$s_i(Z, R_n) = \sqrt{\sum_{k=1}^{\text{No. of bins}} \left(R_i(k) - \Omega(k)\right)^2}. \quad (11)$$

This strangeness measure is valid for an image representation R_i in the form of either R_{cc} or R_{ee} as it satisfies the assumption in Section 3. The set of computed strangeness values $\{s_1, \cdots, s_n\}$ for $R_i, i = 1, \cdots, n$ is used to compute the p-value of R_n using (3).

7.2 Characteristics of the Martingale Method for Video Stream

Consider the example shown in Figure 3, we use the image representation R_{cc} and the threshold value $\lambda = 10$ to demonstrate some characteristics of the martingale method.

The characteristic of Ω_c (for color histogram representation, R_{cc}) is shown in Figure 4. The shot change is detected at Frame 158 and all previous information is removed from the memory. Unlike labeled data stream where the data-generating process changes at some particular time instances [Ho, 2005; Ho and Wechsler, 2005b], the unlabeled video sequence is usually non-stationary, with constant small changes. The "cluster" representation Ω_c should be either (i)

Figure 3: Example: Starting at (Left Image) Frame 80 to (Center Image) Frame 200: Human starts appearing in Frame 98 and appearing completely at around Frame 150. The shot change is detected at Frame 158 (Right Image).

insensitive to small changes or (ii) maintain most information from the previously observed data. Ω_c satisfies the two criteria.

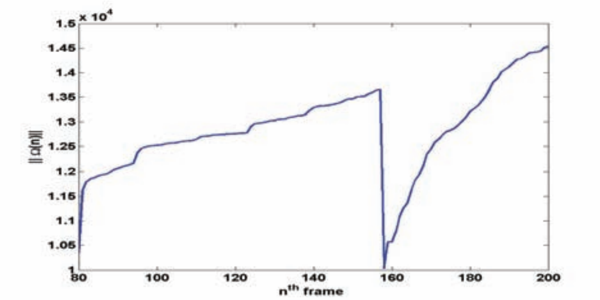


Figure 4: Using the example shown in Figure 3: The characteristic of Ω_c using the image representation R_{cc} when the video sequence is nearly stationary and after a shot change is detected at Frame 158.

We perform the Kolmogorov-Smirnov test (KS-Test) to see whether the p-values computed using (3) is uniformly distributed when the video content is nearly stationary. This important property of p-values computed using (3) does not hold when the video sequence no longer satisfies the exchangeability condition. The mean of the p-values lowers and the martingale values start to increase. In this experiment, we perform the martingale method without reacting to shot change detected, removing data points and resetting variables. In Figure 5, we observed that with high confidence (based on p-values obtained from the KS-Test) the p-values computed using (3) is uniformly distributed before Frame 163. After Frame 163, the p-values are no longer uniformly distributed at 0.05 significance level.

In Figure 6, we show the martingale values computed using the color feature, R_{cc} and the edge feature, R_{ee}. One notes that for this particular example shown in Figure 3, the edge feature is more sensitive than the color feature. The martingale values computed using the edge feature increase faster and higher than the ones computed using the color feature near the image frame where the change occurs. Hence, by using the multiple-martingale test with 2 views, the number of image frames observed before detecting the change point is lowered. Hence, the sensitivity of the martingale method can be increased using multiple views.

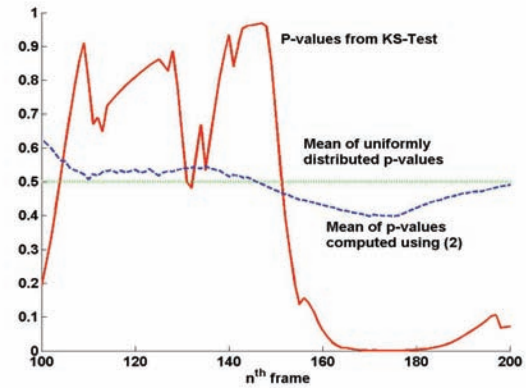


Figure 5: Using the example shown in Figure 3: The characteristics of the p-values computed using (3) and the image representation R_{cc} on the video sequence. We perform the Kolmogorov-Smirnov Test (KS-Test) starting at Frame 100. Frame 177 has the lowest p-value using the KS-Test at 0.000598 significance level and a significance level of less than 0.05 is observed after Frame 163.

7.3 Comparison with Alternative Methods

Experiments were performed to compare the multiple-martingale test with alternative methods. Four sets of videos, consisting of two documentary video streams (anni, ugs) [1] and two video streams from hand-held digital video camera (outdoor, indoor), are used. The "anni" video stream consists of 9 shots and 8 gradual transitions between shots. The "ugs" video stream consists of 13 shots with some fast transitions and moving camera. The "outdoor" and "indoor" video streams consist of continuous changes due to the motion of the camera. The two video streams are captured from the top of a hill and from an apartment, respectively.

The performance of the methods is measured based on the number of detections, miss detections and false detections. They are summarized using the recall, precision and the F_1 measure:

$$\text{Precision} = \frac{\text{Number of Correct Detections}}{\text{Number of Detections}}$$
$$\text{Recall} = \frac{\text{Number of Correct Detections}}{\text{Number of True Changes}}$$
$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Precision is the probability that a detection is actually correct, i.e. detecting a true change. Recall is the probability that a change detection system recognizes a true change. F_1 measure represents a harmonic mean between recall and precision. A high value of F_1 measure ensures that the precision and recall are reasonably high.

In the experiment, we use $\lambda = 20$ which corresponds to the fact that the computed martingale values are unlikely to be higher than 20 with probability bound of 0.05 if no change

[1] The documentary videos can be obtained freely from http://www.open-video.org

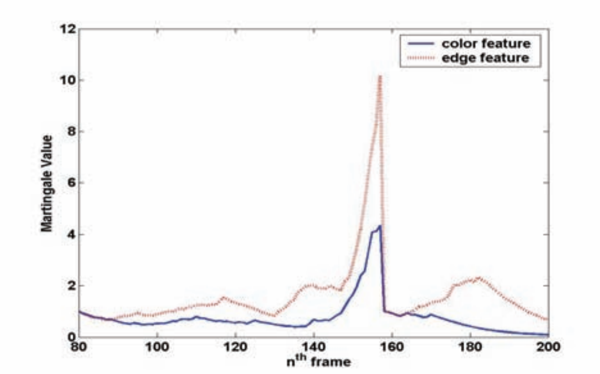


Figure 6: The characteristics of martingale values M_n using both the image representations R_{cc} and R_{ee} on the example in Figure 3. The edge feature is more sensitive than the color feature in this particular example.

occurs. We compare the multiple-martingale test (MT (20)) with the alternative methods. For the alternative methods, the color histogram is used as the image representation and the local threshold is selected based on window averaging [Gargi *et al.*, 2000] on three similarity measures: histogram intersection (HI), chi-square measure (χ^2) and Euclidean distance (ED), to measure the similarity between any two image frames for all the image frames in a video sequence. The parameters, such as the window size, are varied so that the best results can be obtained from the alternative methods. The experimental results on the four video sequences are shown in Table 1.

	Video	anni	ugs	outdoor	indoor
Method	Length	0:30	1:11	1:20	2:27
	Frames	913	2145	1202	2210
	Shots	17	13	18	26
MT (20)	Precision	0.933	0.857	0.824	0.686
	Recall	0.824	0.923	0.778	0.923
	F_1	0.875	0.889	0.800	0.787
HI	Precision	0.813	0.733	0.889	0.622
	Recall	0.765	0.846	0.889	0.885
	F_1	0.788	0.785	0.889	0.731
χ^2	Precision	1.000	0.706	0.553	0.648
	Recall	0.765	0.923	0.889	0.923
	F_1	0.807	0.800	0.682	0.761
ED	Precision	0.765	0.520	0.524	0.548
	Recall	0.765	1.000	0.647	0.885
	F_1	0.765	0.684	0.579	0.677

Table 1: Experimental Results.

Based on the F_1 measure, the multiple-martingale test performed the best three out of the four video sequences. The main difference between the alternative methods and ours is the fact that our threshold is chosen a prior with theoretical consideration while the thresholds selected by the alternative methods are heuristics such that the window size and other parameters need to be tuned to achieve good result. Moreover, the martingale method is a one-pass algorithm while the alternative methods are not.

8 Conclusions

In this paper, we extend the martingale methodology [Ho, 2005] to be used on unlabeled data. We also proposed the multiple-martingale test based on building different martingales using multiple views (features) to enhance the performance of the martingale methodology. We apply this multiple martingale test method to the video-shot change detection and show that it compares favorably with alternative methods.

References

[Basseville and Nikiforov, 1993] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, 1993.

[Blum and Mitchell, 1998] Avrim Blum and Tom M. Mitchell. Combining labeled and unlabeled sata with co-training. In *COLT*, pages 92–100, 1998.

[Domingos and Hulten, 2001] Pedro Domingos and Geoff Hulten. Catching up with the data: Research issues in mining data streams. In *DMKD*, 2001.

[Gargi *et al.*, 2000] Ullas Gargi, Rangachar Kasturi, and Susan H. Strayer. Performance characterization of video-shot-change detection methods. *IEEE Trans. Circuits Syst. Video Techn.*, 10(1):1–13, 2000.

[Ho and Wechsler, 2005a] Shen-Shyang Ho and Harry Wechsler. Adaptive support vector machine for time-varying data streams using martingale. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *IJCAI*, pages 1606–1607. Professional Book Center, 2005.

[Ho and Wechsler, 2005b] Shen-Shyang Ho and Harry Wechsler. On the detection of concept change in time-varying data streams by testing exchangeability. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence*, pages 267–274, 2005.

[Ho, 2005] Shen-Shyang Ho. A martingale framework for concept change detection in time-varying data streams. In Luc De Raedt and Stefan Wrobel, editors, *ICML*, pages 321–327. ACM, 2005.

[Lefevre *et al.*, 2003] Sébastian Lefevre, Jérôme Holler, and Nicole Vincent. A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. *Real-Time Imaging*, 9:73–98, 2003.

[Muslea *et al.*, 2002] Ion Muslea, Steven Minton, and Craig A. Knoblock. Active + semi-supervised learning = robust multi-view learning. In Claude Sammut and Achim G. Hoffmann, editors, *ICML*, pages 435–442. Morgan Kaufmann, 2002.

[Page, 1954] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[Vovk *et al.*, 2003] Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Testing exchangeability on-line. In Tom Fawcett and Nina Mishra, editors, *ICML*, pages 768–775. AAAI Press, 2003.

[Vovk *et al.*, 2005] Vladimir Vovk, Alex Gammermann, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005.

[Wald, 1947] A. Wald. *Sequential Analysis*. Wiley, N. Y., 1947.

[Zhai and Shah, 2005] Y. Zhai and Mubarak Shah. A general framework for temporal video scene segmentation. In *Proc. 10th International Conference on Computer Vision, Beijing, China*, 2005.