# Comparative Study for Digital Video Automatic Segmentation Algorithms

**Abdul Monem S. Rahma, Hilal Hady Saleh and Faten Hsan AL-Kathy ***
*Department of Computer Science, University of Technology. Baghdad-Iraq.*
** Department of Computer Science, Karkuk Technical College. Karkuk-Iraq.*

**Abstract**

While several methods of automatic video segmentation for the identification of shot transition have been proposed, they have not been systematically compared. In this paper, several segmentation techniques cross different type of video are examined. Each of these techniques defines a measure of dissimilarity between successive frames, which is then compared to a threshold .Dissimilarity values exceeding the threshold identify shot transition.
The techniques are compared in terms of the percentage of correct and false identifications for various thresholds, their sensitivity to the threshold value their performance across different type of video, their ability to identify complicated transition effects, and their requirements for computational resources. Finally, the definition of a priori set of values for the threshold parameter is also examined. Most techniques can identify over 90% of the real shot transitions but have a high percentage of false positives. Reducing the false positives was a major challenge; this paper introduced a local filtering technique that was fairly effective.

**الخلاصة**

أقتـرح العديـد مـن طـرق تقطيـع أفـلام الفيـديو للكـشف عـن هويـة انتقـالات الكـاميرا (Shot Transitions) والتي لم تقارن بشكل منظم , هذة الورقة تتضمن اختبار لعدة تقنيات للتقطيع طبقت على أنواع مختلفة من أفلام الفيديو كل من هذة التقنيات اعتمدت مقياس للتباين بين الإطارات المتتالية (Successive Frames)التي قورنت بقيم عتبة الحد الفاصل (Threshold) فالقيم التي تتجاوز عتبة الحد الفاصل ستحدد انتقالات الكاميرا .
تم مقارنة التقنيات المقترحة للتقطيع وفق نسبة الخطأ والصواب ولعدة قيم من الحد الفاصل لبيان مدى تأثرها بقيم عتبة الحد الفاصل ,أدائها لأنواع مختلفة من الفيديو,قدرتها على التحقق من هوية مؤثرات الانتقالات المعقدة و متطلباتها من الموارد الحسابية .
أغلب تقنيات التقطيع حققت أكثر من 90% من الانتقالات الحقيقة للكاميرا لكنها تحوي نسبة عالية من الخطـأ والـذي تقليلـه يعتبـر اكبـر رهـان لنجـاح عمـل التقنيـة,تـم اقتـراح تقنيـة التـصفية الموقعيـة (Filtering Techniques Local) التي أعطت نتائج فعلية جيدة.

## Introduction

To extract useful information from a digital video source, the information in it need to be indexed and annotated. The services that can be provided after indexing include video database, video browser capable of presenting information in an organized manner. To facilities browsing and retrieval of digital video, it is necessary to be able to detect logical scene breaks. Automatic video segmentation is a first step towards video indexing. Digital video segmentation problem is defined as the automatic detection of shot boundaries, just as a text can be divided into sentences, a video sequences can be hierarchically divided into smaller components such as scenes and shot, as shown in (Figure 1) [ENG 96][ 9] .

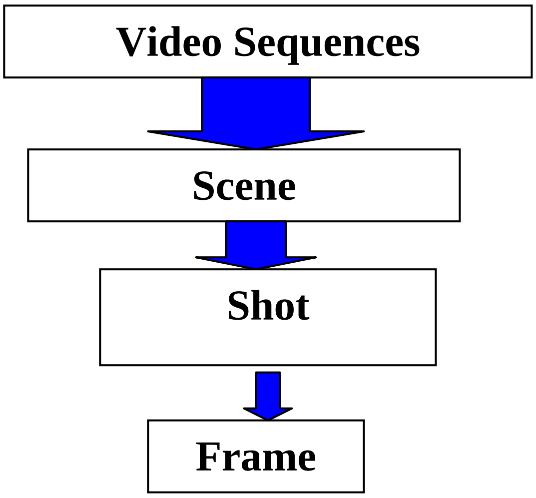


**Figure 1: Show structural view of a video sequences**

Typically the annotation of the information in the video has to be associated with shot transitions where a shot is defined as the sequences of successive frames from the moment a camera starts recording until it stop (is a sequences of images taken from a single operation of the camera and that depicts a continuous action in time and space) [14].

While scene is a group of related shots(is defined as a collection of semantically related and temporally adjacent shots depicting and conveying a high-level concept or story.

During the shot duration ,both the camera and the objects in the image can move ,but the sequences of frames is part of the same shot .Therefore the first step in indexing digital video is the identification of shot transitions.

There are four different types of shot change:

- **Cut:** is an instantaneous transition from one scene to the next, and it occurs over two frames.
- **fade**: is a gradual transition between a scene and a constant image (fade out) or between a constant image and a scene (fade in).A fade involves slowly changing the shot intensity to give the impression that a shot is appearing or disappearing.
- **dissolve**: is a gradual transition from one scene to another in which the first scene fade out and the second fade in. A dissolve combines a fade-out and a fade-in as one shot vanishes another shot appears .In a dissolve for a short period of time both shots are super imposed.
- **wipe**: occurs as a line moves across the screen, with the new scene appearing behind the line. A typical wipe involves one shot replacing the other as a line crosses the screen. As one side of the line shows the old shot the other shows the new shot. As the line crosses the screen the old shot disappears. In a wipe, shots are not superimposed but occupy different areas of the screen.

Examples of some of the common shot transitions are illustrated in Figure2 [14].

## Segmentation Methods

The methods define below is examine two frame f and f' and define a measure of dissimilarity d (f, f') whose value can be used to identify shot transition.

### 1- Absolute Frame-Difference Segmentation [5]

The simplest measure of the difference between two consecutive frames is the absolute difference of the sum of the intensities of all pixels in the frame .The gray –level intensity of a pixel is defined as 0.299r+0.587g+0.114b for a color image where r, g and b are the intensities for the three basic colors each represented by a one byte number.

### 2- Histogram Based Segmentation [7]

Among the simplest, most effective and most commonly used methods are the histogram based method and its many variations which rely on the following basic idea. The number of values each pixel can have is discretized and a histogram is created for a frame by counting the number of times each of the discrete values appear in the frame. Then this histogram is compared with the histogram of the next frame .Histogram of the frames within the same shot should be very similar to each other even in the case of camera or object motion because the method is relatively insensitive to the position of object within the frame.

Let H (f, k) be the value of the histogram f and for the discrete value of intensity k .The value of k is in the range [0, N], where N is the number of discrete values a pixel can have. There are usually three histogram for each frame one for each color. Several variation of the histogram method are described in the following subsections.

### 2.1 Simple Histogram Difference [3]

The metric of the dissimilarity between frames f and f' is given by

$$d(f, f') = \sum_{J=0}^{N} |H(f, j) - H(f', j)|$$

### 2.2 Weighted Histogram Difference [11]

In a frame there might be some dominant color which should be given a greater weight in the comparison between two frames thus developed a formula for the histogram difference which is defined as

$$d(f,f') = r/s.\ d(f,f') + g/s.\ d(f,f') + b/s.\ d(f,f')$$

where r, g and b are the luminance for the red, green and blue components of the picture respectively and s is defined as (r + g + b)/3

### 2.3 Histogram Difference after Equalization [5] [1]

The goal of histogram equalization is to produce a uniform histogram $H^e(f,j)$ for the output frame.

$$d(f,f') = \sum_{J=0}^{N} |H^e(f,j) - H^e(f',j)|$$

Where the equalized $H^e(f,j)$ is obtained in the same way as $H(f,j)$ after transforming the value v of a pixel to $Veq$ through the following procedure

$$V_{eq} = int\ [w - w_{min} / 1 - w_{m\,n}\ .\ (L-1) + 0.5\ ]$$

Where $L$ is the number of levels of the pixel value $w$ is given through

$$w = 1/ \sum_{J=0}^{N-1} H(f,j)\ .\ \sum_{J=0}^{V} H(f,j)$$

And $w_{min}$ is smallest positive value of $w$.

### 2.4 Intersection of Histograms [10] [11]

The intersection of two histogram which also serve as the metric of the similarity $s(f,f')$ between two frames is defined as:

$$IS^{(color)}(f,f') = s(f,f') = \sum_{J=0}^{N} min\,(H(f,j), H(f',j))$$

When taking the intersection of two identical frames the above value is maximum and equal to the number of pixels in the frame, whereas in dissimilar frames this value is generally much lower. Uniform treatment of the techniques requires the definition of dissimilarity metric. Letting M represent the maximum value of similarity between any two consecutive frames in the video d(f ,f') can be defined as:

$$d(f,f') = M - s(f,f')$$

### 2.5 Squared Histogram Difference [7]

Compared to the pure histogram difference ,this method tries to amplify the difference of two frames using the following formula:

$$d(f,f') = \sum_{J=0}^{N} (H(f,j) - H(f',j))^2 / H(f,j)$$

The division by $H(f,j)$ serves as a normalization factor for the significance of the square of the difference on the nominator. The preliminary tests showed that a more effective variation of this formula is :

$$d(f,f') = \sum_{J=0}^{N} ((H(f,j) - H(f',j))^2 / max\,(H(f,j), H(f',j)))$$

## 3- Segmentation Based on Moment Invariants [2]

Moment invariants have properties such as invariance to scale change ,rotation and transition ,which make them good candidates as a representation of the frame, the(p+q)th order moments of an image $f(x,y)$ is defined as:

$$m_{pq} = \sum_x \sum_y x^p\ .y^q\ .\ f(x,y)$$

where p, q are integers , p +q = 0,1,2... The infinite set of moments $\{ m_{pq}, p+q = 0,1,2...\}$ uniquely determine $f(x,y)$.

Moment invariant are derived from normalized central moments defined as

$$n_{pq} = 1/m_{00}^{\gamma}\ .\ \sum_x \sum_y (x-x')^p.\ (y-y')^q.\ f(x,y)$$

where $\gamma = 1 + (p+q)/2$,
$x' = m_{10}/m_{00}$ and $y' = m_{01}/m_{00}$

In this study we applied the first three moment invariants. These are defined as:

$$\Phi_1 = n_{20} + n_{02}$$

$$\Phi_2 = (n_{20}-n_{02})^2 + 4n_{11}^2$$

$$\Phi_3 = (n_{30}-3n_{12})^2 + (3n_{21}-n_{03})^2$$

The Euclidean distance can be used as the metric of the difference between two frames

$$d(f,f') = |\, б_f - б_{f'} \,|, \quad where$$
$$б = \{ \Phi_1, \Phi_2, \Phi_3 \}$$

## 4- Segmentation Based on the Range of Pixel-Value Change[1]

This method is models the difference between the values of a pixel in two successive frames as the combination of three factors. The first is a small amplitude additive zero-centered Gaussian noise modeling camera ,tape and digitizer noise. The second is the change of the pixel value resulting from object or camera motion ,change of focus and lighting at a given time in a given time in a given shot The third is the change caused by a cut, wipes dissolve or fade to /from black or white. According to the analytical models for each of the three factors above cuts can be found by looking at the number of pixels whose difference of value in two consecutive frames falls in the range of [128,255] where as dissolve and fade to/from color can be identified by the number of pixels whose change of value between consecutive frames falls in the range of [7,40] for 8-bit coded gray–level image.

Finally wipes are also identified by the number of pixels with value change in the range of [128,255], but their identification is not reliable. Histogram equalization is performed for cut and wipes detection but not for fade detection. The method is not intended for static detection of shot transitions by comparing two frames .Instead it incorporates temporal filtering of the values in the above mentioned ranges over a sequence of consecutive frames.

## 2.5 Segmentation Based on Edge Detection [6] [13]

This method is based on the observation that during a shot transition new intensity edge appear far from the locations of old edges and old edges disappear far from the location of new edges. Therefore shot transition can be identified by comparing the edges in two consecutive frames. The method also uses a registration techniques to compute an overall motion between frames which is taken into account in the computation of the percentage of new edge far away from the old ones. Letting pin denote the percentage of edges pixels in the frame f which are more than affixed distance r from the closest edge pixel in the frame f' and pout the percentage of edge pixels in frame f' which are farther than r away from the closest edge pixel in frame the metric used to measure dissimilarity between two frames is $d(f,f')=max\ (pin, pout)$.

## Comparison of Methods

## 1- Video

Three test video is used in our comparison, include 2 news broadcast and a training video.

## 2- Segmentation by Human Observers

It is necessary to have some independent standard by which to compare the segmentation algorithms. Thus we had a human observer determine where the segment boundaries are .The human-observer segmentation is subjective because the shot transitions are often a ambiguous. For example the news broadcast or the training video, inset of material into the frame ,changes only part of picture but that part changes extensively .Such effect are identified as shot transitions by some observers but not by others. As shown in figure 3 the training video have very different characteristics. The news video are hardest to segment. The most difficult parts of the news broadcast to segment are the commercials .Some of the commercials have very short shots dissolved with one another and sometimes have more than two shot overlapping with each other in a dissolve. The training video was easy to segment while the feature movie was mostly straightforward but some fades were extremely long. Table 1 presents the output of the human segmentation which is the identification of shot transition and their classification as either cuts or other effects(e.g. fades, dissolves and wipes).

## 3- Setting Thresholds for Automatic Segmentation

For all of the methods except those presented in sections 2.4 and 2.5, we adopted the following automatic shot-transitions identification method. A threshold is specified and the values of the similarity metric are compared against it. Whenever the values cross the threshold from below to above a shot transition is identified. The methods in sections 2.4 and 2.5 use prespecified parameters to determine whether a shot transition has occurred.

## 4- Filtering Algorithm

Initial tests of the segmentation methods showed that many of the shot transitions triggered spurious shot transition identifications. One of the most common cases of spurious shot transition was caused by the threshold being crossed multiple times for a single transition due to variations of the value of the metric. Although little can be done to identify shot transitions which were not found due to producing small values for the metric of dissimilarity between consecutive frames, a lot can be done to reduce false shot transition identifications.

A simple filter is design to process the sequences of values of the dissimilarity for the different methods and produces anew sequence with the following procedure.

The value is compared with the k previous values and the k following values. If any of these 2k values is greater than the current value replace the current value with in the last local minimum detected in the sequences of metric values.

Two conflicting requirements govern the choice of k . Shot transitions that take place over a sequences of frames (e.g. fades ,dissolves ,wipes) can cause big oscillations in the sequences of values of the metric of dissimilarity resulting in several crossings of the threshold for the same transitions effects. The value of 2k should ideally be larger than the longest of those transition effect. On the other hand there exist shots whose duration is shorter than the longest transition effect. The ability to identify them ,required that k is smaller than the duration of the shortest shot. A reasonable value of k can be obtained from examination of figure 3 where the cumulative histogram of the duration of shots and the duration of fades is shown for the two video. Based on this figure and the fact that the abc news video was digitized at 15 frames per second k has been set to 5 .Notice that the video have very different values for the shot duration .The choice of k for the filter is selected according to the statistics of the news video because a larger k suggested by the statistics of the training video would filter out real shot transition that are close to each other.

An alternative to the filter proposed in this section ,would be a moving average window ,with the window size specified with reference to figure 3. The moving average technique is used can shift the peak values of the metric introducing some in accuracy in the detection of the shot transition instant. Furthermore , the problem of multiple crossing of the threshold for the same transition in alleviated but not completely gone ,because for small window sizes as those suggest by figure 3 the variations in the value of the metric may still exit.

Another promising technique trying to overcome the false identifications related to a global threshold ,is local thresholding .In local thresholding the value of the threshold changes over time according to the characteristics of the metric within an observation window .The problem of multiple crossing of the threshold can easily be overcome by setting the local threshold to an appropriate value beyond the local average of the metric values .Both global and local thresholding can easily be coupled with simple filters to suppress the identification of spurious transitions .

## 5- Comparison of Automatic Methods

This section compares the performance of the video segmentation techniques for the abc news video which was chosen because it contains a variety of shot transition effects such as cuts, fades ,dissolves, and wipes .It also contain s advertisements which are generally hard to automatically segment due to the short duration of the shots which are dissolved with each other and also because there is often fast object or camera motion . Finally some parts of the video come old black and white films which are very noisy.

The filter has been applied to all of the methods where the segmentation is based on the use of a global threshold. Figure 4 and table 2 shows the percentages of correctly identified shot transition and the percentage of the false identification for each of the methods. The horizontal axis is the value of the threshold used as the parameter to decides where shot transition occurs. It ranges from 0 to 1 with 1 corresponding to the highest value found in the metric .The vertical axis shows the percentage of correctly or falsely identified shot transitions relative to the real number of shot transition which corresponds to 100%.

**Table 2: correct and false identification for abc news video**

| Method | % correct Identification | %False Identification |
|---|---|---|
| Range of Pixel-Value Changes | 95 | 68 |
| Edge Detection | 92 | 59 |

While separate histogram could have been computed for each color. we focus on red histogram because they are representative of, or perhaps slightly better than, the other colors[11] .Notice the very high percentage of false identification for low thresholds ,even after the application of filtering algorithm. Table 3 compare the different methods in terms of the percentage of correct and false identification for the threshold that maximize the percentage of correctly identified shot transitions. It can be seen that several method identify correctly 94% to 95% of the real transitions.

**Table 3: Maximum % of correct identifications and corresponding % of false identifications for ABC news video**

| Methods | %Correct Identification | %False Identification |
|---|---|---|
| Absolute Frame Difference | 73 | 56 |
| Red Histogram Difference | 94 | 175 |
| Weighted Histogram Difference | 94 | 135 |
| X^2Red Histogram Difference | 95 | 137 |
| Red Histogram Intersection | 94 | 174 |
| Pure Moment Invariants | 54 | 105 |
| Range of Pixel-Value Change | 95 | 68 |
| Edge Detection | 92 | 59 |

The corresponding high percentage of false identification are not very discouraging since incorrect identification can easily be filtered out by the human who annotates the important shot transition and moreover this percentage reduce much faster than the percentage of correct identification as the value of the threshold is increased (for example for the red histogram difference method in Figure 4 the percentage of correct and false identification for the first point to the right of the point where the percentage of correct identification is maximum are 93% and 88% compared to 94% and 175% for its adjacent point).
The 68% false identification for the range of pixel value change method appear unusually low and indeed we did not find such good performance with algorithm on other video even though the percentage of false identification for this method was always among the lowest the result for the edge detection method have been obtain using the code we were granted by the author of 6 the code was still under development so better result may be expected with newer version.
Figure 5 and table 4 discriminates the correctly identified shot transitions between cuts and other shot transition effects (including fades, dissolves, and wipes).Although according to most of the metrics ,fades, dissolves and wipes should be much harder to detect because the changes happen gradually it turns out that most of the methods do comparably well for cut and fade detection with the exception of the absolute frame difference method which performs very well for cuts but very poorly for other effects. Notice also that the threshold that maximizes the correctly identified cuts is different from the one for fades ,dissolves and wipes with the later being lower as expected thus in practice we believe that separate thresholds should be employed for the two types of effects .

**Table 4:Correctly identified cuts and other shot –transition effects for abc news video**

| Method | %Correct identification of cuts | %Correct identification of other effects |
|---|---|---|
| Range of pixel-value changes | 97 | 91 |
| Edge detection | 92 | 91 |

The evaluation of the performance of the segmentation algorithms depends on several factors .the first of these is how the segmentation is going to be used. An application involving annotation by a human observers would care more for a high percentage of correct identification than for a low percentage of incorrect identification .When a human does not intervene ,a low percentage of incorrect identification becomes Increasingly important. A second factor for evaluating to the result obtain in an experimental stage where the best threshold is specified for several videos. Therefore as should be clear by looking at Figure 4 the sensitivity to the threshold parameter is an important criterion .A final factor is the type of video to be segmented also influences the choice of algorithm.

## 6- Variation of Segmentation Performance for Different Videos

This section compares the performance of each segmentation method for different video material. Figure 6 and table 6 compares the performance of each method for two of the videos the horizontals axis is the threshold value to the maximum value of the method for a particular video. The vertical axis is the percentage of correctly identified shot transition.
Figure 6 and table 6 clearly shows that the *abc* news video is the hardest to the segment. The percentage of correct identification is not only lower than in the rest of the videos but it is also very sensitive to the threshold parameter implying that the threshold parameter does not transfer well across different types of video. For the difference is performance, consider the absolute frame difference method which performs very well for the training videos but not for the news broadcast. The difference in the threshold that maximizes the percentage of correct identification and also the different sensitivity to the value of this threshold can be seen in the Weighted Histogram Difference method. The methods that depend on the Range of Pixel-Value Changes and on the Edge Detection seem to perform uniformly well across the different types of videos.

## 7- Automatically Setting Threshold Parameters

Section 3.6 clearly shows that the threshold where the percentage of correct identification is maximized (optimal threshold) is video dependent. The values for the optimal threshold are depicted in the Table 5 for each of the methods. The third column in the Table 5 is the average of the first two column and tests the value of previously determined threshold for segmentation anew video .In order to appreciate the suitability of these values as prior set threshold and test whether the thresholds would be more stable within a type of video ,we applied the automatic segmentation methods on the lbc news video . The results presented in Table 7 show that average optimal thresholds from the three column of Table 5 give results which are close to optimal for the lbc news video .

## 8- Reasons for Misses and False Positives

Misses of shot transitions and false positives were examined in order to get a better understanding of the reasons why the methods fail to produce the desired results. The most common causes of misses are the following :

- Wipes are quite difficult to identify .None of the segmentation methods is able to distinguish clearly between a wipe and a sequences of images where objects are moving around quickly.
- The value of the specific metric is sometimes lower for a shot transition than for anon-shot transition. This is particularly true for cuts where the frames of the two shots are very close according to the metric. It also occurs for slow fades.

The most common causes of false positive are the following:

- High values of the dissimilarity metric for rapidly moving images.
- Sudden variations of the luminance of the image ,either due to lightning effects(e.g. lighting )or to poor video quality.
- Fade out and then fade in (e.g. ,between advertisements) may be identified as two shot transitions although there is really only one shot transition.
- Variation s in the value of the metric of dissimilarity between consecutive frames in the case of fades ,dissolves ,and wipes ,cause the upward crossing of the threshold more than once for a single shot transition.

Local threshold techniques can help to reduce misses by small values for the metric of dissimilarity between consecutive frames .False shot transition identification can be greatly reduced with the filtering technique or averaging of the value of the metric within a sliding window.

## Computational Resources

When real –time segmentation of videos is required an estimate of the number of operations to evaluate the measure of dissimilarity between two frames is required Table 8 presents a rough estimate of the required computations per frames under the assumption that addition subtraction and multiplication required time equivalent to one operation whereas divisions take approximately four times more .Implementation issues such as assignment of variables to registers use of pointer or arrays ,memory access time and others ,are ignored. The variable N is the number of levels (bins)of the pixel value and P is the number of pixel per frame In both the absolute frame difference and the *Range of Pixel-Value Changes* methods, we assume that grey –scale pixel values are provided by the video encoding. No estimate for the requirements of the *Edge Detection* algorithm was computed, because its slow and still under development.

## Discussion

Because all of the methods studied here have high false –identification rates they should be thought of as providing suggestions to human observers and not as an ultimate standard of performance . For instance, they could provide input to an interactive interface for specifying the structure of a video. The high percentage of correct identification of these methods implies that for applications involving human interaction computationally more expensive methods such as those involving object recognition could be avoided.

The choice of the best segmentation method is not straightforward. The time requirements the sensitivity to the threshold parameter the percentage of correct and false identification and the type of video to be segmented should by taken into account .Also for time-critical application the performance of each method when spatial (within a frame)or temporal(across consecutive frames) under sampling is used, should be studied .It should be noted that some methods are robust and can even improve by working at low resolution. Finally the type of the application segmentation is needed for should also be considered For example an application for retrieval of shot based of the video similarity could take advantage of the histogram information that can be stored during preprocessing of the video for representative frame of each shot to perform very fast retrieval of related shots (e.g. for the red histogram difference method each such comparison would require approximately 2N computations).

Saving method-dependent information during preprocessing for the other methods can also help but they are still going to be much slower than histogram-based methods. Also methods like the Range of pixel-value changes that was not intended to be applied for static detection between two frames would be slow to use in such an application since its decision is based on the temporal characteristics of some statistical measure over a sequence of consecutive frames.

There are still problem that the methods described above fail to address .One problem is how to distinguish between a fast change of the image in the same shot caused by movement of the camera (e.g. filming from within a moving car )and a cut or dissolve .Methods based on only simple statistical characteristics of the frame sequences might not be sufficient in this case and some more involved processing like object recognition might be necessary.

Another important problem is the identification of semantic transition not associated with shot transition .This situation is very common in news programs ,where in switching between topics, only a small portion of the frame change(usually square in the background of the anchor)and this would normally not result in a transition identification.

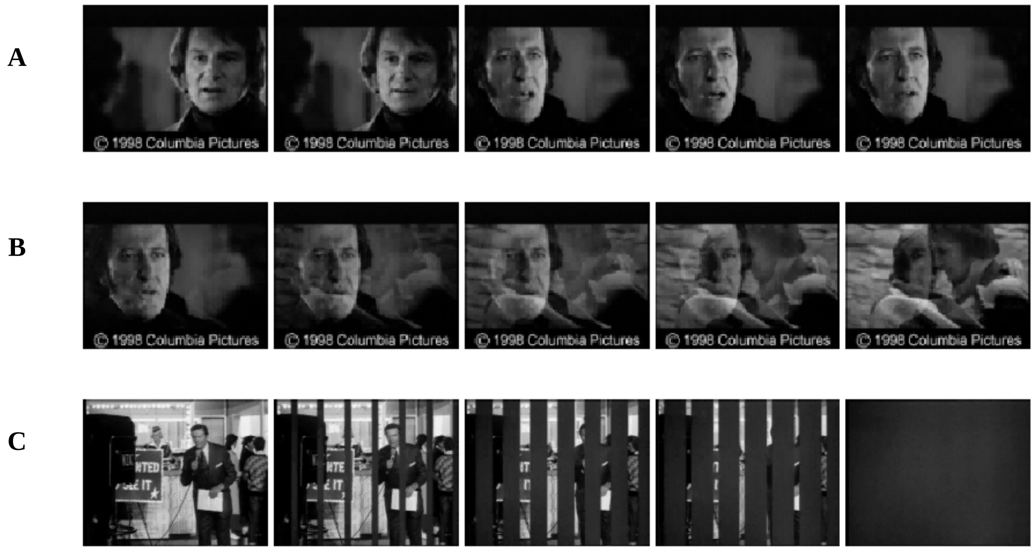


**Figure 2 : Shot transitions :A shows a cut: B shows a dissolve :C shows a wipe**

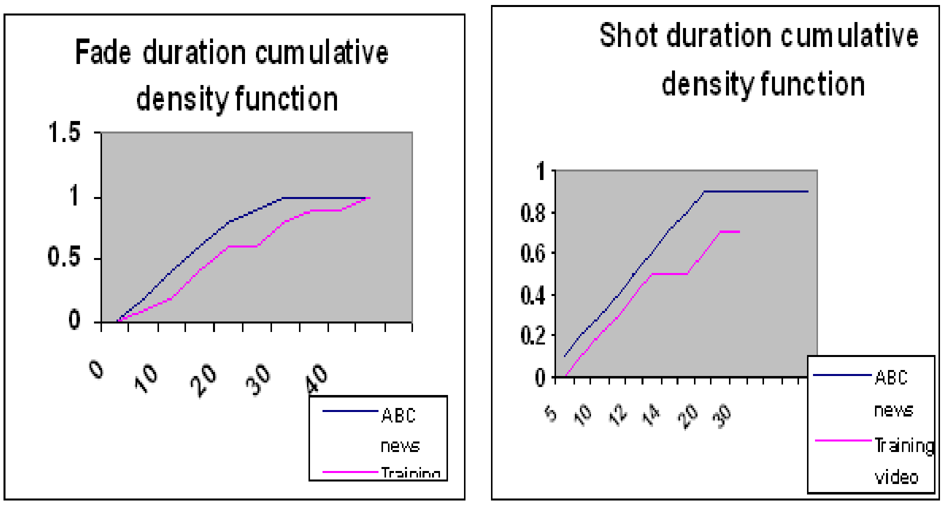


**Figure 3: Shot and fade duration**

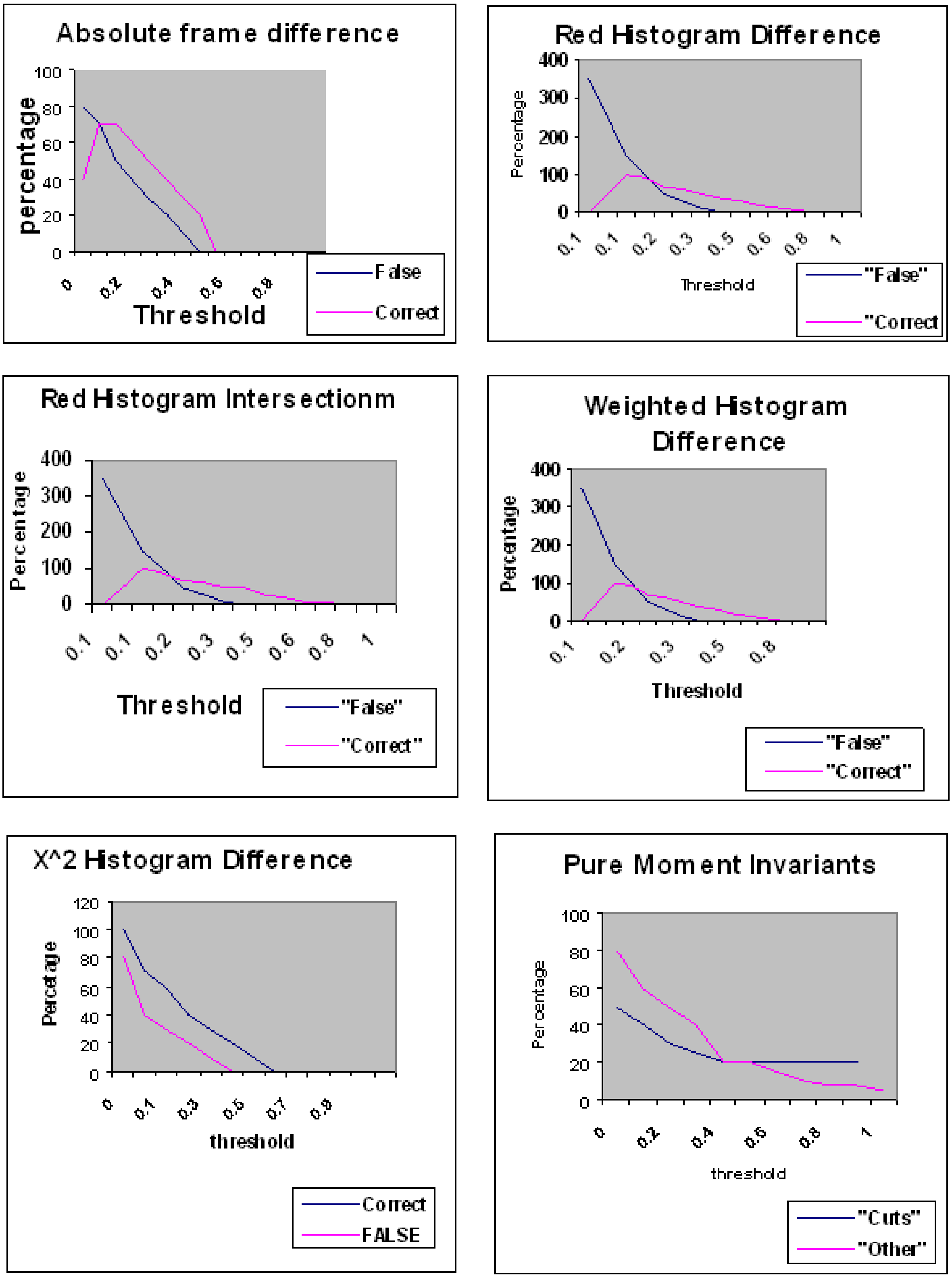


**Figure 4: correct and false identification for abc news video**

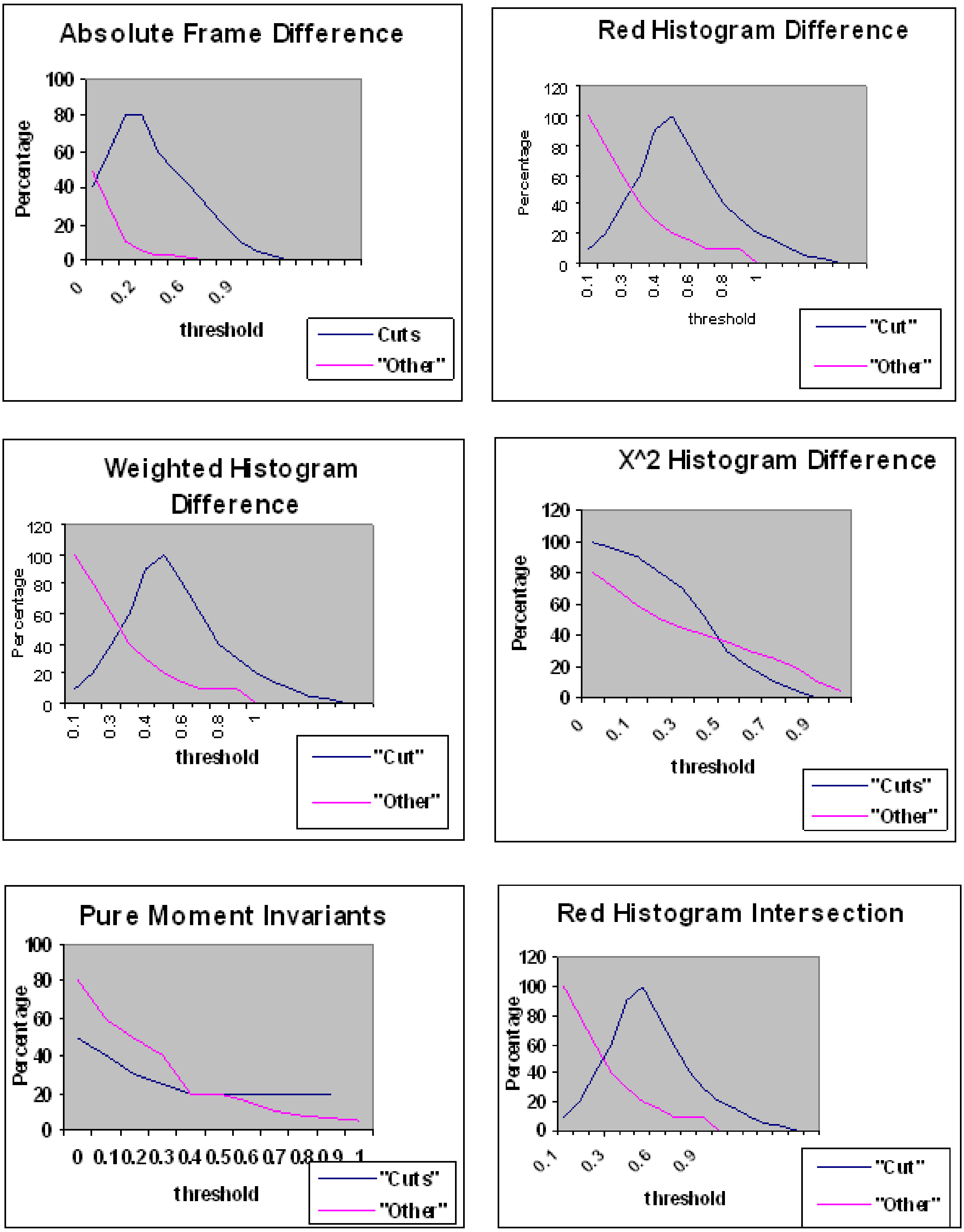


**Figure 5:Correctly identified cuts and other shot –transition effects for abc news video**

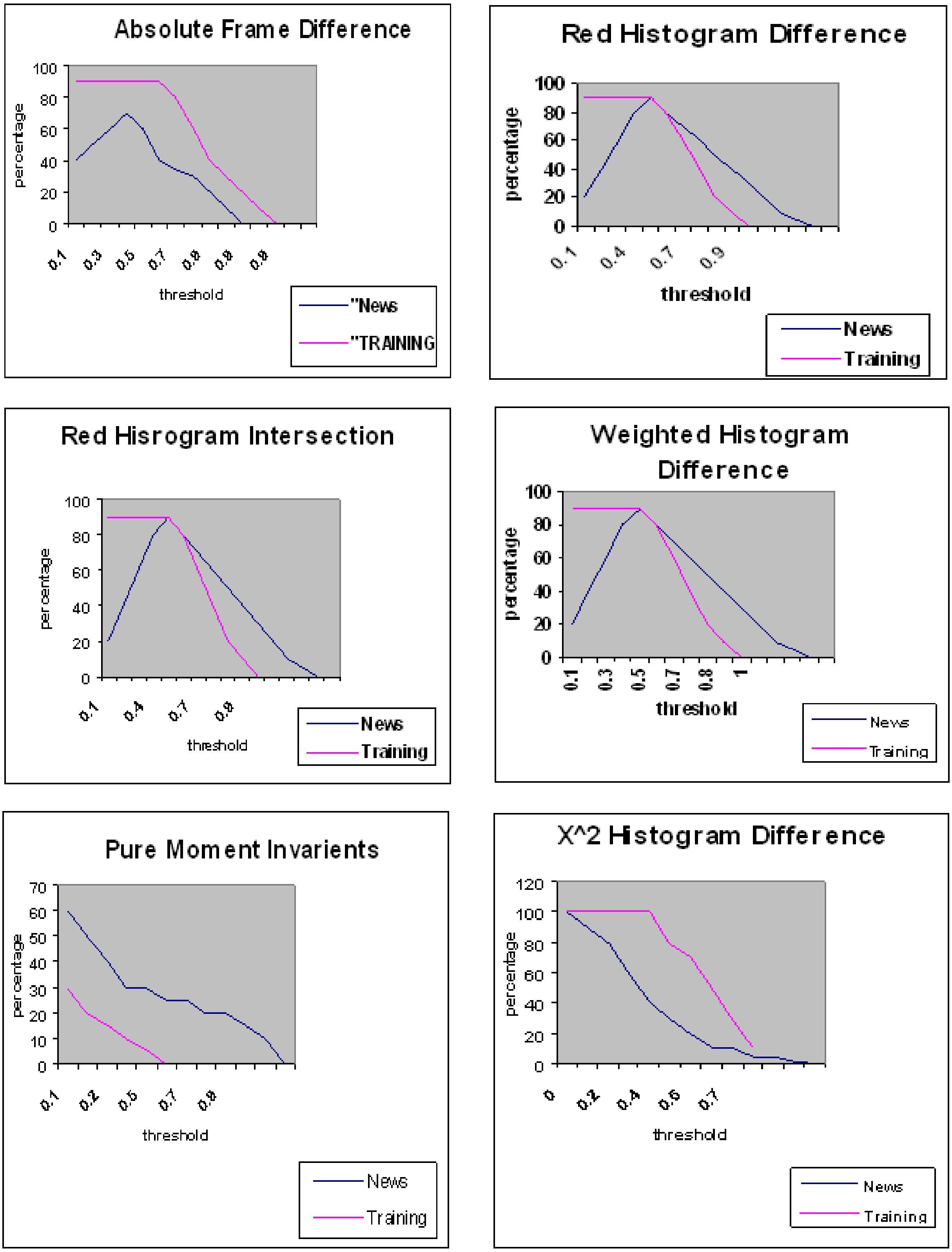


**Figure 6: Correct shot –transition identifications across videos**

**Table 1: Characteristics of the videos**

| Video | digitization rate (frame per second) | duration(min) | cuts | fades, dissolve and wipes |
|---|---|---|---|---|
| ABC news | 15 | 28.5 | 353 | 160 |
| LBC news | 12 | 30 | 366 | 115 |
| Training vid | 12 | 60 | 147 | 7 |

**Table 5: Optimal thresholds for two videos and the average optimal threshold**

| Method | ABC news | Training video | Average optimal threshold |
|---|---|---|---|
| Absolute Frame Difference | 0.10 | 0.10-0.17 | 0.15 |
| Red Histogram Difference | 0.10 | 0.17-0.23 | 0.14 |
| Weighted Histogram difference | 0.10 | 0.17 | 0.15 |
| X^2 Red Histogram Difference | 0.03 | 0.07 | 0.05 |
| Red Histogram Intersection | 0.10 | 0.13-0.20 | 0.14 |
| Pure Moment Invariants | 0.03 | 0.03 | 0.03 |
| Range of pixel-value Changes | Independent | Independent | Independent |
| Edge detection | Independent | Independent | Independent |

**Table 6:Correct shot –transition identifications across videos**

| Method | ABC News | Training video |
|---|---|---|
| Rang of pixel-value changes | 95 | 96 |
| Edge detection | 92 | 96 |

**Table 7: Test of various threshold values for lbc news video**

| Method | LBC news optimal threshold | %Correct identifications for optimal threshold | %Correct identifications for average optimal threshold | %Correct identifications for ABC news optimal threshold |
|---|---|---|---|---|
| Absolute Frame Difference | 0.10 | 81 | 79 | 81 |
| Red Histogram Difference | 0.13-0.17 | 92 | 92 | 90 |
| Weighted Histogram difference | 0.13 | 92 | 91 | 89 |
| X^2 Red Histogram Difference | 0.07 | 92 | 91 | 90 |
| Red Histogram Intersection | 0.13-0.17 | 92 | 92 | 90 |
| Pure Moment Invariants | 0.03 | 60 | 60 | 90 |

## References


1. Aigrain P. and Joly P., **1994**, "*The automatic real-time analysis of film editing and transition effects and its applications*", Computer& Graphics, Vol . 18,No,1,pp. 93-103,.
2. Arman F., Depommier R., Hsu A., and Chiu M-Y., (Nov.**1994**) "*Content- based browsing of video sequences*",Proceedings of ACM Multimedia San Francisco pp.97-103.
3. England P. Allen R.B., Dailianas A., Bianchi M., Sullivan M. and Heybey A., (Jan.**1996**) "*The Video Library Toolkit - A system for Indexing and Browsing Digital Video libraries*" Proceedings SPIE Photonics Electronic Imaging Science and Technology96, Storage and Retrieval for Image and Video Database IV, san Jose.
4. Hampapur A., Jan R., and Weymouth T., (Nov.**1994**) "*Digital Video Segmentation*", proceedings ACM Multimedia, San Francisco pp.357-364.
5. Jain A .K., **1989** "*Fundamentals of Digital Image Processing*, prentice Hall.
6. Mai K ., Miller J., and Zabih R., (Nov.**1995**) "*A Robust Method for Detecting Cuts and Dissolve in Video Sequences* " proceedings of ACM Multimedia.
7. Nagasaka A. and Tanaka Y. **1992**, "*Automatic Video Indexing and Full-Video Search for Object Appearances*", Visual Database System II ,Elsevier Science Publisher (North-Holland),.
8. Otsuji K. and Tonomura Y., (Aug **1993**) "*Projection Detecting Filter for Video and Cut Detection,* "Proceedings of ACM Multimedia Anaheim, pp.251-257.
9. Smoliar S . and Zhang II., **1995** " *Content-Based Video Indexing and Retrieval* ", IEEE Multimedia ,Vol.2, pp62-72.
10. Swain M. and Ballard D., **1991** "*Color indexing* "International Journal of Computer Vision, Vol.7, pp.11-32.
11. Yeung ,B.L Yeo, Wolf W., and liu B., **1995** "*Video Browsing using Clustering and Scene Transitions o Compressed Sequences*, "IS&T /SPIS Multimedia Computing and networking.
12. Zhang H..S. Tan Y., Smoliar S.,and Yihong G., **1995** "*Automatic Parsing and Indexing of News Video*" ,Multimedia Systems ,. vol2,pp256-266.
13. Zabih R., Miller J., and Mai K.. **1995** "*A Feature-Based Algorithm for Detecting and Classifying Scene Breaks*", Technical report cs-tr- 951530,computer science department,university Cornell.
14. Lawrence S., Ziou D., and Wang S.. **2001** "*Motion Insensitive Detection of Cut and Gradual Transitions in Digital Video*", international conference on multimedia, modeling,Ottawa.