Key Frame Estimation in Video Using Randomness Measure of Feature Point Pattern

Dipti Prasad Mukherjee, *Senior Member, IEEE*, Sitansu Kumar Das, and Subhra Saha

***Abstract*—In this paper, a generalized statistical tool is introduced to estimate key frames in a video sequence. The tool works based on the inter-relationship between different features of image frames in a video. The image feature vectors are plotted in feature space as points and a randomness measure is determined from the distribution of these points. The randomness measure of the feature vectors is defined with respect to simulated random point patterns and expressed as a probability value of a frame being a key frame. Since, depending on the video content more than one inter-relationship of features can be used to determine a single key frame, different probability values are derived to support a frame as a key frame. To integrate these probability values a *combiner* model is designed to uniquely decide the status of a key frame. The combiner model is based on the Dempster–Shafer theory of evidence. To demonstrate the idea, randomness measures, and consequently the probabilities of a frame to be a key frame, are obtained separately from spatial domain and frequency domain features. The combined probability value enhances the confidence in selecting a frame as a key frame. The result is tested on a number of standard video sequences and it outperforms the related approach.**

***Index Terms*—Dempster–Shafer (DS) theory of evidence, key frame, spatial randomness measure.**

I. Introduction

THE estimation of key frame in a video sequence is an important research topic for various reasons. The content in key frames can be used for video indexing for applications like content-based video retrieval. Key frames are also used in designing MPEG based codecs and streaming media applications. In this paper we present a key frame estimation technique.

As the name suggests, key frames in a video sequence are those frames where there are *significant* changes in the image content. Therefore, methodologies developed for key frame estimation have attempted to capture the significant changes between consecutive frames [2], [6], [8]. Naturally, the techniques for key frame estimation focus on defining image features that can estimate significant changes in between-frame image contents. For example, while one methodology utilizes optic flow [4] to estimate change in motion energy between frames, another relies on an ensemble of features like the change in mean and variance of a local image neighborhood, motion vectors etc. [5]. However, most of the existing techniques for key frame estimation are application specific and select features suitable for that particular application, for example, detecting key frames in videos of technical presentation [2] or for a surveillance system [8]. In contrast, our focus is to develop a generalized application-independent tool for key frame estimation, where a statistical technique is used suitable for a wide range of image features and a variety of applications. Some of the relevant existing techniques for key frame estimation are discussed and our contribution is highlighted in the next section.

For the proposed technique, inter-relationship between different image features, which estimate intra-frame or inter-frame significant changes, is visualized as point patterns distributed in the feature space. We have quantified the measure of significant change of image content through a measure of *randomness* of these point pattern in the feature space.

The spatial distribution of feature vectors is tested against a set of simulated feature population generated randomly [1]. In case the distribution of feature vectors agrees with that of the simulated random points, we conclude that the relationship between features is random. The randomness of every image frame is quantified after calculating the extent of the deviation of the feature vector distribution from the simulated random point patterns. Naturally, the higher the randomness, the higher is the chance of the frame to be a key frame.

For a robust inference of a randomness measure, a number of inter-related feature modules are used to generate the randomness measures. It is as if a number of experts are looking at the video and different experts are assigning different randomness measures to a potential key frame. Naturally, there exists a need to design an infrastructure to unify these different randomness measures. In order to do that, we have used the Dempster–Shafer (DS) theory of evidence [3]. The DS theory combines different randomness measures into one confidence value of randomness based on which key frame is detected.

The primary contribution of this work is in relating the randomness measure of the spatial point pattern with the key frame estimation problem. The quantified randomness measure is expressed as a probability value that indicates the potential of a video frame as key frame. The second contribution is in the use of DS theory in unifying randomness measures or key frame probability values derived from different feature vectors. Note that unlike Bayesian techniques, no *a priori* knowledge or model of the key frame is used in the decision fusion process.

In the next section, we discuss some of the related works on key frame estimation that motivates us to develop an application-independent generalized framework for key frame estimation. In Section III, we introduce the model to measure randomness that quantifies the significant change in the content of video

frames. In Section IV, the randomness measures, derived from different sets of inter-related features, are integrated using DS theory. Results and discussions of the proposed approach are given in Section V followed by conclusions.

II. Related Works

Existing techniques for key frame estimation exploit different image features to measure changes in the content of an image frame in a video. The most common feature is to quantify the difference of image frame information and this is utilized by Stringa and Regazzoni [8] to detect key frames in a movie sequence for a surveillance system. They have determined key frame on the basis of the change of feature values in some regions of interest in consecutive video frames. The pixels in the frame difference image are used as features. If the changes in features remain *almost* unaltered for a certain period of time then the initial frame where the feature values have changed is taken as key frame. In a related context, Ju *et.al* [2] have designed a key frame detection system for movie sequences showing technical presentations. Detected key frames are the *unique slides* shown in the presentation whose contents do not vary significantly over a certain time period barring few minor changes. The authors assume an affine transformation model for unskewing between-frame changes. The key frames are detected after combining heuristics related to presentation technology and comparing changes in unskewed frames.

From a different perspective, changes in video content can be analyzed by extracting the between-frame motion information of significant image segments. Wolf has estimated the key frame by analyzing between-frame motion information using optical flow analysis of the image sequence [10]. The sum of magnitudes of the optical flow components of each pixel is computed for each frame and plotted in a graph against their respective frame number. Key frames correspond to the local minima points of optic flow magnitude graph having neighboring maxima on both sides.

Optic flows are also used by Lui *et.al* [6] to detect key frames in a long video sequence. The average magnitude of the motion vector is multiplied with the dominant motion direction obtained from the optic flow to estimate the perceived motion energy (PME), which is taken as the main feature for key frame detection. The frame from which PME starts to accelerate, or the frame from which PME starts deceleration, is taken to be a key frame.

The scheme in [4] equally subdivides the total span of a video into subshots such that the between-frame feature variations within subshots become minimum. As per definition in [4], a key frame has the least feature variation with respect to features of other frames within a subshot. Optical motion, mean, and variance of pixels in a local neighborhood are taken as features for comparison of frames within a subshot. In a partial modification to this approach, an iterative scheme is implemented where the number and length of the subshots are decided iteratively to make feature variation within a subshot minimal [5].

Overall, key frame estimation is implicitly motivated by the segmentation of between-frame information. The features of this segmentation and related motion information govern the determination of key frames. In [8], the values of the pixels in the areas of interest are compared against a threshold. In [2], [6], and [10], consistency of the optical motion is checked whereas in [4] and [5], feature variation within a video shot is used for key frame estimation. The application domains of [2] and [8] are very focused whereas treatments in [4]–[6] are more generalized. In the proposed approach, our focus is to exploit the ensemble of features and their relations so that a generalized tool for key frame estimation can be developed. The key question is how this generalization could be achieved. The use of a spatial randomness measure [1] of feature point pattern provides a generalized tool to quantify the randomness of an image content. To the best of our knowledge this is the first attempt where the randomness measure is used in video content analysis. Since our approach of assigning a frame as key frame depends on the randomness measure of the frame, an important question is how to set a threshold value for the randomness measure to mark a frame as a key frame. We have used the concept of DS theory to answer this problem by combining the decision making process of several randomness measure modules. The design of our system is such that it is capable of accepting and unifying different heterogeneous feature modules compared to making a decision based on any specific set of features or using an application-specific threshold. Before we explain the randomness measure unification in Section IV, we present our approach of deriving randomness measure from image features.

III. Randomness Measure

As noted in the introduction, we derive a set of image features either from a single video frame or by comparing a pair of temporally apart video frames. The relationship between the image features is visualized as point pattern in the feature space. Randomness of this point pattern is measured as described next.

A. Model of Randomness Measure

Let us present the concept assuming two image features f^1 and f^2. Assume a 2-D feature space where orthogonal axes are represented using f^1 and f^2. The spatial distribution of feature vectors is given by $\left(f^1_{(p,q)}, f^2_{(p,q)}\right)$ where $(p, q) \subset Z^2$ is an image point, $0 < p < r$ and $0 < q < c$ for an image I having r rows and c columns. Naturally, $\left(f^1_{(p,q)}, f^2_{(p,q)}\right)$ is a point in 2-D feature space. For every feature point, its distance from rest of the points in the feature space is compared against a set of preset distances. Let y_i be the nearest neighbor distance of a feature point from the ith point of the remaining population and the feature population at a distance less or equal to a preset distance y from the ith point is given by $\#(y_i \leq y)$. For n such feature points the total feature population within a preset distance y is given by $\sum_n \#(y_i \leq y)$. Given this, the spatial distribution of the feature points is defined as

$$G(y) = n^{-1} \sum_n \#(y_i < y). \tag{1}$$

The feature space is normalized to a unit square region. Hence, the distance y is incremented by a preset value from 0 to $\sqrt{2}$ (which is the length of the diagonal of a unit square). Note that y_i includes duplicate distances as every feature point pair is considered twice in calculating neighborhood distances.

Equation (1) finds the ratio of number of points within a given neighborhood of feature points with the total number of feature points n. Therefore, $\hat{G}(y)$ is m-length vector where y is incremented m times between 0 and $\sqrt{2}$. The vector $\hat{G}(y)$ derived from image data is known as empirical distribution function [1]. The extension of the analysis based on (1) to more than two features is straightforward. The algorithm below can illustrate the above process.

step $= \sqrt{2}/m$; $y = 0$;

for $i = 1$ to m

$y = y +$ step;

$G(i) = 0$;

for $j = 1$ to n

$G(i) =$ # (points within the circle of radius y centred at point j)+ $G(i)$;

end

$G(i) \quad G(i)/n$;

end

$\hat{G}(y) : \{G(i), i \quad 1 \ldots m\}$

As mentioned earlier, we need to correlate the empirical distribution function $\hat{G}(y)$ against a set of simulated feature points generated randomly. The spatial distribution of these simulated random points can also be analyzed in the same way as (1) based on inter-point nearest neighbor distance measure. The distribution of random point pattern obviously depends on the total number of points n within a given feature space (in this case a unit square or hypercube).

Given the area of feature space as $|A|$, the distribution of n arbitrary simulated points within a distance y (or within a circle of area πy^2) of a specified point is modeled as [1]

$$G_m(y) = 1 - (1 - \pi y^2 \ A|^{-1})^{n-1}. \tag{2}$$

The occurrence of points is assumed independent. As in (1), the distance y also ranges between 0 to $\sqrt{2}$. Assuming n to be fairly large and $\lambda = n\,|A|^{-}$, the model of (2) can be redrafted as [1]

$$G_m(y) \approx 1 \quad \exp(\ \lambda \pi y^2). \tag{3}$$

Intuitively, as the preset distance y is increased in the model (3), more and more points in the feature space are being included within the distance and $G_m(y)$ approximates towards unity. Also, as number of points n is increased in a given feature space A, the value of λ increases, and for a given preset distance y, $G_m(y)$ should saturate towards unity. This scenario is well captured by the model shown in (3). For key frame estimation, we correlate empirical distribution $\hat{G}(y)$ against $G_m(y)$ for randomness measure of the feature points. If $G_m(y)$ represents an ideal random population, the question is whether $\hat{G}(y)$ is compatible with $G_m(y)$ or not. This issue is discussed next.

B. Test of Randomness

While populating $G_m(y)$ following (3), a number of simulations are executed for each of the neighborhood distance specified by y. Since n numbers of feature vectors are derived from the image data, each simulation generates n random populations distributed in a unit square (assuming feature space is 2D). If there are s numbers of simulations, then distribution function for simulated data points is expressed as the mean of $G_j(y)$, $j \quad 1, 2, 3, \ldots, s$

$$\overline{G}(y) \quad s^{-1} \sum_{j=1}^{s} G_j(y). \tag{4}$$

$G_j(y)$ is calculated following (1) but for the random population. Again the range of y is fixed between 0 to $\sqrt{2}$.

With this observation what we get is two sets of distribution: m-length $\hat{G}(y)$ derived purely from image feature vectors and m-length $\overline{G}(y)$ derived from s number of random point sets closely approximating (3). To test the randomness, we need to correlate $\hat{G}(y)$ and $\overline{G}(y)$. The higher correlation between $\hat{G}(y)$ and $\overline{G}(y)$ indicates that feature calculated from a particular video frame or from a pair of frames are spatially random in feature space.

For each of m neighborhood distances of y for the random points, we can evaluate

$$U(y) = \max\{G_j(y)\} \tag{5a}$$

$$L(y) = \min\{G_j(y)\} \quad \text{for} \quad j = 1, 2, 3, \ldots, s. \tag{5b}$$

The parameters $U(y)$ and $L(y)$ are upper and lower envelops of $\overline{G}(y)$, respectively, derived from the simulated data. For a particular y, if $\hat{G}(y)$ is either less than $L(y)$ or higher than $U(y)$, $\hat{G}(y)$ is considered significantly away from the $\overline{G}(y)$ and not considered random. This may be inferred as the feature vectors having *clustering tendency*. In case $\hat{G}(y)$ is within the envelopes that is having close correlation with $\overline{G}(y)$, the image features are randomly distributed, which we need to quantify as discussed in the next section.

C. Quantification of Randomness Measure

As noted in Section III-B we are correlating values of $\hat{G}(y)$ and $\overline{G}(y)$ at different preset distances. If their values agree (that is closer to each other) at different preset distances, we assume that the feature point distribution generating $\hat{G}(y)$ is random. Note that our objective is not to find an overall relation between $\hat{G}(y)$ and $\overline{G}(y)$ but to look at their values at discrete intervals of y. The objective in this section is to quantify the agreement or closeness of values between $\hat{G}(y)$ and $\overline{G}(y)$. Of course, as noted in Section III-B, $\hat{G}(y)$ values outside the upper and lower envelopes are not at all considered as random.

The measure of departure $u(1)$ of $\hat{G}(y)$ from the mean spatial randomness characterized by $\overline{G}(y)$ can be expressed as [1]

$$u(1) = \int \{\hat{G}(y) - \overline{G}(y)\}^2 dy. \tag{6}$$

Extending (6), the measure of departure u can be recalculated for the entire simulated feature points generated for $\overline{G}_y$

$$u(j) = \int \{G_j(y) \quad \overline{G}_j(y)\}^2 dy \text{ for } j = 1, 2, 3, \ldots, s. \quad (7)$$

The modified mean of distribution of random population $\overline{G}_j(y)$ is calculated excluding the data from jth simulation: $\overline{G}_j(y) = (s \quad 1)^{-1} \sum_{i=1, i \neq j}^{s} G_i(y)$. The measure of departure $u(j)$ gives the degree of departure of jth simulation from the mean of all the other simulations (that is barring jth simulation). To determine the extent of randomness in the image features, we would like to investigate the rank r of $u(1)$ in all the $u(j)$ values.

To calculate r, unique $u(j)$ values are sorted in ascending order. The position of $u(j)$ closest to $u(1)$ magnitude is the rank of $u(1)$. The rank r of $u(1)$ is expressed as the probability of the image features in kth frame being random

$$p(k) = \frac{(s_u - r)}{s_u}. \quad (8)$$

If the image data is derived from kth frame in the video sequence[1], $p(k)$ is the probability that the kth frame is random given that s_u numbers of unique $u(j)$ values are obtained in total s number of simulations.

Before we explain how $p(k)$ can be utilized to decide key frame let us take an example to understand how inter-relationship of features can be used to measure randomness. Ideally, the use of both spatial and frequency domain features are a better choice to measure the significant change in the content of image frames. In the following sections, this is analyzed.

D. Estimation of Randomness Using Spatial Domain Feature

Let us take the example of billiards video sequence. Fig. 1(a)–(c) are the first, third, and the tenth frames of the sequence, respectively. Clearly, there is significant change in image content between Fig. 1(a) and (c) compared to almost no change between Fig. 1(a) and (b). Fig. 1(d) is the absolute temporal difference of the frames of Fig. 1(a) and (b). The spatial domain features, average intensity and busyness of intensity defined within a 3×3 image mask, are evaluated on the frame difference data of Fig. 1(d). The busyness of intensity is the absolute sum of consecutive pixel differences within 3×3 image mask. The consecutive pixel differences are calculated both along horizontal and vertical directions [7]. The frame matrix is of size 73×110 pixels and it generates 7668 number of 2-D feature vectors after ignoring border rows and columns. These image features are plotted as point patterns as shown in Fig. 1(f) with x and y axes representing average intensity and busyness, respectively. The randomness of the frame is estimated from the distribution of these point patterns.

Following (1), $\hat{G}(y)$ is calculated after varying y between 0 to $\sqrt{2}$ taking $m \quad 100$. Again, following Section III-B, after 100 simulations to generate random point pattern, $\overline{G}(y)$ is evaluated and plotted (along y axis) as broken-line curve against

[1]In case k and $(k+1)$th frames are used to derive feature vectors, $p(k)$ indicates randomness of $(k-1)$th frame.

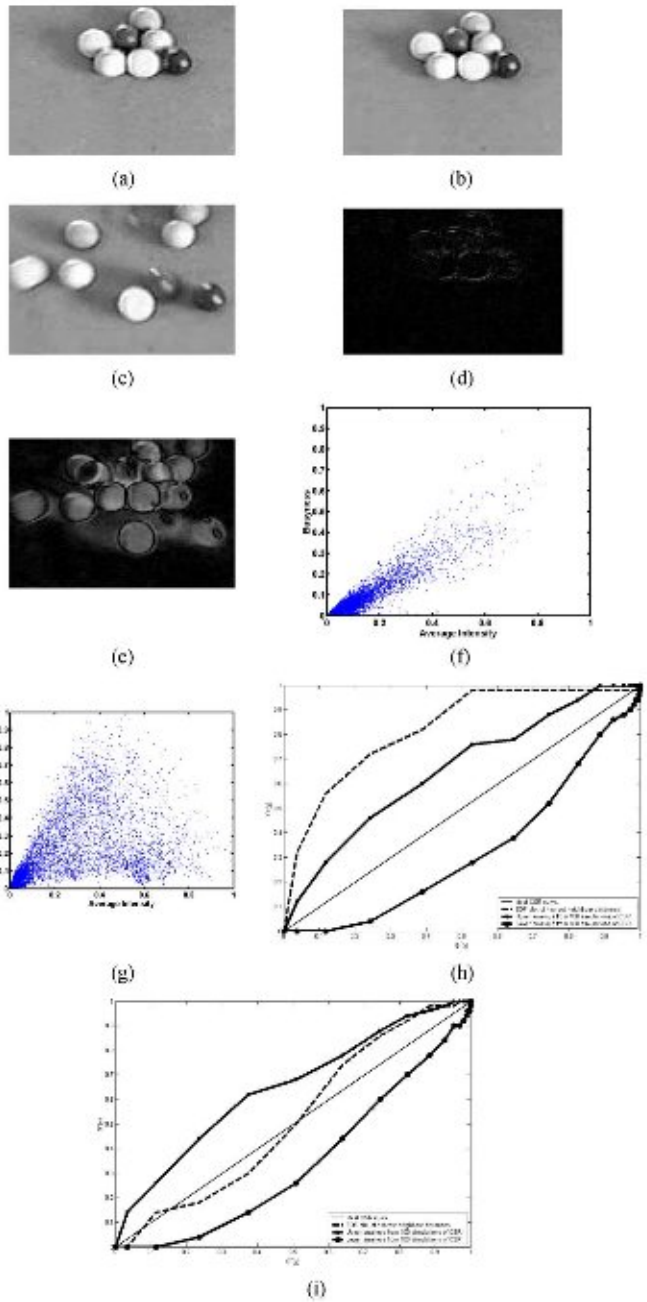

Fig. 1. (a) Frame #01. (b) Frame #03. (c) Frame #10. (d)-(e) Absolute temporal difference of Frame #01 and #03 and Frame #01 and #10, respectively. (f)-(g) Distribution of 2-D spatial domain feature vectors of Fig. 1(d) and (e), respectively. (h)-(i) Plot of $\hat{G}(y)$ versus $\overline{G}(y)$ for Fig. 1(f) and (g), respectively. Black diagonal curve: complete spatial randomness (CSR). Star-marked curve and circle-marked curve: upper and lower envelops. Broken-line curve: empirical distribution function (EDF) plot.

$\hat{G}(y)$ (along x axis) as shown in Fig. 1(h). Ideally the diagonal line in Fig. 1(h) signifies maximum randomness or complete spatial randomness (CSR). For every nearest neighbor distances (which is 100 numbers in between 0 and $\sqrt{2}$), the maximum and minimum of $G_j(y)$ for $j = 1, 2, 3, \ldots, 100$ as derived in (5) gives upper and lower envelopes. The broken-line curve of Fig. 1(h) furthest from the diagonal and above the upper envelope, shows the relationship between feature patterns derived combining frames 0 and 3 versus $\overline{G}(y)$ based on 100 numbers

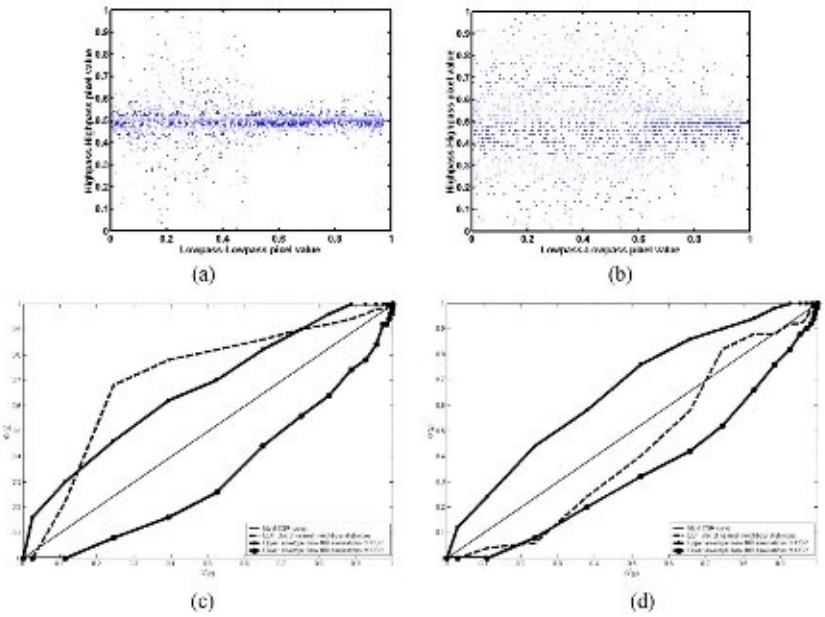

Fig. 2. (a)-(b) Distribution of 2-D wavelet domain feature vectors of Fig. 1(d) and (e), respectively. (c)-(d) Plot of $\dot{G}(y)$ versus $\overline{G}(y)$ for Fig. 2(a) and (b), respectively.

of random simulations. Clearly, the observation rejects the hypothesis that the temporal variation between frames 0 and 3 is random. This is corroborating with the visual interpretation.

The identical experiment is carried out between the frames 1 and 10. The temporal difference image between Fig. 1(a) and (c) is shown in Fig. 1(e). The distribution of 2-D feature point is shown in Fig. 1(g). The corresponding plot of $\dot{G}(y)$ versus $\overline{G}(y)$ is shown in Fig. 1(i). Notice the broken-line curve is well within the upper and lower envelope, closer to diagonal. This is a clear indication that the change in the scene content between frame 1 and 10 is *significantly* random. This is also conforming to the visual test and the objective measure of randomness is detailed below.

We have evaluated the measure of departure from ideal spatially random point patterns as given in (6). For the temporal difference image of Fig. 1(d), the departure measure is 1.54×10^{-2} while that for the Fig. 1(e) is 4.9138×10^{-1}. The corresponding rank of the measure of departure with respect to the simulated point patterns is 100 and 48 for Fig. 1(h) and (i), respectively. The rank 100 of measure of departure of feature patterns of Fig. 1(d) means that the pattern are least random corresponding to all the 100 simulated set of point patterns.

We now show the use of randomness measure for frequency domain features in wavelet space.

E. Estimation of Randomness Using Frequency Domain Features

In this module we have used both low and high frequency feature spaces of wavelet transform [9]. Each frame of the video sequence is subjected to Haar wavelet transform. After first level of decomposition, low-pass–low-pass (representing spatial information of the image after low-pass filtering along row and then along column) and high-pass–high-pass (representing edge information of the image after high-pass filtering along row and then along column) frequency space points are taken as feature vectors. For $N \times N$ image, two $N/2 \times N/2$ feature spaces are obtained after using low-pass wavelet masks $[1 \quad 1]$ and $[1 \quad 1]^T$, and high-pass masks $[1 \quad -1]$ and $[1 \quad -1]^T$ followed by down sampling of number of rows and columns by 2:1.

Fig. 2(a) and (b) are distribution of point patterns where points are feature vectors representing low-pass-low-pass (plotted along x-axis) and high-pass-high-pass (plotted along y-axis) coefficients of images of Fig. 1(d) and (e), respectively, after one level of Haar wavelet transformation. Fig. 2(c) and (d) are corresponding plots of randomness measure with respect to 100 sets of simulated random feature vectors. Clearly, plot of $\dot{G}(y)$ versus $\overline{G}(y)$ in Fig. 2(d) represents higher randomness compared to that in Fig. 2(c). The ranking of the randomness measure using wavelet features for Fig. 1(d) is calculated as 96 while that for Fig. 1(e) is 47 testing against 100 sets of random simulations.

As mentioned in the Introduction, various feature modules similar to the ones developed in this or previous section may be employed to derive randomness measures and corresponding probability values to declare a frame as key frame. In the next section, we investigate how these probability values are integrated so that key frame can be detected with higher confidence.

IV. Integration of Randomness Measure

In the previous section, randomness of image content is quantified through the estimation of ranking parameter $p(k)$. Since, the randomness measure depends on the type of features and the methodology used to calculate randomness, it is logical to experiment with a number of inter-related features to measure randomness. Alternately, we may hypothesize that no one particular set of feature measure and/or randomness test is absolute to determine a key frame; rather a combination of features should work better. Determining and subsequently fusing $p(k)$ based on different sets of features add on to the robustness of the

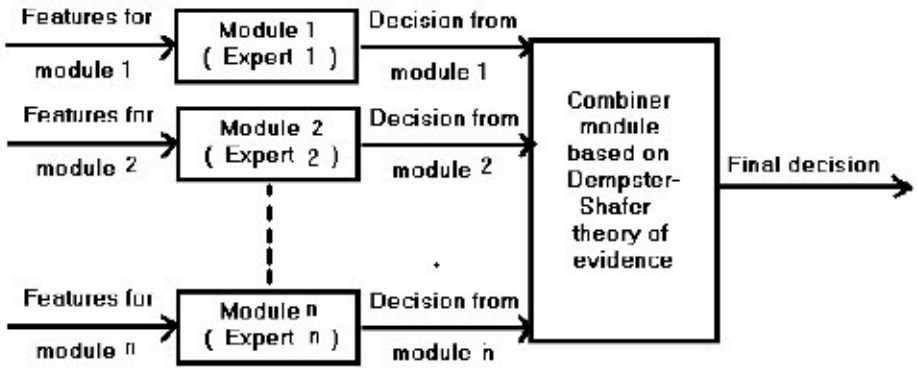

Fig. 3. Schematic for combination of different feature modules using DS Theory.

inferencing system. Moreover not all features may be sensitive to randomness test all the times and this may as well depend on varied contents of image sequence.

This issue can be viewed as if a number of *experts* are determining the randomness of the image content. And the final decision is taken only after fusing the opinions of all the experts as visualized in Fig. 3. We have used the DS theory of evidence to integrate different $p(k)$ values due to different set of features in order to get a unique randomness ranking of each frame of the video sequence. Note that we have not used any *a priori* knowledge of the possible location of key frame in video sequence. This is unlike using Bayesian techniques where *a priori* knowledge is used. The DS theory provides an advantage that accuracy of each expert or feature-processing module may not be known precisely. The DS theory can combine evidences or beliefs so long at least one evidence is common between the experts [3]. The different decision making modules for key frame estimation also generate evidences or beliefs independent of each other, which is why DS theory should be fit for combining beliefs from multiple independent sources. This is very relevant for image feature based key frame estimation problem.

Given a set of randomness ranking $\{p(k)\}$ for kth frame we first map them into a belief function that determines support for kth frame to be a key frame. This is described in the next section. In Section IV-B, the belief functions are combined based on which key frame is determined.

A. Mapping Randomness Measure to a Belief Function

We have referred each set of inter-related features as a module. Each module generates $p(k)$ which is the degree of support given by the particular module for selecting kth frame as key frame. Naturally $(1 - p(k))$ is the lack of confidence in declaring kth frame as key frame. The *possibilistic* vector representing decision alternatives for selecting a video frame as a key frame is $\{p(k), \overline{p}(k)\}$ where $\overline{p}(k) = (1 - p(k))$ is complement of $p(k)$. This possibilistic vector will be used to assign a belief or probability to an event where events are generated due to the uncertainty associated with declaring a frame as a key frame. Note that this uncertainty is due to the performance of a particular randomness measure module.

Let θ be this set of events, often referred as a *frame of discernment*. In this context θ consists of two events $\{c_1, c_2\}$ where c_1 and c_2 represent that a particular frame is a key frame or not, respectively. While assigning a belief or probability to these events as noted in the last paragraph, three situations can happen due to the uncertainty associated with the result of the randomness measure module. A probability can be assigned to the occurrence of event $\{c_1\}$, or $\{c_2\}$, or $\{c_1, c_2\}$, the last one being the degree of ignorance as from this particular assignment, belief or evidence in individual assignment of $\{c_1\}$ or $\{c_2\}$ cannot be inferred deterministically. All these three possibilities can be defined using the power set $P(\theta)$ of θ. So, the belief or basic probability assignment function χ can be given by [3]

$$\chi : P(\theta) \rightarrow [0, 1] \tag{9}$$

where $\chi(\phi) = 0$, $\sum_{\beta \subset \theta} \chi(\beta) = 1$, $\beta \subset P(\theta)$. The probability assignment $\chi(\beta)$ signifies the degree of evidence supporting the claim that the true hypothesis belongs to the set β.

To consider kth frame as key frame, given $\theta = \{c_1, c_2\}$ and possibilistic vector $\{p(k), \overline{p}(k)\}$, the algorithm for belief or probability assignment is given by the following.

If $(p(k) > \overline{p}(k))$, $\chi(\{c_1\}) = p(k)$, $\chi(\theta) = \overline{p}(k)$ and $\chi(\beta) = 0 \forall \beta \in P(\theta) \setminus \{\theta, \{c_1\}\}$.

Note, $\chi(\theta) = \overline{p}(k)$ signifies the extent of ignorance and to minimize this ignorance, we take the help of another expert or feature module for the key frame estimation problem. Similarly, if $(p(k) < \overline{p}(k))$, $\chi(\{c_2\}) = \overline{p}(k)$, $\chi(\theta) = p(k)$ and $\chi(\beta) = 0 \forall \beta \in P(\theta) \setminus \{\theta, \{c_2\}\}$.

For the situation, $p(k) = \overline{p}(k)$, $\chi(\{c_1\}) = \chi(\{c_2\}) = 0$ and $\chi(\theta) = 1$.

Obviously, the last situation points to most confusing state. Once the beliefs are assigned from different independent modules, the beliefs need to be combined together as discussed in the next section.

B. Combination of Belief Functions

In Section IV-A, the belief or probability assignment function χ^i is evaluated from possibilistic vector $\{p(k), \overline{p}(k)\}$ of the ith feature module. For different feature modules, the integrated joint probability assignment function χ^G is defined combining results from independent feature modules [3]

$$\begin{aligned} \chi^G &= \chi^1 \oplus \chi^2 \oplus \cdots \oplus \chi^k \\ &= \oplus_{i=1}^{k} \chi^i = (((\chi^1 \oplus \chi^2) \oplus \chi^3) \oplus \cdots \chi^k) \end{aligned} \tag{10}$$

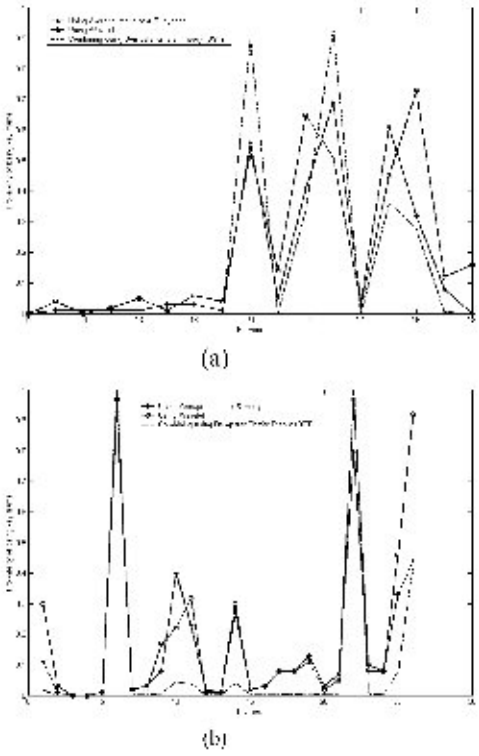

Fig. 4. (a)-(b) $p(k)$ and $\pi\{c_2\}$ are plotted against image frames of Billiard and News Video Sequence.

where, joint probability assignment function of χ^1 and χ^2 is given by

$$\chi^1 \odot \chi^2 = \begin{cases} \dfrac{\sum_{\alpha\cap\gamma=\beta} \chi^1(\alpha)\chi^2(\gamma)}{1-K}, & \text{if } \beta \neq \phi \\ 0, & \text{if } \beta = \phi. \end{cases} \tag{11}$$

Sets α, β, and γ are the members of $P(\theta)$ and $\alpha \cap \gamma = \beta$. Note that in (11), $\chi^1(\alpha)$ and $\chi^2(\gamma)$ are known from randomness measure modules 1 and 2, respectively. Integration of probability assignment does not arise in case $\beta = \phi$. The normalization parameter is defined as $K = \sum_{\alpha\cap\gamma=\phi} \chi^1(\alpha)\chi^2(\gamma)$. This information aggregation rule is commutative and associative.

Given the integrated probability assignment χ^G, the task is to recalculate the probability of each event, c_1 and c_2 for the problem of key frame estimation. This is achieved through pignistic probability assignment [3]

$$\pi(c_i) = \sum_{\beta\subseteq\theta, c_i\in\beta} \frac{\chi^G(\beta)}{|\beta|}. \tag{12}$$

The $\pi(c_i)$ is the probability assigned to element c_i, a member of the frame of discernment from integrated body of evidence χ^G. Note that as detailed in Section IV-A, the frame of discernment contains two events c_1 and c_2 representing a frame as key frame or not, respectively. A frame is declared key frame if $\pi(c_1) > \pi(c_2)$. In case $\pi(c_1) \quad \pi(c_2)$, no conclusive decision can be taken and we have considered such situation to be same as $\pi(c_1) < \pi(c_2)$.

The result of integration of randomness measures is shown in Fig. 4(a) for billiard video sequence. The randomness measure from spatial domain and frequency domain feature modules as calculated in Sections III-D and E are combined in this example. The x axis of Fig. 4(a) represents frame numbers whereas $p(k)$ (for individual feature module) or $\pi(c_i)$ (after combining two feature modules) values are plotted along y axis. The star-marked and circle-marked curves of Fig. 4(a) represent $p(k)$ values for average-busyness module and wavelet based feature module, respectively. Both the spatial and frequency domain features are evaluated on difference images in the consecutive frames. The video frames are numbered from 1 and $p(k)$ values are evaluated starting from frame 2.

The integrated probability $\pi(c_1)$ values supporting the frame as key frame are plotted in broken-line. Notice that for frame positions close to 9 and 12, individual $p(k)$ value obtained from each feature module has reinforced the final decision and the combined probability is enhanced increasing the confidence in the hypothesis that the respective frames are indeed key frames. For the frame positions around 14, the decision however reflects that combination of feature increases the confusion. This is because two modules are arriving at a conflicting decision when treated individually. To declare a key frame we have taken the strict condition $\pi(c_1) > \pi(c_2)$. Note that as noted in Introduction, we do not need any explicit use of threshold to find a key frame. The process is also repeated for News video sequence and the integrated probability shown as broken-line graph in Fig. 4(b) agrees with the ground truth observation.

Next we take examples from a number of video sequences and demonstrate the overall performance of our system.

V. Results

The proposed methodology is tested on a number of standard video sequences as listed in Tables I and II. Out of the six video sequences whose results are presented in this paper, only the particle video sequence is a synthetic video sequence. In each case, spatial domain feature module comprising average intensity and busyness and wavelet based frequency domain feature modules are used. Consecutive frame difference images are used to calculate the features. The results shown in Table I are obtained after combining $p(k)$ values using the DS theory as derived in Section IV. The process of declaring a frame as key frame from integrated probability is detailed in Section IV-B.

We have compared our approach with the PME based approach [6] where between frame motion is estimated from optic flow. As noted in Section II, PME is calculated multiplying dominant motion direction with motion magnitude. The local maximums of PME values indicate potential key frame [6]. The comparison of the proposed and PME based approach is described in terms of accuracy of detection of key frames with respect to ground truths. The accuracy is specified as the number of correctly identified key frames and accuracy in correctly identifying the spatial location of the respective key frames within a video sequence. Number of key frames incorrectly identified is categorized either as false positive or false negative. False positives are those video frames, which are erroneously marked as key frames. False negative counts number of true key frames missed by the detection technique. The comparison shown in Table I clearly shows that the proposed approach outperforms the PME based approach. The proposed approach has both low false positive and false negative values than that of PME based approach. Also, total count

TABLE I
PERFORMANCE OF THE PROPOSED APPROACH VIS-À-VIS GROUND TRUTH AND [6]

	GROUND TRUTH		PME BASED APPROACH [6]				PROPOSED APPROACH			
Video Sequence	# of key frames	Position of key frames	# of key frames	Position of key frames	# of false positive	# of false negative	# of key frames	Position of key frames	# of false positive	# of false negative
Billiards video sequence (total 20 frames)	6	02, 10, 12, 13, 15, 16.	5	02, 05, 14, 15, 16.	2	3	6	02, 10, 12, 13, 15, 16.	0	0
Particle video sequence (total 40 frames)	11	02, 07, 10, 13, 16, 19, 25, 28, 31, 34, 38.	12	02, 05, 07, 08, 10, 11, 16, 25, 28, 31, 34, 38.	3	2	10	02, 07, 10, 13, 16, 19, 28, 31, 34, 38.	0	1
Tennis video sequence (total 20 frames)	4	02, 06, 12, 19.	5	02, 07, 09, 16, 19	3	2	4	02, 06, 12, 19.	0	0
Harvey video sequence (total 60 frames)	23	03, 09, 10, 12, 15, 17, 21, 23, 28, 29, 30, 31, 32, 33, 35, 38, 41, 44, 46, 49, 51, 52, 53.	23	03, 09, 12, 18, 19, 22, 23, 28, 29, 32, 33, 35, 37, 38, 41, 44, 45, 47, 49, 51, 52, 53, 57.	7	7	24	03, 09, 10, 12, 13, 15, 16, 17, 21, 23, 28, 29, 31, 32, 33, 35, 38, 41, 44, 46, 49, 51, 52, 53.	2	1
News video sequence (total 30 frames)	13	02, 03, 07, 10, 11, 12, 15, 18, 19, 20, 23, 26, 27.	11	02, 03, 04, 07, 11, 15, 19, 20, 21, 23, 27.	2	4	12	02, 07, 10, 11, 12, 15, 18, 19, 20, 23, 26, 27.	0	1
Foreman video sequence (total 30 frames)	13	04, 05, 09, 11, 16, 18, 19, 20, 21, 23, 24, 25, 26.	14	02, 03, 05, 07, 09, 11, 16, 19, 20, 21, 23, 25, 26, 29.	4	3	14	04, 05, 09, 10, 11, 16, 18, 19, 20, 21, 23, 24, 25, 26.	1	0

of key frames using randomness measure is more consistent to ground truth than that using PME.

For further objective comparison, false positive and false negative counts are expressed in terms of percentage error with respect to the total number of ground truth key frames. This is shown in Table II. In case of false positives, where additional key frames detected erroneously are absent in the ground truth data, a spatial accuracy is measured. The spatial accuracy is in terms of number of frames by which the false positive frame is shifted with respect to nearest ground truth key frame. Naturally, a low value in this accuracy measure shows that false positive frames are comparatively better approximations (that is temporally closer) to actual key frames. Table II further supports that the integrated randomness measure has identified better sets of key frames compared to PME based approach.

The proposed approach has the promise of real time use. For demonstrating the idea the feature detection module employed in this paper is $O(n^2)$. However, use and integration of randomness measure can take the help of more computationally and space efficient features having better discrimination power. The generation of random numbers for simulated data can be implemented offline and only once for a series of video frames.

VI. CONCLUSION

The randomness measure derived from spatial arrangements of point pattern is successfully used for key frame estimation problem in video. The feature vectors derived from video sequence are used to generate the spatial point pattern. The efficacy of the approach is in integrating randomness measure of different feature modules. This is more akin to practical decision

TABLE II
OBJECTIVE COMPARISON OF THE PROPOSED APPROACH VIS-À-VIS [6]

Algorithms	Video Sequences	Accuracy in detecting key frames		False detection (%).	
		Accuracy in Number (%)	Accuracy in Position	False positive	False negative
PMF based approach [6]	Billiards video sequence	50.00	2	33.33	50.00
	Particle video sequence	81.82	1.33	27.27	18.18
	Tennis video sequence	50.00	2.33	75.00	50.00
	Harvey video sequence	69.57	1.57	30.43	30.43
	News video sequence	69.23	1	14.28	30.77
	Foreman video sequence	76.92	2		
Proposed approach	Billiards video sequence	100	0	0.00	0.00
	Particle video sequence	90.90	0	0.00	9.10
	Tennis video sequence	100	0	0.00	0.00
	Harvey video sequence	95.65	1	8.70	4.35
	News video sequence	92.31	0	0	7.69
	Foreman video sequence	100	1	7.69	0

making process where robust inferencing is the outcome of combination of evidences or beliefs. This is particularly relevant for key frame estimation or similar such problems where no one set of feature can always have the most discriminating power. Comparison of the proposal with similar approach shows promise. We are now investigating this technique for shot boundary and fade in and fade out detection in a video. Also, we are exploring the use of randomness measure in identifying the most active content in a scene so that region specific coding scheme can be employed.

REFERENCES

[1] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press, 2003.

[2] S. Ju, M. Black, S. Minneman, and D. Kimber, "Summarization of videotaped presentations: Automatic analysis of motion and gesture," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 686–696, Sep. 1998.

[3] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty and Information*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[4] H. C. Lee and S. D. Kim, "Rate-Driven frame selection using temporal variation of visual content," *Electron. Lett.*, vol. 38, no. 5, pp. 217–218, Feb. 2002.

[5] ——, "Rate-Constrained key frame selection using iteration," in *Proc. ICIP2002*, Rochester, NY, 2002, pp. 928–931.

[6] T. M. Liu, H. J. Zhang, and F. H. Qi, "A novel video key frame extraction algorithm based on perceived motion energy model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 10, pp. 1006–1013, Oct. 2003.

[7] D. P. Mukherjee, P. Pal, and J. Das, "Sodar image segmentation by fuzzy c-means," *Signal Process.*, vol. 54, pp. 295–301, 1996.

[8] E. Stringa and C. S. Regazzoni, "Real-Time video-shot detection for scene surveillance applications," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 69–79, Jan. 2000.

[9] P. Umbugh, *Computer Vision and Image Processing: A Practical Approach using CVIPtools*. Englewood Cliffs, NJ: Prentice Hall, 1998.

[10] W. Wolf, "Key frame selection by motion analysis," in *Proc. ICASSP'96*, Atlanta, GA, 1996, pp. 1228–1231.

Dipti Prasad Mukherjee (M'99–SM'05) is a Professor in the Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, India.

He has published three books and more than 70 peer-reviewed papers. He has held visiting faculty positions at Oklahoma State University, University of Virginia, and University of Alberta.

Dr. Mukherjee is the recipient of a predoctoral fellowship to the University of Oxford, Oxford, U.K., and UNESCO-CIMPA fellowships to INRIA, France and to ICTP, Italy. He is the senior member of the Computer Society of India and had served on the Editorial Board of IEEE SIGNAL PROCESSING LETTERS.

Sitansu Kumar Das received the B.E. degree in mechanical engineering from Bengal Engineering and Science University, Bengal, India, in 1997 and the M.Tech. degree in computer science from the Indian Statistical Institute, Kolkata, India, in 2003.

Currently, he is Senior Research Fellow in the Electronics and Communication Sciences Unit of the Indian Statistical Institute. His research interests are in the areas of computer vision, pattern recognition, and image processing.

Subhra Saha received the B.Tech. degree in computer science and technology from the University of Kalyani, Kalyani, India, in June 2004.

He was with the Indian Statistical Institute, Kolkata, from July 2004 to November 2004. Since December 2004, he has been working with the Tata Consultancy Services Ltd., Mumbai, India, as an Assistant Systems Engineer in a mission critical project as a Software Developer. He is interested in applying advanced pattern recognition techniques to model semantics in image data.