

REVIEW OF VIDEO SEGMENTATION APPROACH

¹Dattatraya A. Jadhav,²Dr. Parul S. Arora,³Dr. Yogesh Kumar Sharma

¹Research Scholar, ²⁻³Associate Professor

¹⁻²J.J.T.University, Churella, Jhunjhunu (Raj.).

³JSPM's ICOER Wagholi, Pune, India.

Abstract: A Present investigation topic focus on video comprises of video abstraction, video classification, video annotation and content based video retrieval. In all these applications one needs to identify shot and key frames in video which will correctly and briefly indicates the contents of video. This paper presents a brief overview on shot boundary detection and its techniques.

Summary: The paper is organized with introduction then review of category of basic shot boundary detection with their parameter evaluation.

1. Introduction:

The augmented accessibility and practice of online digital video has evolved a necessity for automatic video content analysis techniques. Detection of shot boundaries is an initial and fairly important processing step in video analysis, especially for indexing and effective retrieval applications. [1-22]

2. Shot Boundary Detection:

A video is usually structured into frames, shots, sub shots or scenes as explained in Figure 1. Formerly performing any kind of processing, usually, the initial step of CBVIR (concept of Content-Based Video Indexing and Retrieval) systems is to segment a video into its main components. A shot is measured as the elementary unit of a video, and is well-defined as a continuous sequence of frames taken from a single camera, serves an action during the time. Detecting shots in a video is acknowledged as the shot boundary detection (SBD) problem. A sub-shot takes place when the visual content of the present shot changes dramatically. A scene is a set of decor that signifies the place of the action (i.e. beach, forest, countryside, building, etc.). In a video sequence, a scene may comprise of several shots.

Detecting sub-shots and scenes are denoted to micro-segmentation and macro-segmentation, respectively. Mostly both fields need shot boundary detection first. Dissimilar types of transitions are added between shots to develop the video sequence. Segmenting a video into shots is same as detecting the transitions between shots. Once a video is categorized into shots, further auxiliary applications can be performed. Usually treated as a first step in CBVIR, SBD is a fundamental step towards subsequent high-level applications.[6]

2.1 Shot Boundary Detection challenges:

Video shot boundary detection algorithms have to encounter the difficulty of finding shot boundaries in the presence of camera and object motion and illumination variations. Furthermore, different video shot boundaries may present very different appearances like abrupt temporal changes or smooth temporal transitions. Shot boundary detection has been an area of active research. Many automatic techniques have been developed to detect frame transitions in video sequences. [1] The goal of SBD is to split an edited video into consecutive frames which show a continuous progression of video, as shown in this illustration below as shown in the figure below

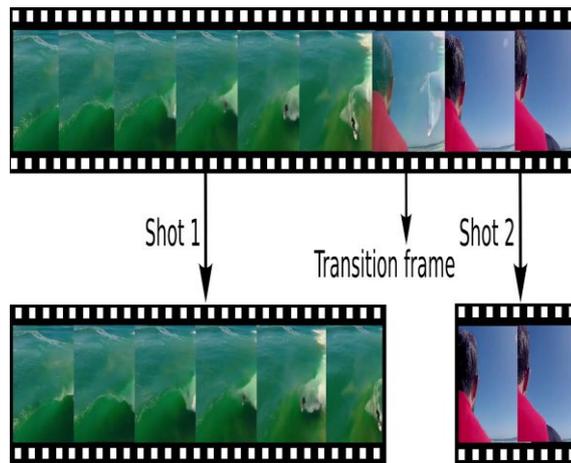
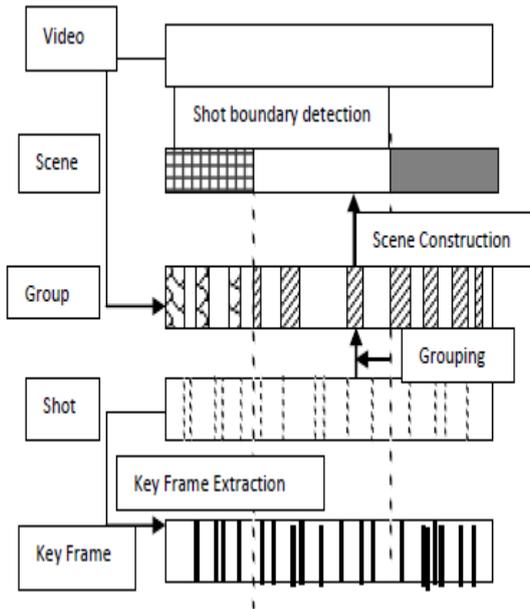


Figure 1: Hierarchical structure of a video [1]

Figure 2: splitting a video into a set of shots [1]

2.2 Shot Transition types

Abrupt Transition: Rapid transition from one shot to the subsequent shot results in abrupt transitions. Here there is a quickly change in the aspect of the last frame belonging to the previous shot compared to the aspect of the first frame of the consequent shot.[5]

Gradual Transition: The gradual transition spans over a sequence of multiple frames. In general, the resemblance between frames within a shot is high while the similarity is low between the frames of different shots. It is again classified into 3 types. They are

1. Dissolve
2. Fade
3. Wipe

A dissolve is a gradual transition from one scene to another in which the first scene fade – out and the second scene fade – in. so it is a combination of fade – in and fade – out .[4]



Figure 3: shows the Dissolve transition [2]

A fade – out is a sluggish decrease in brightness resulting in a black frame and A wipe is a gradual transition in which a line moves across the scene, with the new scene appearing behind the line



Figure 4: shows the Fadeout transition[2]



Figure 5: shows the wipe transition [2]

2.3 Attributes used for illustration of video frame

Almost all

shot change detection algorithms diminish the large dimensionality of the video domain by extracting a small number of features from one or more regions of interest in each video frame. Such features include the following.[2]

2.3.1 Luminance/color histogram

The simplest feature that can be used to describe an image is its average grayscale luminance. Moreover, is vulnerable to changes in illumination. A substantial feature for an image is the grayscale or color histogram. Its advantage is discriminant, easy to compute, and mostly unresponsive to translational, rotational, and zooming camera motions. For these reasons, it is extensively used. However, it does not characterize the spatial distribution of color in an image.[12]

2.3.2 Image edges

Alternative choice for characterizing an image is its edge information. The gain of this feature is that it is adequately invariant to illumination changes and numerous types of motion, and is associated to the human visual perception of a scene. Its main hindrance is computational cost, noise sensitivity, and when not post-processed, high dimensionality. A number of researchers have used a Gaussian smoothed step edge (an error function) as the simplest extension of the ideal step edge model for modeling the effects of edge blur in practical applications. Thus, a one-dimensional image which has exactly one edge placed at $x=0$ may be modeled as:[11-12]

$$f(x) = \frac{I_r - I_l}{2} \left(\operatorname{erf} \left(\frac{x}{\sqrt{2}\sigma} \right) + 1 \right) \quad --1 \quad [3-5]$$

2.3.3 Features in transform domain

The information present in the pixels of an image can also be signified by using transformations such as discrete Fourier transform, discrete cosine transform and wavelets. Such transformations also lead to representations in lower dimensions. Drawbacks include high computational cost, effects of blocking while computing the transform domain coefficients, and loss of information caused by retaining only a few coefficients.

2.3.4 Motion

This can be used as a feature for identifying shot transitions, but it is usually united with other features, since motion itself can be extremely discontinuous within a shot (when motion changes abruptly) and is not suitable when there is no motion in the video.[18]

2.4 Spatial domain for feature extraction

The extent of the region from which individual features are mined plays an important role in the overall performance of algorithms shot change detection. A small region inclines to reduce detection invariance with respect to motion, while a large region might lead to missed transitions between similar shots. In the following, we will describe various possible choices:

2.4.1 Single pixel

Some algorithms originate a feature for each pixel such as luminance and edge strength. However, such an approach results in a feature vector of very large dimension, and is very sensitive to motion, unless motion compensation is subsequently performed.

2.4.2 Rectangular block:

Another technique is to segment each frame into equal-sized blocks and extract a set of features (e.g., orientation, color histogram) from these blocks. This approach has the benefit of being invariant to small motion of camera and object, as well as being adequately discriminant for shot boundary detection. This approach divides a picture into blocks of regions. Every region follows approximate object boundaries, and is formed of rectangular blocks. The dimensions of the blocks could vary among constant region to raised approximate the particular object boundary. Block-based segmentation algorithms are unit developed principally for grayscale or color compound pictures.[21]

2.5 Arbitrarily shaped region

Feature extraction can also be applied to arbitrarily shaped and sized regions in a frame, derivative by spatial segmentation algorithms. This permits the derivation of features based on the most homogeneous regions, thus facilitating a better detection of temporal discontinuities. The main hindrance is the high computational complexity and instability of region segmentation.

2.5.1 Whole frame

The algorithms that extract features from the whole frame have the improvement of being robust with respect to motion within a shot, but tend to have reduced performance at detecting the change between two similar shots.

2.5.2 Similarity Measure

To estimate discontinuity between frames based on the selected features, an appropriate similarity /dissimilarity metric needs to be chosen. A wide variety of divergence measures has been used in the literature. Some of the commonly used measures are Euclidean distance and log-likelihood ratio.

2.6 Temporal domain of continuity metric

Another significant characteristic of shot boundary detection algorithms is the temporal window that is used to perform shot change detection. In general, the objective is to select a temporal window that contains a representative amount of video activity. The following cases are typically used. [6]

2.6.1 Two frames

The humblest way to detect discontinuity between frames is to look for a high value of the discontinuity metric between two successive frames. However, such method can fail to discriminate between shot transitions and changes within the shot when there is momentous variation in activity among different parts of the video or when certain shots contain events that cause brief discontinuities (e.g., photographic flashes). It also has difficulty in detecting gradual transitions.[19]

2.6.2 N-frame window

One technique for improving the above problems is to detect the discontinuity by using the features of all frames within a suitable temporal window, which is centered on the location of the potential discontinuity.

2.6.3 Interval since last shot change

Another method for detecting a shot boundary is to calculate one or more statistics from the last detected shot change up to the recent point, and to check if the next frame is consistent with them. The tricky with such approaches is that there is often great variability within shots, such that statistics computed for an entire shot may not be representative of its end.

2.7 Shot change detection method

Static thresholding

This is the basic elementary decision method, which involves relating a metric expressing the resemblance of the features computed on adjacent frames against a fixed threshold. This only accomplishes well if video content exhibits similar characteristics over time, and only if the threshold is manually adjusted for each video. Adaptive Thresholding The recognizable solution to the problems of the static thresholding is to vary the threshold liable on a statistic of the feature difference metrics within a temporal window. Probabilistic Detection Perhaps the furthestmost vivid way to detect shot changes is to model the pattern of specific types of shot transitions and, presupposing precise probability distributions for the feature difference metrics in each shot, achieve optimal a posteriori shot change estimation.

2.8 Shot Boundary detection techniques:

2.8.1 Pixel based shot boundary detection

It is the simplest method for determining shot boundaries. The difference between corresponding pixels of two consecutive frames is computed. If the difference is greater than some threshold, then shot boundary is assumed. Illustration done using below test scenarios. The stately distance value between two frames governs the shot change. If the measured distance is superior to the threshold TH, it is regarded as shot change, as shown in Figures 3 and 4; otherwise, the two frames are similar.[3-4]

$$D(i, i + 1) = \sum_{x=1}^x \sum_{y=1}^y |P_i(x, y) - P_{i+1}(x, y)| / xy \quad \text{-----2}$$

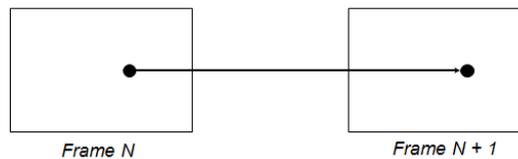


Figure. 6: Pixel Comparison [3]

As the conversion of frame is not always sudden conversion, the conversion of some frames is slow. If only the difference between two consecutive frames is used for detection, the detection of shot will be influenced. Therefore, the shot boundary detection proposed in this paper can be divided into three steps to detect the boundary of shot change.

2.8.2 Block based shot boundary detection

In this practice individually frame is distributed in fixed number of blocks and difference between blocks at consecutive position in frame i and i+1 has been used to find out difference between frames. If this frame difference is greater than particular threshold value then break is detected. [7]

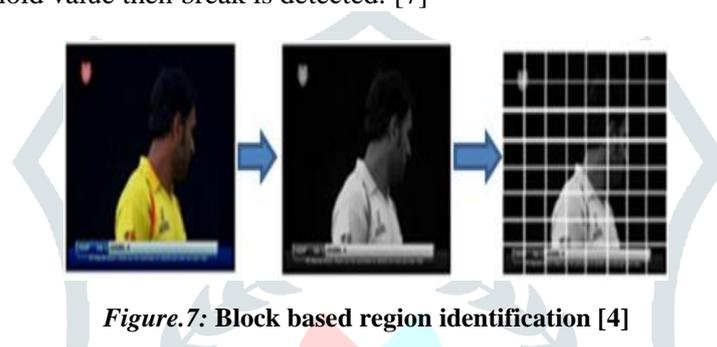


Figure.7: Block based region identification [4]

The formula is used for block based shot boundary detection.

$$D(i, i + 1) = a_0 + \sum_{k=1}^b C_k |DP(i, i + 1, k)| \dots 3 \quad [10]$$

Where each frame is divided into b blocks and DP(i , i+1 , k) indicated difference of kth block between ith and (i+1)thframe and C_k is predetermined coefficient for block k. The absolute variance between all the blocks of two consecutive frames is added to find out the difference between the two frames. This technique is relatively slow due to complication of formulas. This method can't identify dissolve, fade or fast moving objects. But computationally better than pixel based shot detection.[4]

2.8.3 Shot boundary detection using motion activity descriptor

The basis of any video segmentation method consists in detecting visual discontinuities along the time domain. During this process, it is required to extract visual features that measure the degree of similarity between frames in a given shot. This measure is related to the difference or discontinuity between frame n and n+j where j>= 1. The figure explains the Different steps for motion intensity extraction. [5]

Method	Advantages	Disadvantages
Pixel-Comparison	Simple, easy to implement	Computationally heavy,Very sensitive to moving object or camera motion
Block based	Performs better than pixel	Can't identify dissolve, fade, fast moving objects
Histogram comparison	Performance is better Detects hard-cut, fade, wipe and dissolve	Fails if the two successive shots have same histogram. Can't distinguish fast object or camera motion
Edge Change Ratios	Detects hard-cut, fade, wipe and dissolve	Computationally heavyFails when here is large amount of motion

Table 1: Comparison of different methods

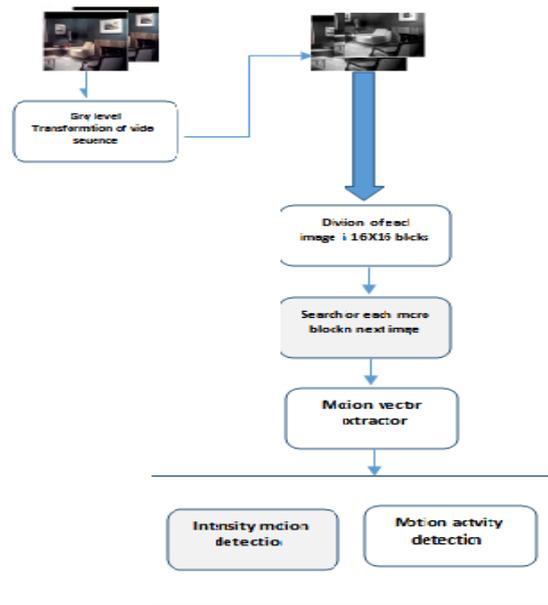


Figure.8: Flow chart

EXPERIMENTAL RESULTS-I

VIDEO TEST DATA [5]

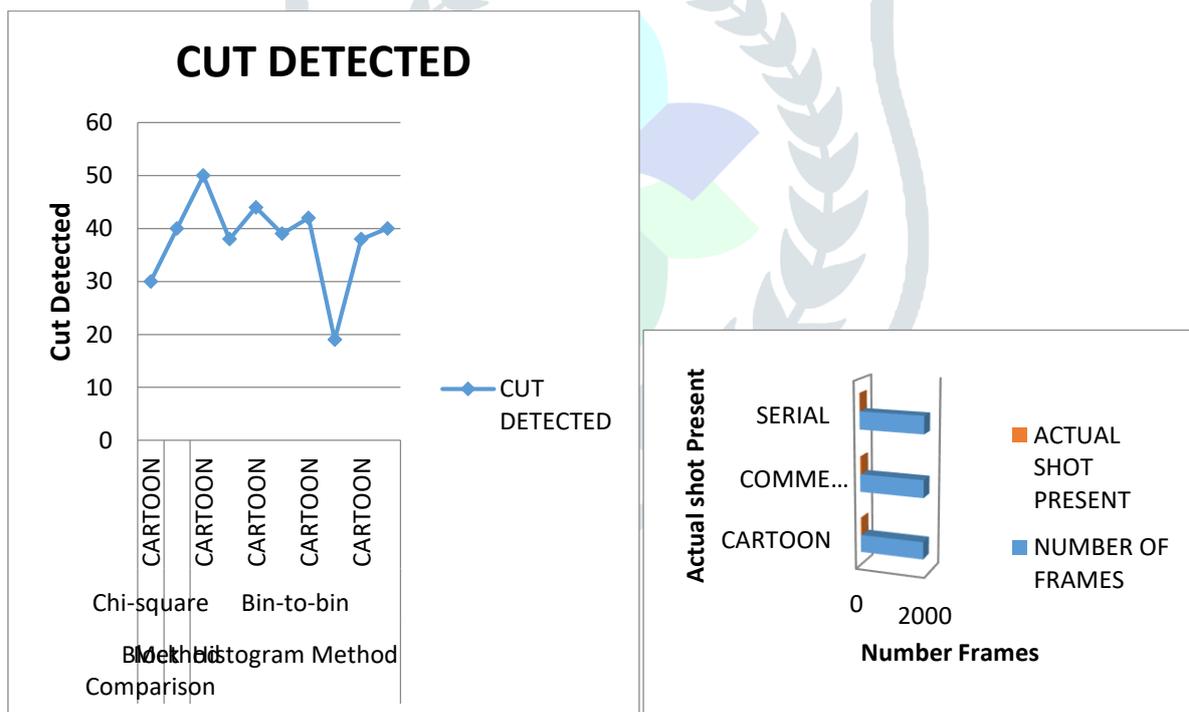


Figure.10: Experimental Results-I [5]

3. Conclusion and Future development

The different techniques are discussed to identify a shot boundary subject upon the contents and the change in contents of video. We propose new technique for video short boundary detection using complex text wavelet transform.

References:

- [1] Chongke Bi, Ye Yuan, Jiawan Zhang, Yun Shi, Yiqing Xiang, Yuehuan Wang, And Ronghui Zhang (2018), “Dynamic Mode Decomposition Based Video Shot Detection”, “IEEE Access”, Volume 6, DOI 10.1109/Access.2018.2825106.
- [2] Asirvatham Kethsy Prabavathy, and J. Devi Shree (2018), “Gradual transition detection in shot boundary using gradual curve point”, Sci. Foundation Sri Lanka. DOI: 10.4038/jnsfsr.v46i3.8491. 46 (3): 393 – 398.
- [3] Rui Liang, Qingxin Zhu, Honglei Wei, Shujiao Liao (2017), “A Video Shot Boundary Detection Approach based on CNN Feature”, IEEE International Symposium on Multimedia, 978-1-5386-2937-6/17, DOI 10.1109/ISM.2017.97.
- [4] Sawitchaya Tippaya, Suchada Sitjongsataporn, Tele Tan, Masood M. Khan, Kosin Chamnongthai (2017), “Multi-modal Visual Features Based Video Shot Boundary Detection”, IEEE Access , DOI 10.1109/ACCESS.2017.2717998.
- [5] Huang-Chia Shih (2017), “A Survey On Content-Aware Video Analysis For Sports, IEEE Transactions On Circuits And Systems For Video Technology”, IEEE Transactions On Circuits And Systems For Video Technology. Vol. 99, No. 9. DOI 10.1109/Tcsvt.2017.2655624.
- [6] Annamaria Mesaros, Toni Heittola, Aleksandr Diment, Benjamin Elizalde, Ankit Shah, Emmanuel Vincent, Bhiksha Raj, Tuomas Virtanen (2017), “Tasks, Datasets And Baseline System, Detection and Classification of Acoustic Scenes and Events”, Dcase 2017 Challenge Setup, , Munich, Germany.
- [7] Renzhen Ye , And Xuelong Li (2017) , “Collective Representation For Abnormal Event Detection”, Journal Of Computer Science And Technology, 32(3): 470–479 May 2017. DOI 10.1007/S11390-017-1737-8.
- [8] Yuming Fang, Yuan Yuan, Leida Li, Jinjian Wu, Weisi Lin, And Zhiqiang Li (2017), “Performance Evaluation Of Visual Tracking Algorithms On Video Sequences With Quality Degradation”, IEEE Access, DOI 10.1109/Access.2017.2666218.
- [9] Zhe Yang, Lihua Tian, Chen Li (2017), “A Fast Video Shot Boundary Detection Employing OTSU’s Method and Dual Pauta Criterion”, IEEE International Symposium on Multimedia. 978-1-5386-2937-6/17 IEEE DOI 10.1109/ISM.2017.114.
- [10] Rui Liang, Qingxin Zhu, Honglei Wei, Shujiao Liao (2017), “A Video Shot Boundary Detection Approach based on CNN Feature”, International Symposium on Multimedia IEEE 2017. 978-1-5386-2937-6/17 DOI 10.1109/ISM.2017.97.
- [11] S. Kaavya, Lakshmi Priya GG (2017), “Local Binary Pattern based Shot Boundary Detection for Video Summarization”, Second International Conference on Recent Trends and Challenges in Computational Models 2017, 978-1-5090-4799-4/16 IEEE DOI 10.1109/ICRTCCM.2017.64.
- [12] Mehrnaz Fani, Mehran Yazdi (2016), “Shot Boundary Detection with Effective Prediction of Transitions’ Positions and Spans by Use of Classifiers and Adaptive Thresholds”, 24th Iranian Conference on Electrical Engineering (ICEE), 978-1-4673-8789-7/16 ©2016 IEEE.
- [13] T. Kar, P. Kanungo (2016), “A Texture Based Method for Scene Change Detection, Communication and Information Technology”, Conference (PCITC) Siksha ‘O’ Anusandhan University, Bhubaneswar, India. 978-1-4799-7455-9/16/IEEE.
- [14] Soyoung Park, Jeongwoo Son, Sun-Joong Kim (2016), “Effect of Adaptive Thresholding on Shot Boundary Detection Performance”, IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) 2016. 978-1-5090-2743-9/16/2016 IEEE.
- [15] Yi Huo, Yanfeng Wang , Haihe Hu (2016), “Effective Algorithms For Video Shot And Scene Boundaries Detection”, ICIS 2016, June 26-29, 2016, Okayama, Japan. IEEE 978-1-5090-0806-3/16.
- [16] Xiaojun Chang, Yao-Liang Yu, Yi Yang, and Eric P. Xing (2016), “Semantic Pooling for Complex Event Analysis in Untrimmed Videos”, IEEE Access, DOI 10.1109/TPAMI.2016.2608901, IEEE.
- [17] Xin Liu, Jin Dai (2016), “A Method of Video Shot-Boundary Detection based on Grey Modeling for Histogram Sequence”, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.9, No.4 (2016), pp.265-280, 2016.
- [18] Mr. Naim G. Fanan, Dr. Aatish S. Khobragade (2016), “A Fast Robust Technique for Video Shot Boundary Detection” , Online International Conference on Green Engineering and Technologies (IC-GET), 978-1-5090-4556-3/16/\$31.00 ©2016 IEEE.

- [19] Y ouxian Zheng, Yuan Zhang (2016), “Abrupt Shot Boundary Detection with Combined Features and SVM”, 2nd IEEE International Conference on Computer and Communications, 978-1-4673-9026-2116 IEEE.
- [20] Huayong Liu, Tao Li (2016), “An Adaptive threshold Shot Detection Algorithm Based on Improved Block Color Features”, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 978-1-5090-4093-3/16 IEEE.
- [21] Wenjing Tong, Li Song, Xiaokang Yang, Hui Qu, Rong Xie (2016), “CNN-Based Shot Boundary Detection and Video Annotation”, Cooperative Medianet Innovation Center, Shanghai China 2016.
- [22] B.H.Shekar, K.P. Uma, Raghurama Holla (2016), “Shot Boundary Detection Using Correlation based Spectral Residual Saliency Map”, Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India, 978-1-5090-2029-4/16 IEEE.

