International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 08 - Issue 09 || September 2022 || PP. 17-22

Video Shot Transition Detection Using Convolutional Neural Network (CNN), Euclidean Distance Algorithm and Change Point Analysis Algorithm

Usha Matta[1], Lakshmanarao Battula[2]
[1](Assistant Professor, cse, Visakha Institute of Engineering and Technology, Visakhapatnam, AP)
[2](Assistant Professor, cse, Centurion University, Vizainagaram, AP)

Abstract: Multimedia streams usage increases nowadays and that creates the scope of development of efficient and effective methodologies for manipulating different image databases storing this type of information. Any content-based access to video data always requires parsing of each video stream into its building blocks. Any video stream consists of a number of shots; each one is a sequence of frames. Shot boundary detection is the very first step in any video stream-analysis system and there are numbers of proposed techniques are available for solving the problem of shot boundary detection, but the major limitation to them are their inefficiency, lack of reliability and less trustworthy. Here, proposes to learn shot boundary detection end-to-end, from pixels to final shot boundaries. For training such a model, we created our own dataset and automatically generated transitions such as cuts, dissolves and fades. Here we propose a Convolutional Neural Network (CNN) and Euclidean Distance algorithm and Change Point Analysis algorithm to make the system more efficient and accurate in nature.

Keywords: Convolutional Neural Network (CNN), Euclidean Distance algorithm.

I. INTRODUCTION

All digital video information consists of a series of many frames or images. Over the years image processing technology has developed comprehensive and complete measures and techniques to index, store, edit, retrieve, sequence and present video material. To develop any content-based manipulations on digital video stream information, this information must first be structured and broken down into different components. The basic structural building blocks are called shots and the boundaries between shots need to be determined automatically

A shot in video stream information may be defined as continuous images (i.e. frames) from a single camera at a time. A shot boundary is defined the gap between two shots. A cut is a type of shot boundary where one shot abruptly changes to another shot. An example of a shot cut is where the last frame in one shot is followed by the first frame in the next. Examples of other different types of shot boundary are fades (where the frames of the shot gradually change from or to black), dissolves (where the frames of the first shot are gradually morphed into the frames of the second) or wipes (where the frames of the first shot are moved gradually in a horizontal or vertical direction into the frames of the second).

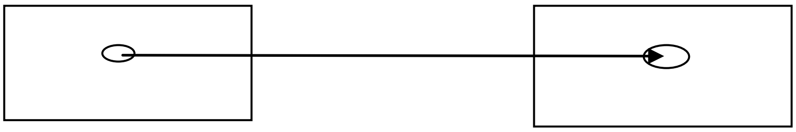
The main reason why automatic shot boundary detection is difficult is the fact that any kind of shot transition can be easily confused with camera and object motion which occurs in video anyway. A shot with much object motion throughout the frame such as a sports or action shot or a clip from a music video, can cause the false recognition of a shot boundary. Conventionally, if there exist frames that are merged by the adjacent shots but belong to neither of them, the transition is called a gradual one; otherwise, it is called a cut.[1]

II. PROPOSED ARCHITECTURE

For any video indexing, browsing, retrieval, representation and other video analysis technologies video shot boundary detection is the first and fundamental step.
To identify the transition between every two adjacent shots, video shot boundary detection is the process.

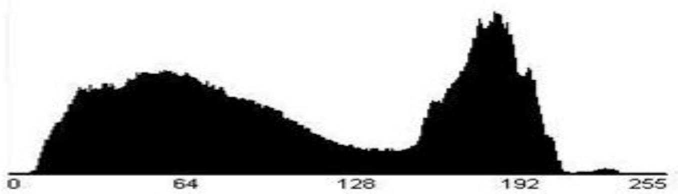
1. Pixel Comparison: In Pixel Comparison, if there two frames are significantly different and to count the number of pixels that change in value more than any threshold. This method is sensitive to camera motion. We note that manually adjusting the threshold is unlikely to be practical. This commonly used matching process duplicates the process used to extract motion vectors from an image pair. Then the pixel differences for each region were sorted, and then the weighted sum of the sorted region differences. The Gradual transitions were detected by generating a cumulative difference measures from consecutive values of the image differences. During dissolves and fades, this chromatic image assumes a reasonably constant value.

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 08 - Issue 09 || September 2022 || PP. 17-22



Frame N Frame N +1

2. Histogram Comparison: The histogram comparison methods are the most common method used to detect shot boundaries. The simplest histogram method computes, two types, gray level or color level histograms of any of the two images. If the bin-wise difference between the two histograms is above a threshold, a shot boundary is assumed

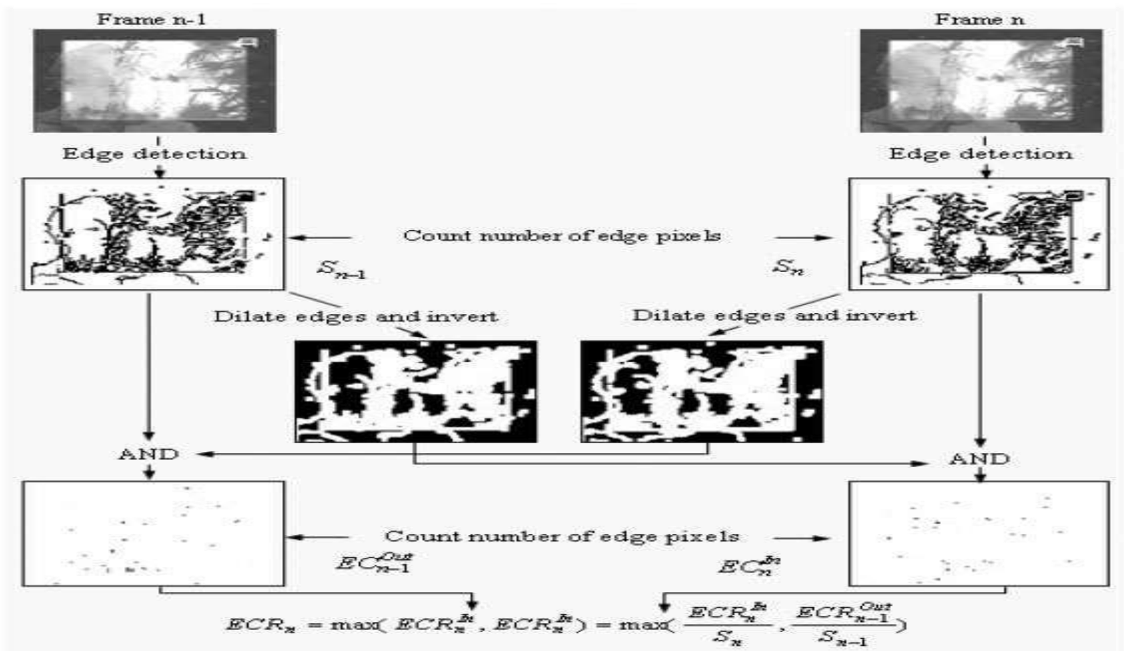


3. Statistical Differences: The statistical method is nothing but the idea of pixel differences by breaking the images into regions/blocks and comparing statistical measures of the pixels in those regions. It divides the frames into small regions. Then compares some of the few properties of every pixel in those regions between successive frames using the measurable statistical computation parameters.

4. Motion Vectors: MPEG compressed video sequences can also contains Motion vector information. The block matching performed as part of MPEG encoding based on compression efficiency and thus often selects inappropriate vectors for image processing purposes.

5. Edge Change Ratio: To detect, if any new edges have entered the image or if some old edges have disappeared, commonly uses the edges of successive aligned frames are detected first and then the edge pixels are paired with nearby edge pixels in the other image. The main reason why automatic shot boundary detection is difficult is the fact that any kind of shot transition can be easily confused with camera and object motion which occurs in video anyway. A shot with much object motion throughout the frame such as a sports or action shot or a clip from a music video, can cause the false recognition of a shot boundary.

In this project we report on shot boundary detection by using Convolutional Neural Network (CNN).



III. EUCLIDEAN DISTANCE ALGORITHM AND CHANGE POINT ANALYSIS ALGORITHM

Euclidean Distance algorithm:

The Euclidean distance is the simple distance of straight line between two points on a Euclidean space. Then the Euclidean space becomes a metric space with this distance. The associated norm is called the Euclidean norm. Older literature refers to the metric as Pythagorean metric. The Euclidean distance between points p and q is the length of the line segment that connects them In Cartesian coordinates, if p = (p1, p2, ..., pn) and q = (q1, q2, ..., qn) are two points in euclidean e-space, then the distance (d) of p to q, or q to p is given by the formula of Pythagoras:

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}.$$

The position of a point in a Euclidean n-space is a Euclidean vector. Then, p and q are Euclidean vectors, starting from the origin of space, and their suggestions indicate two points. The Euclidean norm, or Euclidean length, or the magnitude of a vector measures the length of the vector:

$$\|\mathbf{p}\| = \sqrt{p_1^2 + p_2^2 + \cdots + p_n^2} = \sqrt{\mathbf{p}\cdot\mathbf{p}},$$

whereas the last equation includes the dot product. A vector can be stated as a line segment started from the origin of the Euclidean space to vector point. If we consider that its length is really the distance from its tail to its tip, then the Euclidean norm of a vector is only a special case of Euclidean distance, whereas the Euclidean distance between its tail and its tip. The distance between points p and q can have an address (for example, from p to q), so it can be represented by another vector, given by

$$\mathbf{q} - \mathbf{p} = (q_1 - p_1, q_2 - p_2, \cdots, q_n - p_n).$$

One dimension

In Euclidean geometry, setting two points on a line and choosing one to be the origin to established metric. The length of the line segment between these points defines the unit of distance and the direction from the origin to the second point is defined as the positive direction. This line segment can be translated along the line to construct longer segments whose lengths correspond to multiples of the distance of the unit. In this way, real numbers can be associated with points on the line (such as the distance from the origin to the point) and these are the Cartesian coordinates of the points on what can now be called the real line. As an alternative way to set the metric, instead of choosing two points on the line, choose a point to be the origin, a unit of length, and an address along the line to call positive. The second point is determined uniquely as the point on the line that is at a distance from a positive unit of the origin. The distance between any two points on the real line is the absolute value of the numerical difference of its coordinates. It is common to identify the name of a point with its Cartesian coordinate. Therefore, if p and q are two points on the real line, then the distance between them is given by:

$$\sqrt{(q-p)^2} = |q-p|.$$

Two dimensions In the Euclidean plane, if p = (p1, p2) and q = (q1, q2) then the distance is given by

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}.$$

Three dimensions In three-dimensional Euclidean space, the distance is

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

***n* dimensions**

In general, for an n-dimensional space, the distance is

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2 + \cdots + (p_i-q_i)^2 + \cdots + (p_n-q_n)^2}.$$

Change Point Analysis:

Abrupt changes respect of time series is known as change point and detecting this point is Change point detection. The change point detection is applied in weather forecasting, detecting bank transaction fraud, image analyses, and classified data in data analysis, human activity and medical science. Many methodologies are applied to detect change point in time series. Since all methods are capturing Change Point, it is difficult to detect which method is appropriate for a particular field. So, that we are learning and getting the basic concept behind those methods.

E-divisive:

E-divisive detects change points by quantifying how different are the characteristic functions of the distributions of later segments of the time series. In fact, the characteristic functions uniquely describe a probability distribution, the changes in the distribution of characteristic distribution of the signal. This method combines the measure of multivariate divergence. First we discuss multivariate divergence measure here.

Measuring Differences in Multivariate Distributions:

For complex-valued functions φ(·), the complex conjugate of $\bar{\varphi}$ is denoted by φ, and the absolute square |φ|2 is defined as φφ The Euclidean norm of x ∈ Rd is |x|d, or simply |x| when there is no ambiguity. A primed variable such as Xj is an independent copy of X; that is, X and Xj are independent and identically distributed.

$$\int_{\mathbb{R}^d} |\phi_x(t) - \phi_y(t)|^2 \, w(t)\, dt,$$

In which w(t) denotes here an arbitrary positive weight function, for which the above integral exists. We use the following weight function,

$$w(t;\alpha) = \left(\frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^{\alpha}\Gamma((d+\alpha)/2)} |t|^{d+\alpha} \right)^{-1},$$

For some fixed constant α ∈ (0, 2). Then, if E|X|α, E|Y |α < ∞, a characteristic function based divergence measure may be defined as

$$\mathcal{D}(X,Y;\alpha) = \int_{\mathbb{R}^d} |\phi_x(t) - \phi_y(t)|^2 \left(\frac{2\pi^{d/2}\Gamma(1-\alpha/2)}{\alpha 2^{\alpha}\Gamma((d+\alpha)/2)} |t|^{d+\alpha} \right)^{-1} dt.$$

Suppose X, Xj , Fx and Y, Y j Fy, and that X, Xj, Y, and Y j are mutually independent. If E|X|α, E|Y |α < ∞, then we may employ an alternative divergence measure based on Euclidean distances,

$$\mathcal{E}(X,Y;\alpha) = 2E|X-Y|^{\alpha} - E|X-X'|^{\alpha} - E|Y-Y'|^{\alpha}.$$

Lemma 1, for any pair of independent random vectors X, Y ∈ Rd, and for any α ∈ (0, 2), if E(|X|α + |Y |α) < ∞, then E(X, Y ; α) = D(X, Y ; α), E(X, Y ; α) ∈ [0, ∞), and E(X, Y ; α) = 0 if and only if X and Y are identically distributed. The Lemma 1 motivates a simple empirical divergence measure for multivariate distributions based on a U -statistics. Let Xn = {Xi : i = 1, . . . , n} and Y m = {Yj : j = 1, . . . , m} be independent iid samples from the distribution of X, Y ∈ Rd, respectively, such that E|X|α, E|Y |α < ∞ for some α ∈ (0, 2). Then an empirical divergence measure analogous to Equation (4) may be defined as

International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 08 - Issue 09 || September 2022 || PP. 17-22

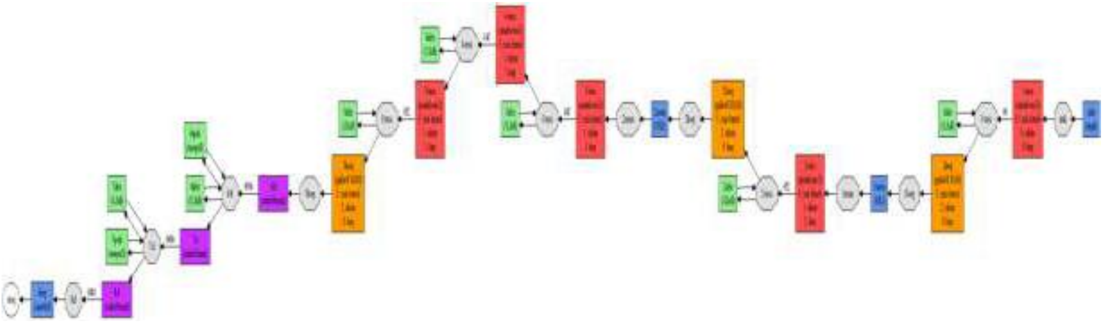
$$\widehat{\mathcal{E}}(\boldsymbol{X}_n, \boldsymbol{Y}_m; \alpha) = \frac{2}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}|X_i - Y_j|^\alpha - \binom{n}{2}^{-1}\sum_{1\le i<k\le n}|X_i - X_k|^\alpha - \binom{m}{2}^{-1}\sum_{1\le j<k\le m}|Y_j - Y_k|^\alpha.$$

Additionally, under the null hypothesis of equal distributions, i.e., E(X, Y ; α) = 0, we note that 2 mn E (Xn, Y m; α) converges in distribution to a non-degenerate random variable as m ∧ n → ∞. Further, under the alternative hypothesis of unequal distributions, i.e., E(X, Y ; α) > 0, we note that mn E (Xn, Y m; α) → ∞ almost surely as m ∧ n → ∞. These asymptotic Estimating the Location of a Change Point

$$\widehat{\mathcal{Q}}(\boldsymbol{X}_n, \boldsymbol{Y}_m; \alpha) = \frac{mn}{m+n}\widehat{\mathcal{E}}(\boldsymbol{X}_n, \boldsymbol{Y}_m; \alpha)$$

IV. PROPOSED CUSTOM CNN MODEL AND EXPERIMENTAL RESULT

Custom CNN model:



Video segment detection can be determined online and offline both ways. Here we proposed a new and efficient way of segment detection online/automatic. A video is consists of different segment of frames. However, each segment can be classified any of the following, i) Sharp/Hard transition ii) Gradual transition and iii) No transition

Experimental Result:

We used *Caffe* to extract and prepare image DB as learning data set. Also the code development of Change Point Detection algorithm is done partially.
Step 1: Extract video and prepare image DB(Train & Test data set)
Step 2: Calculate distance between two images by Euclidean Distance algorithm
Step 3: Apply Change Point Detection algorithm to determine the Hard-cut automatically

The graph generated on Euclidean Distances data from the above table mentioned in Appendix 1 in comparison with another graph generated by Change Point Detection algorithm on the above Euclidean Distances as input. This is to measure and show the efficiency and accuracy of our proposed approach.

Frame No.	Frame Name	method	Euclidean Distance
1	anni001_1.jpg	Cut	2.372311149
23	anni001_23.jpg	Hard cut	3.277518624
31	anni001_31.jpg	Gradual	1.766347775

Hard Cut detect

Gradual Cut detect

V. CONCLUSION

This work introduces an efficient and robust system for detecting video scene changes, an essential task in fully content analysis systems. Our module receives frame differences as inputs, then recalls the information stored into the neural network weights to determine the outputs. The algorithm has been tested on varieties of videos. Better generalization of the neural network can be achieved by increasing the number of video clips used in the training phase and by varying their contents. The effectiveness of the proposed paradigm has been proven as a robust and efficient way to identify scene changes in any type of compressed video streams.

REFERENCES

[1]. Nikita Sao, Ravi Mishra, A survey based on Video Shot Boundary Detection techniques https://github.com/BVLC/caffe/
[2]. http://caffe.berkeleyvision.org/
[3]. Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv preprint arXiv:1408.5093, 2014.
[4]. [4]Waleed E. Farag, University of Pennsylvania, USA Hussein Abdel-Wahab, Old Dominion University, USA, Video Shot Boundary Detection
[5]. http://mathonline.wikidot.com/the-distance-between-two-vectors
[6]. Christian Szegedy, Wei Liu, Yangqing Jia , Pierre Sermanet, Scott Reed, Dumitru Erhan, Vincent Vanhoucke. Andrew Rabinovich ,Going deeper with convolutions
[7]. http://searchnetworking.techtarget.com/definition/neural-network
[8]. https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm
[9]. https://www.slideshare.net/nilmani14/neural-network-3019822
[10]. John S. Boreczky Lawrence A. Rowe , Comparison of video shot boundary detection techniques , Journal of Electronic Imaging 5(2), 122–128 (April 1996)