ARPN Journal of Engineering and Applied Sciences

FUZZY RULE BASED MODEL FOR SEMANTIC CONTENT EXTRACTION IN VIDEO BIG DATA

A. Manju and P. Valarmathie
Department of Computer Science and Engineering, Saveetha Engineering College, Chennai, India
E-Mail: manjuappukuttan1985@gmail.com

ABSTRACT

Recent increment in the utilization of video-based applications has unveiled the requirement for extracting the substance in videos. Street crime is expanding as of late, which has requested more solid and smart open conservative framework. Raw information and low-level elements alone are not adequate to satisfy the client's needs that is, a more profound comprehension of the substance at the semantic level is needed. Manual procedures, which are wasteful, subjective and expensive in time and limit the questioning abilities, are being utilized to bridge the gap between lower-level delegate components and higher-level semantic substance. It is fundamental to portion the video information into important pieces as image frame using image processing. To recognize important video data as useful big data, it is necessary to associate information from every methodology. In order to achieve this, Video Semantic Substance Extraction Framework was initiated to extract the objects, events and ideas consequently from videos through the previously mentioned procedure. With video analytics it is possible to track movement, size, speed, shape and directions of objects. In this video semantic substance model fuzzy rule based procedures are used to accomplish preferable outcome.

Keywords: video analytics, video semantic substance model, fuzzy rule, image processing.

1. INTRODUCTION

Huge numbers of videos are uploaded everyday into sites like YouTube, Facebook, and Whatsapp from devices like mobile phones, personal computers and home surveillance cameras. With the new technology, it is possible to mine visual data to obtain valuable insights about world. At present, extracting data from video was done manually through human observation. Current technology use metadata or tags with videos which are stored with videos when the video was uploaded. Picture annotation has reflected the semantic gap between video data and original data. Picture annotation was categorized between two principle classifications: idea based picture recovery and substance based picture recovery. The previous spotlights were on recovery by picture articles and higher-level ideas, while the last spotlight on the lower-level visual element of the picture [1]. Division by region intends to separate picture parts into various areas sharing regular properties. These techniques register a general similitude among pictures in light of statistical picture properties and basic cases of such properties are surface and color where these strategies are observed to be strong and proficient [2].

Semantic understanding of scenes remains an essential research challenge for the picture and video recovery community [3]. Representation and semantic annotation of multimedia content have been distinguished as critical stages towards more effective control and recovery of visual media. Even though new multimedia media principles, like, MPEG-4 and MPEG-7, give necessary functionalities to the control and transmission of items and related metadata, the extraction of semantic depictions and annotation of the substance with the corresponding metadata is out of the extension. This propels overwhelming examination attempt towards the automatic annotation of multimedia content [4].

The basic contrast between substance based and text based recovery frameworks is that the human collaboration is a key part of the last framework. People tend to utilize higher-level highlights (ideas), like, keywords, content descriptors, to translate pictures and measure their likeness. While the components consequently get separated utilizing computer vision strategies are for the most part lower-level elements (colour, surface, shape, spatial design, and so forth) [5]. Knowledge representation and semantic annotation of multimedia content have been distinguished as necessary stages towards the most effective control and recovery of visual media. Today, new multimedia benchmarks, like, MPEG-4 and MPEG-7, give vital functionalities to control and transfer the articles and related metadata. The extraction of semantic substances and annotation of the substance with the related metadata, however, is out of the extent of these measures and still left to the substance manager. This persuades overwhelming exploration attempt toward programmed annotation of multimedia substance [6].

Extraction of objects and applying semantics can help a wide range of uses in the photograph recovery area, which includes 1] Improved picture search through gathered query semantics; 2] Automated production of place and occasion gazetteer information that can be utilized; 3] Web search by recognizing applicable spatial areas and time ranges for specific keywords; 4] Generation of photograph gathering representations by area and additionally occasion/time; 5] Support for tag recommendations for photographs (or different assets) on the basis of area and time of capture; 6] Automated relationship of missing area/time metadata to photographs, or different assets, is on the basis of the labels or caption content [7].

ARPN Journal of Engineering and Applied Sciences

2. RELATED WORK

Amjad Rehman *et al* [8] have exhibited a state of art which audit the component extraction for soccer video summarization research. The existing methodologies with respect to object recognition, video summarization in light of video stream and utilization of content sources in occasion location have been reviewed. Sound, video feature extraction techniques and their blend with textual strategies were researched.

K. Karsch *et al* [9] has composed a method that naturally creates conceivable profundity maps from videos utilizing non-parametric profundity sampling. The system is applicable to single pictures and additionally videos. Nearby movement cues were utilized to enhance the inferred profundity maps, while optical flow was utilized to guarantee temporal consistency for videos.

Suet-Peng Yong *et al* [10] has initiated a structure that model semantic contexts for key-outline extraction. Semantic context of video frames was extracted and its successive changes were observed so that noteworthy novelties were found utilizing a one-class classifier. Working with wildlife video outlines, the system experiences picture division, highlight extraction and matching of picture blocks, and after that a co-occurrence network of semantic labels was built to show the semantic context inside the scene.

Y. Yildirim *et al* [11] has outlined a semantic content extraction framework that permitted the client to question and recover articles, objects and ideas that are extricated consequently. A fuzzy video semantic substance model based on ontology was presented that utilizes spatial/ temporal relations in occasion and idea definitions. This meta ontology definition gave a wide- domain applicable principle with construction standard that permit to build ontology for a provided domain.

Amjad Altadmri *et al* [12] have built up a system for the Automatic Semantic Annotation of unconstrained videos. The initiated system uses two non- domain particular layers: lower-level visual closeness matching, and a annotation investigation that utilizes common sense knowledge bases. Common-sense ontology was made by joining different organized semantic connections. N. Inoue et.al [13] has presented a quick maximum a posteriori (MAP) adjustment technique for video semantic ordering that utilize Gaussian mixture model (GMM) super vectors. In this technique, a tree-organized GMM was used to reduce the computational cost, where just the yield probabilities of blend parts near an information test were exactly calculated.

3. VIDEO SEMANTIC SUBSTANCE EXTRACTION FRAMEWORK

The enormous development of online videos, video thumbnail, as the basic representation type of video substance, is turning out to be progressively vital to impact client's browsing and searching experience. Nonetheless, conventional techniques for video thumbnail choice regularly neglect to deliver fulfilling outcomes as they overlook the semantic data (e.g., title, depiction, and question) connected with the video. Subsequently, the chosen thumbnail can't generally show the video semantics and the click-through rate is not favoured that is influenced when the recovered videos are related. Hence, a semantic substance extraction framework was built that permits the client to question and recover articles, occasions, and ideas naturally. An ontology-based fuzzy video semantic substance model that utilizes the idea definitions was introduced. Because of absence of the above said issue, inspires us to do research in this region.

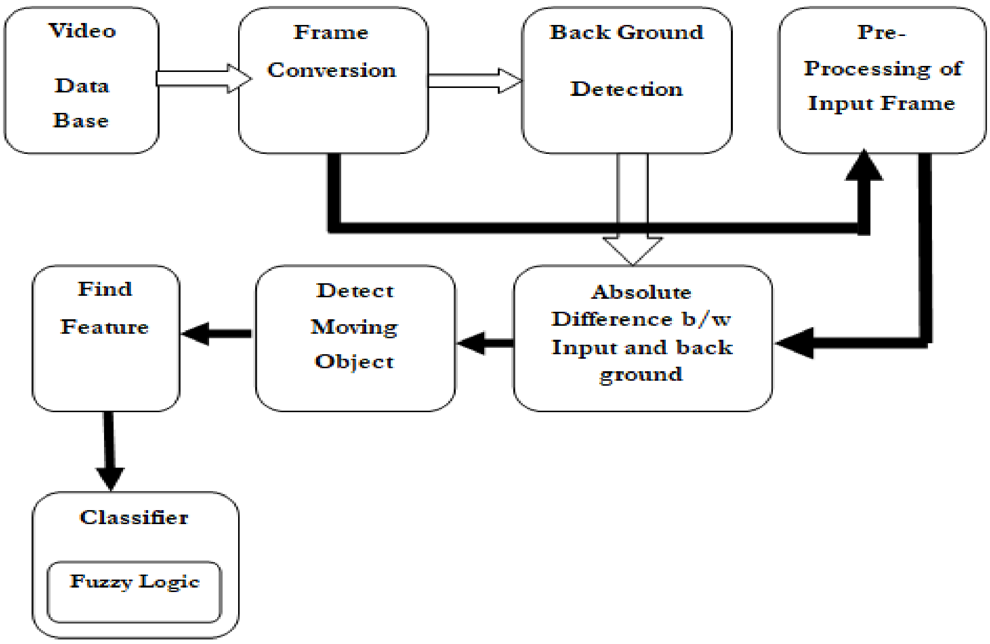

Figure-1. Video semantic substance extraction framework.

VOL. 13, NO. 9, MAY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

In the present days, for the most part video based applications are very alluring and utilized as a part of numerous applications. Subsequently, these applications require semantic search and extraction of multimedia contents. Hence, to audit the video it is essential to distinguish the video occasions. By utilizing raw information and lower-level elements client can't accomplish the entire need and to understand the video profoundly is more vital. Video Semantic Content Meta Model was introduced for extraction of articles, occasions and ideas naturally through the previously mentioned procedure. This video semantic substance model depends on fuzzy rule based process. Extraction procedures begin with object extraction and afterward objects are utilized as input for occasion extraction. Subsequent to creating rules utilizing fuzzy strategy it was necessary to discover the occasion discovery in view of the weights it produces and after that the idea extraction happens. The initiated strategy is executed in the working stage of MATLAB and the outcomes were examined. Additionally, examinations with existing technologies were likewise given to evaluate the execution of initiated strategy.

Database: The database consists of the video for working with particular applications. Database was used to portray accumulations of videos utilized as a part of this initiated work. Beforehand test information was troublesome, yet the invention of modern computerized gadgets has simplified acquiring information.

Object extraction: The extraction process starts with extracting the object. Particularly, an object extraction approach is used for the object extraction and classification needs. From each individual frame, objects and spatial association between objects are extracted. Initially the video contents and components to model are to be identified. Input frames are the basic video units which are in image format, extracted from raw video data that best represent the content of shots in an abstract manner. To extract an object a semiautomatic Genetic Algorithm-based object extraction approach is used.

$$MemberShip : \begin{cases} [\mu \Rightarrow [float]] \\ where \\ 0 \leq \mu \leq 1. \end{cases}$$

$$MBR : \left\{ \begin{bmatrix} x \Rightarrow [integer],\ y \Rightarrow [integer], \\ width \Rightarrow [integer],\ length \Rightarrow [integer] \end{bmatrix} \right.$$

$$ObjectInstance : \begin{cases} \begin{bmatrix} frameNo \Rightarrow [number], \\ minBoundingRectangle \Rightarrow \{MBR_i\}, \\ membership \Rightarrow \{MSV_j\}, \\ objectType \Rightarrow \{O_k\} \end{bmatrix} \\ where \\ ind(O_k, Object), \\ ind(MBR_i, MBR), \\ ind(MSV_j, MemberShip). \end{cases}$$

The approach is a supervised learning approach utilizing eight MPEG-7 descriptors to represent the objects. Throughout the object extraction method, for every delegate key frame in the video clip, above-said object extraction process is performed and a set of objects were extracted and classified. The extracted object instances are stored with their type, frame number, membership value, and Minimum Bounding Rectangle data. Object instances are used as input with the domain ontology's in the event and concept extraction process.

Frame conversion: The video's blend of number of frames that are represented with certain frequency to stay away from flickers. Persistence of vision happens in video which means original perception of a real time object which does not appear for some time. The pictures of the object get saved in human eye for 1/16th of second. At first 24 frames were utilized to make video however event of flickers becomes possibly the most important factor later on frame frequency. Here in Frame transformation reverse procedure happens. Frames are taken out from the video.

A. Background and input frame detection: The foundation and frame data are the two components in video which appears to be single component and need to filter out both for discovery of the frame data. Here frame foundation is recognized on the stage and frame data is recovered from upcoming stage. The identification execution is enhanced by building a decent foundation demonstration. They likewise utilized parametric probability density works as weighted sum of Gaussian models keeping in mind the end goal to section the frontal area pixels from the foundation. Foundation location and frame data is contrasted with the visual distinction between them. This stage separates the moving items from the video.

The identification of moving areas is the underlying procedure in object acknowledgment. The point of moving the item identified is separating moving items in picture sequences which generally join with the foundation pixel. After the discovery of the moving items their highlighted estimations are gotten. For instance contrast, correlation, homogeneity, energy, shape, area, perimeter, filled area and eccentricity. For the division fuzzy logic is actualized.

B. Background model generation: The foundation is displayed as autonomous statistical approach at pixel wise level, compactness, energy of the areas and homogeneity on every picture sequences. The parameter circulation thus fluctuates for each pixel hence the thickness capacity of this distribution is restored by their coordinating colour area and frame a locale map of the picture. The parameter likelihood thickness capacity is evaluated by numbering the occurrence rate of every parameter level in the area to create spatial appropriation. It is further processed to get the aggregate region under the curve equivalent to facilitate the compactness parameter using the territory and perimeter of the area, because of which in this proposed technique most extreme number of the parameters are employed to distinguish the moving component.

In the video scene, the foundation changes relying upon the scene, and it stays new to the framework. Subsequently, at first a foundation model is created through a preparation stage, which requires a few introductory picture sequences with no moving component. Initially the parameters of the method are produced by the framework mean, every pixel is displayed as a distribution of Gaussian blend models. The likelihood value for every pixel can be composed as:

$$P_r(X_t) = \sum_{i=1}^{K} M_i . X_t . C_i . E_{ri} H_i \sum i,t. \tag{1}$$

Where K is the quantity of distribution that determined by the accessible computational memory which default K=3, C is the compactness parameter of the Kth Gaussian model, M is a likelihood thickness capacity, E is the energy that indicate the pixel power value, H signifies the homogeneity and $\sum i,t.$ is the covariance matrix of the Kth distribution.
Gaussian capacity for a parameter can be represented as:

$$M_i(X_t, C_i, \sum i,t.) = \frac{1}{2\Pi^{\frac{D}{2}} \, |\sum i,t.|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - C_i)^t \sum i,t.(X_t - C_i)} \tag{2}$$

The approach is on the basis of the assumption that the data of pixel parameter channel that are autonomous and it possesses the similar differences.

$$\text{Compactness } C_i = \frac{A}{P^2} \tag{3}$$

Where, A is the area, P is the perimeter.

$$\text{Energy } E_r = \sum_{b=0}^{N-1} \varepsilon . e(x)^2 \tag{4}$$

$$\text{Where } e(x) = \frac{n(b)}{m} \tag{5}$$

Where m shows the aggregate quantity of pixels in a nearby window centred about (j, k)and n(b) is the quantity of pixels of amplitude in the similar window.

$$\text{Homogeneity } H = \sum_i \sum_j \frac{p[i,j]}{1 + |i - j|} \tag{7}$$

Where, H represents the homogeneity, P [i, j] is the gray-level co-occurrence matrix.

C. Background subtraction: The parameter's initialization for foundation method has been made; the built model is utilized to compute the distinction with tried frame in which it will relate to the moving area of interest. The foundation pixel will seem more continuous than foreground pixels. On the off chance that the present pixel matches with any circulation method and satisfies the Equation underneath,

$$B = \arg[\min_k (\sum_{i=1}^{k} c_i > TH)] \tag{8}$$

Where, $C_i$is the mixture compactness, TH is the threshold value.

On the off chance that the pixel did not match with any distribution, the rank mode will be supplanted with $X_t = \mu_{i,t}$ a new one with and it will be set as forefront pixel. Every pixel was looked at against the blend of Gaussian method. The threshold in forefront division stage will profit the framework as far as computational cost, which is fundamental with a specific end goal to accomplish the real time execution.

After extracting the object, knowledge conduction during information extraction and the sustainability of object was accessed by applying fuzzy reasoning. The first action line is related to the establishment of techniques for the dynamic management of video analysis based on the knowledge gathered in the semantic network. This technique supports the decisions taken during the analysis process. This procedure is based on a set of rules that are able to handle the fuzziness of the annotations provided by the analysis modules and gathered in the semantic network.

Network architecture of fuzzy logic: Fuzzy Logic (FL) is a technique of reasoning that is as

ARPN Journal of Engineering and Applied Sciences

comparative as human reasoning. The approach of FL emulates the method for basic leadership in humans that includes all middle possibilities between digital values YES and NO. The computational yields of computer on the premise on genuine or false, comparatively people follows if and else criteria.

The designer of fuzzy logic, Lotfi Zadeh, watched that dissimilar computers, the human decision making incorporates a scope of possibilities amongst YES and NO. It can be executed in frameworks with different sizes and capacities extending from micro- controllers to large, networked, workstation-based control frameworks. The learning ability of AND was completely controlled for programmed IF-THEN principles era and parameter optimization of fuzzy framework, which for the most part sees as the basic issue of fuzzy framework. Three interference frameworks were freely created for the expected moving component classes (weight lifter, football, and car).

On the premise of first-order Takagi-Sugeno-Kang (TSK) strategy, the outcome parameter is in linear condition terms. Each of them contained three input nodes of A1-A3, 27 guidelines of TSK, and one yield variable, Z. Every fuzzy is in the structure of IF-THEN, as appeared in condition beneath:

Rule D: If A_i is X_i^D and A_{i+1} is X_{i+1}^D then,

$$Z_D = a_0^D + \sum_{j=1}^{D} a_j^D A_j \tag{9}$$

Where, X_i are the member functions, A_i are the inputs, a_j^D are the parameters of consequent equations.

The structure of the framework for every class is as appeared in below figure.

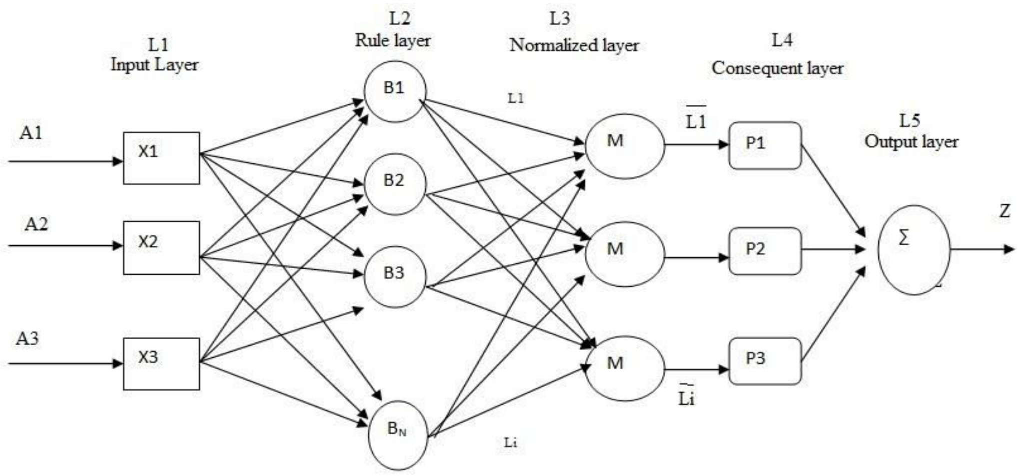

Figure-2. Network structure.

There are five layers of in TSK methods and capacity of every model is characterized underneath.

Layer 1 (L1): The nodes in Input linguistic layer form the linguistic factors, and working as TSK lead bases. Every node plays out a participation value calculation. Gaussian capacity is utilized as membership function to compute the level of enrolment value.

Layer 2 (L2): Rules layer. This layer is playing out the algebraic item operation of the all the capacities acquired from the past layers.

Layer 3 (L3): Normalized layer. This layer constitutes of settled nodes that compute the proportion of the N^{th} terminating quality, L_i to the aggregate of all terminating qualities. The standardized terminating quality is given by,

$$\overline{L}_i = \frac{L_i}{\sum_{k=1}^{D} L_k} \tag{10}$$

Where D is total quantity of rules

Layer 4 (L4): Consequent layer. Each node in this layer is a versatile node which computes the outcome value, O_i given by equation beneath.

$$\overline{L}_i Z_n = \overline{L}_i (a_0 + a_1 A_1 + a_2 A_2 + ... + a_n A_n) \tag{11}$$

Where,

$\overline{L}_i$, the normalized firing strength from layer 3 and $(a_0 + a_1 A_1 + a_2 A_2 + ... + a_n A_n)$ is the parameters of these nodes.

Layer 5(L5): Result linguistic layer. It incorporates a fixed hub indicated as summation of that the capacities as a summation of general yield network. This infers a defuzzification operation so as to acquire the crisp value. By the Weighted Fuzzy Mean (WFM) technique, the general yield Z is gotten by,

VOL. 13, NO. 9, MAY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

$$Z = \sum_i \bar{L}_i Z_n = \sum_i \frac{L_i Z_i}{L_i} \tag{12}$$

The classification algorithm alters the resulting parameters of layer 4 in feed-forward propagation. While, regressive propagations were connected iteratively to reduce the errors.

4. RESULTS AND DISCUSSION

The proposed multi-component tracking mechanism is actualized in MATLAB 7.11.0(R2010b) with i5 processor and 4GB RAM. The datasets utilized for testing the tracking system are football, auto, bar ball video arrangements. The datasets are exceptionally challenging due to the overwhelming inter-person objects and poor picture differentiate amongst components and foundation. The algorithm was accessed on its tracking execution and it was noticed that the detection execution has contrasted our outcomes and existing strategy. Multi-component tracking for the most part confronts three difficulties: component switch among overlapping, new component introduction and re-acknowledgment of re-entering objects. In the accompanying part, it quickly presents two videos and after that discuss about the outcomes as far as previously mentioned challenges.

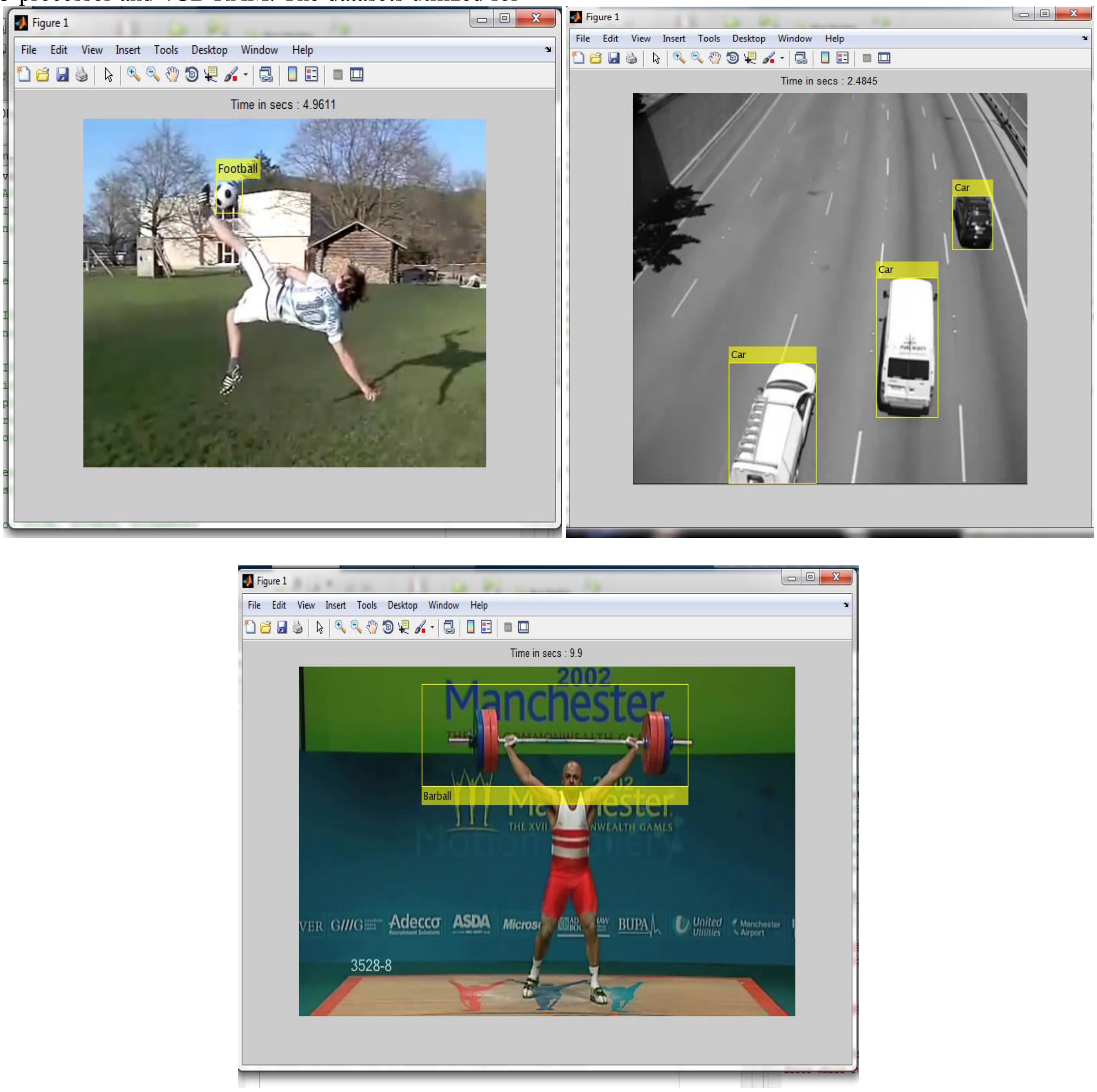

Figure-3. The tracking outcome on football, car, bar ball.

Experimental settings: To track all items all through the benchmark successions, the initiated tracking algorithm depends on a few intuitive parameters. Specifically, the accompanying default parameter settings for our investigations were utilized.

Metrics: The broadly acknowledged execution metrics Multiple Object Tracking Accuracy (MOTA) and Precision (MOTP) method was utilized. The precision metric MOTP assesses the arrangement of genuine positive directions as for the ground truth, though the

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

precision metric MOTA joins 3 error proportions, in particular false positives, false negatives (i.e., missed components), and identity switches.

Give s_n^i be the distance between the assessed outcome and the ground truth for component i at time n and b_n the quantity of matches discovered, further then, MOTP was represented as:

$$MOTP = \frac{\sum_{i,n} s_n^i}{\sum_n b_n} \tag{13}$$

The distance d_n^i is really the covering between the evaluated bounding box and the ground truth. Subsequently, higher estimations of MOTP demonstrate better outcomes.

For the MOTA, let w_n be the quantity of items that exist at moment n. Let likewise k_n, ht_n and kke_n be the quantity of misses, false positives, and jumbles, separately. At that point, the metric can be gotten by,

$$MOTA = 1 - \frac{\sum_n (k_n + ht_n + kke_n)}{\sum_n w_n} \tag{14}$$

The precision metric correlation using initiated fuzzy and existing neural system is represented in Table-1. Precision value acquired for car picture utilizing existing technique is0.897237while initiated has 0.945792 comparatively for Football, Barball pictures precision value computed utilizing existing strategy is0.916324, 0.907507 yet initiated has0.975404, 0.964889 individually. It can be clear from the above discussion that, the initiated technique performs powerfully and it has ability to deliver the outcomes which are near the ground truth, regardless of the identities that the components are swapped.

Table-1. Performance measures of MOTP.

Precision		
Dataset	**Neural**	**Fuzzy**
Car	0.897237	0.945792
Football	0.916324	0.975404
Barball	0.907507	0.964889

The Accuracy metric comparison using initiated fuzzy and existing neural system is represented in Table2. Accuracy value got for car picture utilizing existing strategy is0.912699while initiated has 0.963376 likewise for Football, Barball pictures Accuracy value figured utilizing existing technique is 0.92785, 0.915761but initiated has0.954688, 0.970593respectively. Our initiated technique is better in assigning tracks to the right component, without taking into account how near it really is from the right position.

Table-2. Performance measures of MOTA.

Accuracy		
Dataset	**Neural**	**Fuzzy**
Car	0.912699	0.963376
Football	0.92785	0.954688
Barball	0.915761	0.970593

The underneath graph demonstrate the precision comparison of dataset utilizing initiated fuzzy algorithm with existing neural system procedure for multi component tracking. From the graph precision value of initiated technique is more contrasted with existing strategy. So the quantity of significant objects tracked using initiated technique is better.

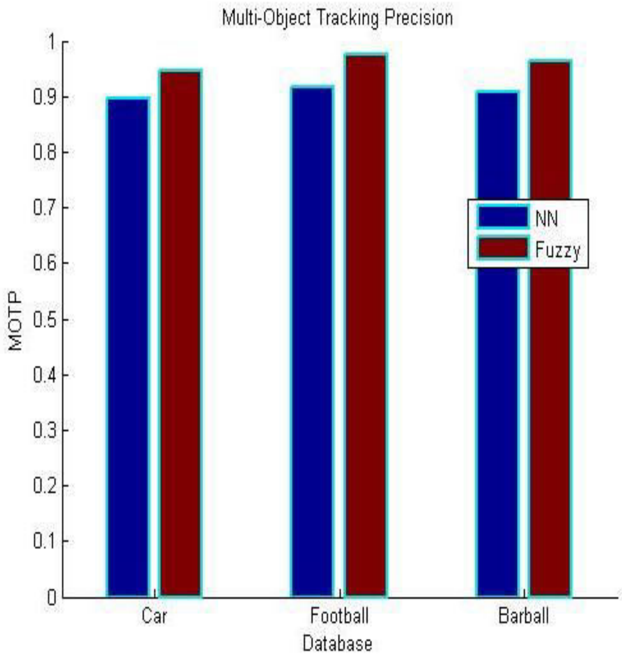

Figure-4. Performance graph of precision.

The beneath graph demonstrate the accuracy comparison of dataset utilizing initiated fuzzy algorithm with existing neural system method for multi component tracking. Our initiated strategy has higher values contrasted with existing technique.

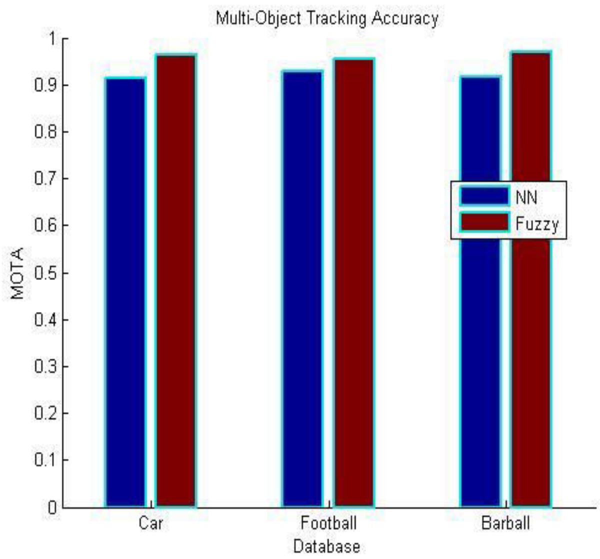

Figure-5. Performance graph of precision.

5. CONCLUSIONS

In this paper, a fuzzy system and its features to new summarization were displayed. Video semantics has been removed utilizing Fuzzy logic in light of its principles. In this framework, the layers of Fuzzy system to give the flow to extraction procedures were represented. The target, recognition of moving component has been performed by two systems, the initiated fuzzy logic system and neural system. The results have been analyzed for both of the methods, where the initiated Video Semantic Substance Extraction Framework have accomplished preferable outcome over existing procedure.

REFERENCES

[1] Amjad Rehman and Tanzila Saba. 2014. Features extraction for soccer video semantic analysis: current achievements and remaining issues. Journal of Medical Imaging and Health Informatics. 41: 451-461.

[2] K. Karsch, C. Liu and S. B. Kang. 2014. Depth Extraction from Video Using Non-parametric Sampling. In the proceedings of: IEEE Transactions on Pattern Analysis and Machine Intelligence. 36: 2144-2158.

[3] Suet-Peng Yong, Jeremiah D. Deng and Martin K. Purvis. 2013. Wildlife video key-frame extraction based on novelty detection in semantic context. Journal of Multimedia Tools and Applications. 62: 359-376.

[4] Y. Yildirim, Ankara, Turkey, A. Yazici and T. Yilmaz. 2013. Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model. In the proceedings of IEEE Transactions on Knowledge and Data Engineering. 25: 47-61.

[5] Amjad Altadmri and Amr Ahmed. 2014. A framework for automatic semantic video annotation. Journal of Multimedia Tools and Applications. 72: 1167-1191.

[6] N. Inoue and K. Shinoda. 2012. A Fast and Accurate Video Semantic-Indexing System Using Fast MAP Adaptation and GMM Supervectors. In the proceedings of: IEEE Transactions on Multimedia. 14: 1196-1205.

[7] W. Liu, T. Mei, Y. Zhang and C. Che. 2014. Multi-Task Deep Visual-Semantic Embedding for Video Thumbnail Selection. In the proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3707-3715.

[8] R. C. F. Wong and C. H. C. Leung. 2008. Automatic Semantic Annotation of Real-World Web Images. In the proceedings of: IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 30.

[9] Julia Vogel and BerntSchiele. 2007. Semantic Modeling of Natural Scenes for Content-Based Image Retrieval. International Journal of Computer Vision. 72: 133-157.

[10] Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis Avrithis, Siegfrie Handschuh, Yiannis Kompatsiaris, and Steffen Staab. 2005. Semantic Annotation of Images and Videos for Multimedia Analysis. In the proceedings of: Semantic Web Conference. 3532: 592-607.

[11] Ying Liua, Dengsheng Zhanga, GuojunLua and Wei-Ying Mab. 2007. A survey of content-based image retrieval with high-level semantics. Journal of Magnetic Resonance in Medicine. 40: 262-282.

[12] A. Cavallaro, O. Steiger and T. Ebrahimi. 2005. Semantic video analysis for adaptive content delivery and automatic description. In the proceedings of: IEEE Transactions on Circuits and Systems for Video Technolog. 15: 1200-1209.

[13] K. Petridis, S. Bloehdorn, C. Saathoff and N. Simou. 2006. Knowledge representation and semantic annotation of multimedia content. In the proceedings

of: IEE Proceedings - Vision, Image and Signal Processing. 153: 255-262.

[14] Tye Rattenbury, Nathaniel Good and Mor Naaman. 2007. Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In the proceedings of: conference on Research and development in information retrieval. pp. 103-110.

[15] R. T. Collins, A. J. Lipton, and T. Kanade. 2000. Introduction to the specialsection on video surveillance. IEEE Trans. Pattern Anal. Machine Intell. 22: 745-746.

[16] J. Steffens, E. Elagin and H. Neven. 1998. Person spotter-fast and robust system for human detection, tracking and recognition. in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition. pp. 516-521.

[17] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson. 2000. A system for video surveillance and monitoring. Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep., CMU-RI-TR-00-12.

[18] Haritaoglu, D. Harwood, and L. S. Davis. 2000. W: Real-time surveillance of people and their activities. IEEE Trans. Pattern Anal. Machine Intell. 22: 809-830.

[19] Chuanping Hu, Zheng Xu, Yunhuai Liu, Lin Mei, Lan Chen and Xiangfeng Luo. 2013. Semantic Link Network based Model for Organizing Multimedia Big Data. IEEE Transactions on Emerging Topics in Computing.

[20] Anil Aksay, Alptekin Temizel and A. Enis Cetin. 2007. Camera tamper detection using wavelet analysis for video surveillance. In IEEE Int. Conf. on Advanced Video and Signal Based Surveillance, AVVS 2007, pp. 558-562.

[21] Irena Koprinska, Sergio Carrato. 2001. Temporal video segmentation: A survey, Elsevier, Signal processing: Image Communication. pp. 477-500.

[22] Moeslund, Thomas B. 2012. Introduction to video and image processing, Springer.

[23] X. Lu, Li Song, Songyu Yu, Nam Ling. 2012. Object Contour Tracking Using Multi-feature Fusion based Particle Filter. IEEE Conference on Industrial Electronics and Applications (ICIEA). pp. 237-242.

[24] Huitao Luo, alexandros Eleftheriadis. 2002. An interactive authoring system for video object segmentation and annotation, Elsevier, Signal processing: Image Communication. pp. 559-572.

[25] L. Wu and Y. Wang. The process of criminal investigation based on grey hazy set. 2010 IEEE International Conference on System Man and Cybernetics, pp.26-28, 2010.

[26] L. Liu, Z. Li, and E. Delp. 2009. Efficient and low-complexity surveil-lance video compression using backward-channel aware wyner-ziv video coding. IEEE Transactions on Circuits and Sys-tems for Video Technology. 19(4): 452-465.

[27] H. Yu, C. Pedrinaci, S. Dietze and J. Domingue. 2012. Using linked data to annotate and search educational video resources for supporting distance learning. IEEE Transactions on Learning Technologies. 5(2): 130-142.

[28] C. Xu, Y. Zhang, G. Zhu, Y. Rui, H. Lu and Q. Huang. 2008. Using webcast text for semantic event detection in broadcast sports video. IEEE Transactions on Multimedia. 10(7): 1342-1355.

[29] H. Zhuge. 2009. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. IEEE Transactions on Knowledge and Data Engineering. 21(6): 785-799.

[30] X. Luo, Z. Xu, J. Yu and X. Chen. 2011. Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering. 8(3): 482-494.

[31] T. Berners-Lee, J. Hendler and O. Lassila. 2001. The Semantic Web. Scientific American. 284(5): 34-43.

[32] Yakup Yildirim, Adnan Yazici and Turgay Yilmaz. 2013. Automatic Semantic Content Extraction in Videos Using a Fuzzy Ontology and Rule-Based Model. IEEE Transactions on Knowledge and Data Engineering. 25(1).

[33] M. Petkovic and W. Jonker. 2000. An Overview of Data Models and Query Languages for Content-Based Video Retrieval. Proc. Int'l Conf. Advances in Infrastructure for E-Business, Science, and Education on the Internet.

[34] M. Petkovic and W. Jonker. 2001. Content-Based Video Retrieval by Integrating Spatio- Temporal and Stochastic Recognition of Events. Proc. IEEE Int'l Workshop Detection and Recognition of Events in Video. pp. 75-82.

[35] G.G. Medioni, I. Cohen, F. Bre´mond, S. Hongeng and R. Nevatia. 2001. Event Detection and Analysis from Video Streams. IEEE Trans. Pattern Analysis Machine Intelligence. 23(8): 873-889.

[36] S. Hongeng, R. Nevatia and F. Bre´mond. 2004. Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods. Computer Vision and Image Understanding. 96(2): 129-162.

[37] A. Hakeem and M. Shah. 2005. Multiple Agent Event Detection and Representation in Videos. Proc. 20th Nat'l Conf. Artificial Intelligence (AAAI). pp. 89-94.

[38] T. Yilmaz. 2008. Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach. Master's thesis, Computer Eng. Dept., METU, Turkey.

[39] G. Salton, A. Wong and C. Yang. 1975. A vector space model for automatic indexing. Communications of the ACM. 18(11): 613-620.

[40] Z. Xu, X. Luo, J. Yu, and W. Xu. 2011. Measuring semantic similarity be-tween words by removing noise. Concurrency and Computation: Prac-tice and Experience. 23(18): 2496-2510.

[41] R. Firth. A synopsis of linguistic theory 1930-1955. In Studies in Lin-guistic Analysis. Philological Society: Oxford, 1957.

[42] M. Vojnovic, J. Cruise, D. Gunawardena and P. Marbach. 2009. Ranking and suggesting popular items. IEEE Transactions on Knowledge and Data Engineering. 21(8): 1133-1146.

[43] H. Rubenstein and B. Goodenough. 1965. Contextual correlates of synon-ymy. Communications of the ACM. 8(10): 627-633.

[44] M. Steinbach, G. Karypis and V. Kumar. 2000. A Comparison of Document Clustering Techniques. KDD Workshop on Text Mining.

[45] L. Wang and S. Khan. 2013. Review of performance metrics for green data centers: a taxonomy study. The journal of supercomputing. 63(3): 639-656.

[46] L.Wang, D. Chen, *et al*. 2013. Towards enabling cyber infrastructure as a service in clouds. Computer & Electrical Engineering. 39(1): 3-14.

[47] L. Wang, J. Tao, et al. 2013. G-Hadoop: MapReduce across distributed data centers for data-intensive computing. Future Generation Computer Sys-tems. 29(3): 739-750.

[48] H. Zhuge. 2010. Interactive Semantics. Artificial Intelligence. 174: 190-204.

[49] H. Zhuge. 2012. The Knowledge Grid -- Toward Cyber-Physical Society, World Scientific Publishing Co., Singapore. 2nd Edition.

[50] H. Zhuge, X. Chen, X. Sun and E. Yao. 2008. HRing: A structured P2P overlay based on harmonic series. IEEE Transactions on Parallel and Distributed Systems. 19(2): 145-158.

[51] A.P. Pons. 2006. Object Prefetching Using Semantic Links. ACM SIGMIS Database. 37(1): 97-109.