International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

An Effective Interest Identification Technique to Enhance Sales Performance in Supermarkets

G. Priyanka, S. Sathya Bama, VJ Aiswaryadevi, M.S Sruthi, S. Soundarya

Abstract: ***Surveillance Camera System is installed in the supermarkets mainly for security purposes. But the main idea of this paper is to use this surveillance camera system to improve the sales performance by targeting a particular stimulus (child) through marketing promotions. The owner of the supermarket monitors the entire store with the help of the security camera system. The owner suddenly finds an abnormal action in a stimulus (child) on looking at a particular product. On observation of the stimulus head and arm movements, the owner concludes the stimulus interest on that product which the parents refuse to buy. This scenario is implemented in this paper using live video analytics which identifies the abnormality. Action recognition is a technique that is used in the classification of actions present in the given video. The Bag of Visual Words Model is implemented for recognizing the action made by the stimulus. This model includes feature extraction, codebook generation and classification. The features from the stimulus such as arm and head are extracted using Speeded up Robust Features (SURF) algorithm. Codebook generation is done by K-means clustering and the histogram of discriminative features is generated and fed as input to SVM classifier which recognizes the action made by the stimulus (child) in order to identify the child's interest factor on a particular product.***

Keywords : Surveillance Camera System, Computer Vision, Scale Invariant Feature Transform, Support Vector machine

I. INTRODUCTION

Surveillance camera systems have been widely used in supermarkets. Some of the reasons for having Surveillance camera systems in supermarkets include theft, damage to products and misconduct of employees. Surveillance camera systems are used mainly to prevent these crimes and improve sales in supermarkets. Obstruction of these crimes helps owners to save thousands of dollars every year. This idea of increasing sales with the help of the Surveillance camera system can be done in a different manner.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
G. Priyanka*, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India.
S. Sathya Bama, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India.
V.J Aiswaryadevi, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India.
M.S Sruthi, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India.
S. Soundarya, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimbatore, India.

Let us consider a situation where the owner of the supermarket sits at the billing section and monitors the entire supermarket through the Surveillance camera system. The owner suddenly observes in one particular section of the supermarket a child looks at a product and asks his parents to buy that product. By analyzing the child's head and arm movements the owner identifies the child's interest on that product. But the parents refuse to buy that product. There are several reasons for not purchasing a product. One of the reasons could be the increase in price of the product. So the owner makes a decision to provide discount for the product that the child was interested.

Humans perceive the world around them by making use of their eyes and their brain. The similar capability to a computer is made available with the help of Computer Vision Technology. There are several tasks involved to achieve automatic visual understanding. The tasks embody strategies for getting, processing, analyzing and understanding digital pictures, and extraction of high-dimensional information from the actual world so as to provide numerical or symbolic data, e.g., within the styles of choices. Some of the recognition problems in Computer vision includes face recognition, identification, detection, content based image retrieval, object recognition etc. In this work Computer Vision technology is used for analyzing the videos of children shopping inside the supermarket.

Image recognition provides the ability for identifying objects, people and actions in images by the software system. Image recognition can be achieved by computers by the combination of camera and artificial intelligence software. In this paper activity of children inside the supermarket especially with respect to their interest on a particular product has been recognized by analyzing their head and arm movements from videos taken inside the supermarket.

The remaining sections of this paper discusses about the related works, implementation details, experimental results, conclusion and future work.

II. RELATED WORK

Since Human behaviour assessment is a broad topic, in this section we discuss on human action recognition. A system was developed by Lee et al. [10] to extract the head regions and body regions for multiple people from surveillance camera. The proposed system performs well in extracting the head regions compared to body regions. A model that jointly estimates body and head orientation in monocular surveillance video was proposed by Chen et al. [11].

The proposed method does not investigate behavioural cues such as head pose.

The head and body orientation is generally detected by using Supervised Learning methods in all the previous works [10-11]. Popa et.al.[12] proposed an approach towards assessment of customers' level of interest. Discriminatory features for basic action detection were extracted. Then non-identical statistical and spatiotemporal classification methods were analyzed, which shows relations between frames, features, and basic actions. Hu et al used the Motion History Image (MHI) along with foreground image obtained by background subtraction and also the bar graph of oriented gradients (HOG) to get discriminatory options for action recognition [14] in the shopping environment. The model proposed by Hu et al is improved by constructing a multiple-instance learning framework SMILE SVM. This model was found to be effective in recognizing customer's interest on a particular product in supermarkets. Although the system was tested on a real world scenario the level of accuracy in recognition of customers' interest was very low. All the methods proposed above detect only the head and body orientation and to some extent pose estimation. The model proposed in this paper mainly aims to identify the child's interest on a particular product by estimating its position near the shelf where the products are displayed. It also considers the head pose which is mainly used for identifying whether the child is looking at the product or not.

III. BAG OF VISUAL WORDS

The Bag of Visual Words model is found to be one of the most reputed concepts in computer vision. It is used to classify the contents of an image and to build highly scalable Content based Image Retrieval Systems. The Fig 1 shows how the proposed system is built. It involves several steps which includes data collection, dividing data into training and test sets, feature extraction, codebook generation, and classifier learning and action classification.

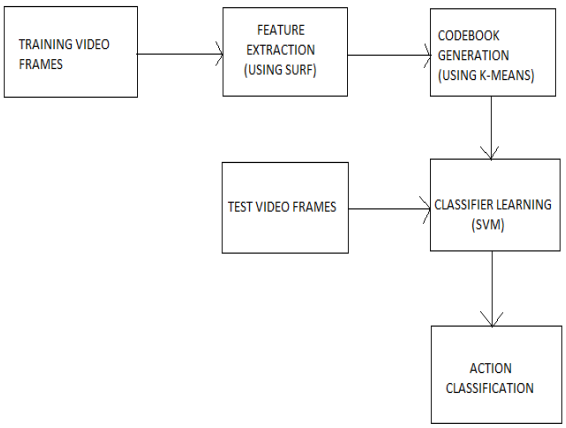

Fig 1. Flow Diagram

A. Data Collection

The dataset consists of videos of children shopping inside the supermarkets with their parents or by themselves. These videos are taken from video portals such as Youtube. The original natural setting recordings of the videos is maintained as no post processing is done. However information such as postures, illumination, clutter backgrounds and multiple objects can be seen indicating real world scenarios. The Fig 2 shows some snapshots from the videos.

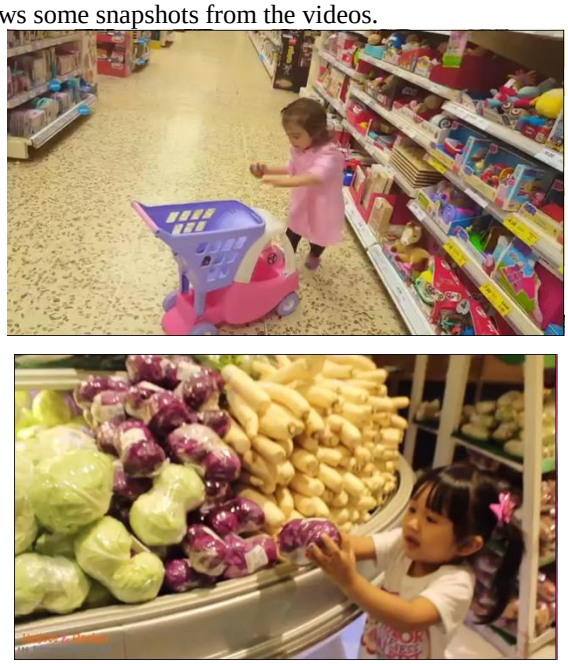

Fig 2. Videos of children shopping inside Supermarket

B. Feature Extraction

Extraction of essential features is the first step in building the Bag of Visual Words Model. Feature extraction is found to be a dimensionality reduction technique that represents attention-grabbing components of an image as a compressed feature vector. This technique is useful when a reduced feature representation is required to quickly complete tasks such as image matching and retrieval. Considering any component in the video frame, attention-grabbing points on the component are excerpted to produce a feature description of the video frame. This description which is excerpted from a training video frame can then be used to identify the object when attempting to locate the object in a test video frame containing many other objects. Reliable recognition is possible only if the characteristics excerpted from the training video frame can be identified even under changes in video frame properties. Here Speeded up Robust Features (SURF) is used for extracting features from the video frames. The SURF is an algorithm to detect and describe local features in video frames. The SURF feature descriptor is invariant to uniform scaling, orientation, illumination changes.

The collected videos are converted into frames and the frames depicting the child's interest on a product as well as doing some other activity are only considered. The other unnecessary scenes inside the supermarket are not considered for building the model.

Fig 3 shows the video frames of children interested in a particular product and Fig 4 shows the video frames of children performing some other activity. These video frames are given as input to the Bag of visual words model.

Fig 3. Input Video Frames depicting the interest of child on a product

Fig 4. Input Video Frames depicting other activity of child

C. SURF Algorithm

Speeded up robust features (SURF) is a local feature detector and descriptor mainly used for feature extraction. Object recognition, image recognition, classification are some of the operations that can be performed using this algorithm. SURF algorithm has powerful features like scale invariance, uniform translation, uniform lighting, contrast invariance, and uniform rotation. It can detect objects in images taken under different environmental settings. Integer approximation is used for finding the determinant of Hessian blob detector, for detecting interest points in SURF algorithm. This is calculated by three integer operations employing an antecedently calculated integral image. The feature descriptor is obtained by calculating the summation of the Haar wavelet response surrounding the point of interest. This feature descriptor can be calculated using integral image. SURF descriptors can help track objects, faces of people and extract points of interest. An image having identical size but with decreased bandwidth is obtained by copying the initial image with Pyramidal Gaussian. A special obscured effect on the initial image is accomplished and it is known as Scale space. It confirms that the interest points do not vary in their scale. Fig 5 provides a clear overview of the feature extraction process done by the SURF algorithm.

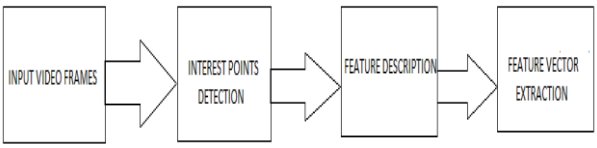

Fig 5. Feature Extraction process

Detection

Square-shaped filters are used for estimation of Gaussian smoothing by SURF algorithm. The integral image is used for filtering the image with a square:

$$S(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j)$$

The integral image and the Rectangle's four corners are evaluated to find the summation of the initial image within a rectangle. To identify points of interest SURF utilizes a blob detector built from Hessian matrix. The local change around the point is measured using the determinant of the Hessian matrix. The points are chosen where this determinant is maximal. SURF utilizes the determinant of the Hessian for determining the scale. When a point p=(x,y) is given for an image I, the Hessian matrix H(p, σ) at point p and scale σ can be determined as:

$$H(p,\sigma) = \begin{pmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yx}(p,\sigma) & L_{yy}(p,\sigma) \end{pmatrix}$$

Here $L_{xx}(p,\sigma)$ is the convolution of the second-order derivative of Gaussian. The image I(x,y) is at the point x.

Representation of Scale-space and Interest Point Location

Interest points are identified at various scales because comparison images are often required for searching correspondences where they are seen at various scales. The feature detection algorithm realizes the scale space as an image pyramid. The Gaussian filter frequently smoothen the images and sub samples them to obtain the next higher level of the pyramid. The scale space can be partitioned into series of octaves. An octave is considered to be a sequence of response maps wrapping a doubling of scale. The output of the 9x9 filters is considered to be the lowest level of the scale space in SURF algorithm. Box filters of different sizes are applied for implementing scale spaces in SURF. Instead of iteratively reducing the image size the scale space is analyzed by up-scaling the filter space. The output obtained from the 9x9 filter is the initial scale layer at scale s=1.2. The maxima of the determinant of the Hessian matrix are interpolated in scale and image space. Scale space interpolation is important because the contrast in scale between the intial layers of every octave is comparatively large.

Descriptor

A descriptor provides a distinctive and strong description about an image feature. An example for this is tracing the intensity dispersal of the pixels within the locality of the point of interest. Since the descriptors are calculated in a confined manner a description is already acquired for every point of interest that has been recognized formerly. Computational complexity and point-matching robustness/accuracy are directly affected by dimensionality of the descriptor.

The first step is as follows: The information is extracted from the circular region surrounding the interest point for fixing reproducible orientation. Then a square region situated to the chosen orientation is constructed and SURF descriptor is extracted from it.

Orientation assignment

The orientation of the point of interest has to be identified in order to attain rotational invariance. Within a circular locality of radius 6s surrounding the point of interest the Haar wavelet responses are computed in both x and y directions. Here s is denoted as the scale where the point of interest has been identified. The responses acquired are adjusted by a Gaussian function centralized at the point of interest. The results acquired are plotted in a 2D space, where abscissa in the horizontal response and ordinate in the vertical response. By summing up all responses present inside a sliding orientation window of size $\pi/3$ the dominant orientation is determined. Then summation of both the horizontal as well as the vertical responses identified inside the window is performed. The summation of responses produces a particular orientation vector. The longest vector determines the orientation of the point of interest. Care must be taken while choosing the sliding window size because it may affect the expected balance between angular resolution and robustness.

Descriptor depending on the sum of Haar wavelet responses

The region surrounding the point is described by extracting a square region. Then the square region is centralized on the point of interest and positioned along the orientation as chosen above. Smaller 4x4 square sub regions are made by splitting the interest region. The Haar wavelet responses are computed for each sub region. The computation is done at exactly 5x5 frequently separated test points. The responses are adjusted by a Gaussian for offering stableness for malformations and translation. Fig 6 depicts the feature extraction process of SURF algorithm.

```
Creating Bag-Of-Features from 2 image sets.
-------------------------------------------
* Image set 1: looking at product.
* Image set 2: Not looking product.

* Extracting SURF features using the Grid selection method.
** The GridStep is [8 8] and the BlockWidth is [32 64 96 128].

* Extracting features from 14 images in image set 1...done. Extracted 201600 features.
* Extracting features from 14 images in image set 2...done. Extracted 199440 features.

* Keeping 80 percent of the strongest features from each image set.

* Balancing the number of features across all image sets to improve clustering.
** Image set 2 has the least number of strongest features: 159552.
** Using the strongest 159552 features from each of the other image sets.
```

Fig 6: Feature Extraction using SURF

D. VOCABULARY CONSTRUCTION

After the extraction of feature vectors from each video frame it is necessary to construct vocabulary of possible visual features. Vocabulary construction is accomplished by means of K-means clustering algorithm where the feature vectors obtained from feature extraction step are clustered. The resulting cluster centers are treated as dictionary of visual words. This step is also called as codebook generation. After completion of extraction of feature vectors from the training images the codebook is ready to be built. The K-means clustering algorithm has the tendency to assemble all the features together and extricates the characteristic centroids. The centroids of the resulting clusters are interpreted to be the codewords. Then clustering process assigns each patch present in the image to a specific codeword. The results of K-Means clustering are shown in Fig 7. The Fig 8 shows how a cumulative vocabulary represents. It comprises of the overall number of individual variety of feature available in the training set.

```
* Using K-Means clustering to create a 500 word visual vocabulary.
* Number of features         : 319104
* Number of clusters (K)     : 500

* Initializing cluster centers...100.00%.
* Clustering...completed 24/100 iterations (~6.67 seconds/iteration)...converged in 24 iterations.

* Finished creating Bag-Of-Features
```

Fig 7. Vocabulary Construction using K-Means Algorithm

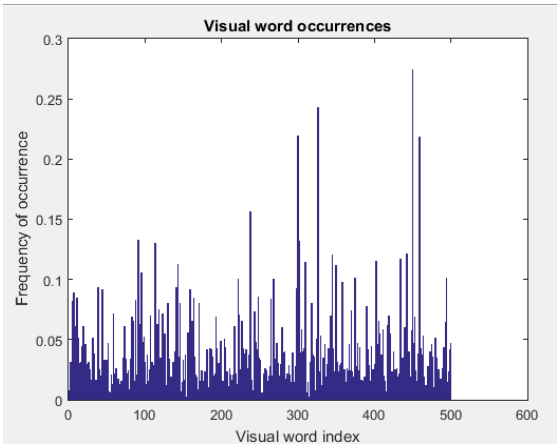

Fig 8. Histogram representing vocabulary of visual words

K-MEANS ALGORITHM

K means is considered to be an unsupervised algorithm. Clustering problem can be solved easily with the help of K means. The K means algorithm divides x observed samples into m clusters where individual sample goes to the cluster having the nearest mean. Non-identical shapes of clusters can be obtained by expectation-maximization mechanism. The K means is an approach of vector quantization, initially from signal processing. The primary method is to train a k-means clustering algorithm using the training data. Labelling of the training data is not necessary. Encoding functions are used for forecasting any input datum into the new feature space. Some examples include the threshold matrix-product of the datum with the centroid locations, the distance from the datum to each centroid, a standard function for the nearest centroid, or some straightforward transformation of the distance.

The K-means algorithm works as follows. If there is a set of N objects which has to be partitioned into m clusters then the input has to be a set of features $N=\{n_1, n_2, \ldots, n_k\}$. The main objective is to reduce the distance between every point in the dispersed cloud and the allocated centroids.

$$\text{Arg}_x \min \sum_{i=1}^{m} \sum_{n \in xi} |n - \mu i|^2$$

Where μ denotes the average of points for every X_i (cluster) and X is the set of points divided into clusters of $\{X_1, X_2, \ldots, X_i\}$.

1. The first step in K –Means is to define an initial random solution which is called as the cluster centroid. These cluster centroids have to be placed arbitrarily confined within the data.
2. Next is the **Assignment Step.** In this step K-Means repeats over every input feature and determines the nearest centroid of the cluster with respect to itself is. As soon as the nearest centroid has been determined, then it is allocated to that specific centroid.
3. Third step is **Mean and Updating**. The centroids of each cluster are relocated after the first clustering is done. The recenly calculated centroids of each cluster is the total of entire members of that specific cluster. Thus the centroid advances within a compactly situated distribution and otherwise outside the distribution.
4. As the averaging step is achieved, and the latest clusters are computed, the similar process is iterated.

E. TRAINING AND TESTING

In most areas of information science, discovering predictive relationships from data is a significant task. A training set is used for initial discovery of relationships. A test and validation set will evaluate whether the detected relationships leverage. The training set can be defined as a collection of data mainly used in the process of finding possibly predictive relationships. The test set can be defined as a collection of data used for evaluating the robustness and benefits of a predictive relationship. Training and testing are considered to be the two popular concepts in machine learning. Training is a process in which correlation between data and attributes is learned from a fragment of the training dataset. Testing is a process of testing predictions of this correlation on some other part of the dataset. In this proposed system the collected videos are converted into frames and these video frames are partitioned into training set and test set. The training and test video frames are shown in Fig 9.

```
Training an image category classifier for 2 categories.
-------------------------------------------------------
* Category 1: looking at product
* Category 2: Not looking product

* Encoding features for category 1...done.
* Encoding features for category 2...done.

* Finished training the category classifier. Use evaluate to test the classifier on a test set.


Evaluating image category classifier for 2 categories.
------------------------------------------------------

* Category 1: looking at product
* Category 2: Not looking product

* Evaluating 14 images from category 1...done.
* Evaluating 14 images from category 2...done.

* Finished evaluating all the test sets.
```

Fig 9. Training and Testing

F. IMAGE CLASSIFICATION

There are several learning methods to leverage the Bag of Visual Words for image related tasks. The Support Vector Machine algorithm is implemented for classifying whether the child is looking at the product or not.

SVM ALGORITHM

Support vector machine is a supervised algorithm. It analyzes classification as well as regression problem. Consider a set of training samples where each sample is assigned to any one of two classes. The SVM algorithm allocates a new sample to either one of the two classes by constructing a model. This makes SVM a non-likelihood binary linear classifier. The samples are represented as points in space in SVM. They are mapped in such a manner that samples of distinct categories are partitioned by a gap that is as broad as possible. The kernel trick is a technique by which the SVMs can accurately accomplish a non-linear classification besides performing linear classification. It is done by absolutely mapping their inputs into high-dimensional feature spaces.

Classifying data is considered to be one of the most popular tasks in machine learning. Consider a set of data points where every data point belongs to either of two classes. The aim here is to determine the class where a new data point should be assigned. In SVM, data point is usually perceived as a p-dimensional vector. It is necessary to identify in case such points can be separated by a (p-1) dimensional hyper plane. . This approach is known as a linear classifier. There are various hyper planes which best categorize the data. The best hyper plane depicts the largest partition among the two classes. So the hyper plane must be selected in a manner where the interval from it to the closest data point on each side is maximized. This hyper plane is said to be the maximum-margin hyper plane. The linear classifier this hyper plane defines is called the maximum margin classifier.

IV. EXPERIMENTS AND RESULTS

The Bag of Visual Words Model explained above is evaluated on the collected dataset. The dataset consists of videos of children shopping inside the supermarket. The collected videos are converted into frames. The input file consists of frames depicting the child's interest on product and doing some other activities like walking or running. It is then divided into training and testing tests. The training set comprises of video frames which are given as input to the Bag of Visual Words Model. Next the features are extracted from each video frame by the SURF algorithm. The extracted features are used for generating the codebook by k-means clustering. This codebook which consists of a vocabulary of visual words is fed into SVM classifier. The SVM algorithm is trained with the codebook that is generated. The test set is evaluated by the SVM algorithm which finally classifies the test video frames as whether the child is looking and pointing his finger to a product or doing some other activity. The overall accuracy of the algorithm was found to be 82% which is shown in Fig 10.

	PREDICTED	
KNOWN	looking at product	Not looking product
looking at product	0.64	0.36
Not looking product	0.00	1.00

* Average Accuracy is 0.82.

Fig 10. Confusion Matrix and Classification Accuracy

V. CONCLUSION AND FUTURE WORK

The bag of visual words model is considered to be a significant concept in computer vision. The bag of visual words model mainly classifies the contents of an image. It's used to build highly scalable Content Based Image Retrieval (CBIR) systems. In this work Bag of Visual Words Model is implemented. This model first creates a bag of features from the video frames in the training set. Then the extracted features are provided as input to the SVM classifier for identifying whether the child is looking at a product and pointing to it or he is doing some other activities. The method has achieved an overall accuracy of 82%. As future work some other model can be used to identify the interest of children on particular product by recognizing their head and arm movements. The method can be applied to more real time videos taken inside the supermarket. .

REFERENCES

1. Manoj Ramanathan,Wei Yun Yau and Eam Khwang Teoh, "Human Action Recognition With Video Data: Research and Evaluation Challenges", *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 5, pp.650-663, 2014.
2. Hossein Rahmani,Ajmal Mian and Mubarak Shah, "Learning a Deep Model for Human Action Recognition from Novel Viewpoints", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp.667-681, 2018.
3. A. F. Bobick and J. W. Davis: "The recognition of human movement using temporal templates",*IEEE Trans, PAMI*, 2001.
4. I.Laptev:" Improving Object Detection with Boosted Histograms", *In Image and Vision Computing*, vol. 27, issue 5, pp. 535-544, 2009.
5. Haritaoglu, I., & Flickner, M. (2001). "Detection and tracking of shopping groups in stores". *In Computer Vision and Pattern Recognition*, 2001. CVPR 2001. Proceedings of the 2001 *IEEE Computer Society Conference on* (Vol. 1, pp. I-431). IEEE.
6. Jun Liu, Gang Wang, Ling-Yu Duan, Kamila Abdiyeva and Alex C. Kot, "Skeleton-Based Human Action Recognition With Global Context-Aware Attention LSTM Networks", *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp.1586-1599, 2018.
7. I. Haritaoglu and M. Flicker, "Attentive billboards: Towards to video based customer behaviour," *In Proc.IEEE Workshop on Applications of Computer Vision*, pp.127-131, Orlando, FL, USA, Dec. 2002.
8. W. Hu,T. Tan, L. Wang, and S, Maybank, "A survey on visual surveillance of object motion and behaviours," *System, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transaction* on.34(3)334-32,Aug 2004.
9. Zhimeng Zhang,Xin Ma, Rui Song, Xuewen Rong, Xincheng Tian, Guohui Tian and Yibin Li, "Deep Learning based human action recognition: A survey", 2017 *Chinese Automation Congress*, pp.3780-3785, 2017.
10. Lee, Kang-Dae, et al. "Context and profile based cascade classifier for efficient people detection and safety care system." *Multimedia Tools and Applications63.1* (2013): 27 44.
11. Chen, C., Heili, A., & Odobez, J. M. (2011, August). "Combined estimation of location and body pose in surveillance video". In *Advanced Video and Signal-Based Surveillance (AVSS)*, 2011 8th *IEEE International Conference* on (pp. 5-10). IEEE.
12. Popa, M., Rothkrantz, L., Yang, Z., Wiggers, P., & Shan, C. (2012, October) "Assessment of Customers' Level of Interest". In Image Processing, 2012 19th *IEEE International Conference* on (pp.41-44) .IEEE.
13. Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang: "Action detection in complex scenes with spatial and temporal ambiguities "*Proceedings of the International Conference on Computer Vision* (ICCV ' 09), 2009.
14. N. Dalal, and B. Triggs: "Histograms of oriented gradients for human Detection", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05), vol. 1, pp. 886-893, San Diego, USA, 2005.
15. M.C. Popa, L.J.M. Rothkrantz, D. Dateu, P. Wiggers, R. Braspenning, and C. Shan: "A comparative study of HMMs and DBNs applied to Facial Action Units Recognition", *Neural Network World*, vol. 6, pp.737-760,2010.
16. M.C. Popa, L.J.M. Rothkrantz,Z. Yang, P. Wiggers,R. Braspenning, and C. Shan: "Analysis of Shopping Behaviour based on Surveillance System",2010 *IEEE Int. Conf. on Systems, Man, and Cybernetics* (SMC'2010),Istanbul, Turkey,2010.
17. Comaniciu D, Ramesh V, Andmeer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25:564–575
18. S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. *NIPS*, 2002.
19. J. Davis and A. Bobick. The representation and recognition of human movement using temporal templates. *CVPR*, 1997.
20. J. Liu, S. Ali, and M. Shah. Recognizing human actions using multiple features. *CVPR*, 2008.
21. C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. *ICPR*, 2004.
22. J. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *BMVC*, 2006.
23. Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. *CVPR, Washington, DC, USA*, pages 66–75, 2004.
24. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
25. K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.
26. P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *The 9th ICCV, Nice, France*, volume 1, pages 734–741, 2003.

AUTHORS PROFILE

G. Priyanka, working as an assistant professor at Sri Krishna College of Technology. She received her B.E in Computer Science and Engineering from Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India in the year 2016 and received the M.E from Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India in the year 2018. Her area of interest is Data Analytics.

S. Sathya Bama, working as an assistant professor at Sri Krishna College of Technology. She received her B Tech in Computer Science and Engineering from Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India in the year 2016 and received the M.E from Anna University Regional Campus, Coimbatore, Tamil Nadu, India in the year 2018. Her area of interest is Data Analytics.

V.J Aiswaryadevi, joined the department of Information Technology in Bannari Amman Institute of Technology in the year 2013 as Assistant Professor until 2015 undergoing research studies on the Specification mining techniques for classification of false positives in the Eclipse packages. From 2017 to 2018 she worked at Sri Guru Institute of technology in the department of Computer Science and Engineering. From 2018 to till date, she works at Sri Krishna College of Technology. Ms. V J Aiswaryadevi obtained her M.E post-graduation in Software Engineering from Sri Ramakrishna Engineering College affiliated by Anna University in the year 2013 and started her doctorate program under Anna University during the year 2017-2018. She is currently working on the deep learning networks and fusion of AI and IOT.

M.S Sruthi, working as an assistant professor at Sri Krishna College of Technology. She received her B.E in Computer Science and Engineering from CMS College of Engineering and Technology, Coimbatore, Tamil Nadu, India in the year 2015 and received the M.E from Anna University Regional Campus, Coimbatore, Tamil Nadu, India in the year 2017. Her area of interest is Data Analytics.

S.Soundarya, working as an assistant professor at Sri Krishna College of Technology. She received her B.E in Computer Science and Engineering from RVS Technical Campus, Coimbatore, Tamil Nadu, India in the year 2016 and received the M Tech from PSG College of Technology, Coimbatore, Tamil Nadu, India in the year 2018. Her area of interest is Data Analytics.