Metadata Based Object Detection and Classification Using Key Frame Extraction Method

S. Vasavi and V. Srinivasa Rao
Department of Computer Science and Engineering, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India
Email: vasavi_movva@hotmail.com, hodcse@vrsiddhartha.ac.in

***Abstract*—Visual surveillance is an active area of research topic. Data that is collected from these cameras have to be monitored by human operators manually for long durations which is not feasible in real time and may lead to inaccurate results. Recorded videos are analyzed only when any unwanted event occurs that may help for recovery and not avoidance. Intelligent video surveillance requires algorithms that are fast, robust and reliable during various phases such as object & shadow detection, classification, tracking, and event analysis. This paper presents metadata based object & shadow detection, classification system for video analytics. Meta data of key frames is stored in the form of database, that helps for object & shadow detection. Gaussian white noise is used for background modeling. Convex non overlapped blobs are identified using LoG (Laplacian of Gaussians). Four channels in four color spaces are used for better removal of shadows. Shadow boundary is detected by growing a user specified shadow outline on an illumination-sensitive image among the four channels. For Object detection we used canny edge detection and Harris corner point detection for detecting the objects, Bag of visual words with SIFT descriptors for extraction of features and KNN classifier for classification. The proposed method is tested on sample CCTV camera video and results are analyzed using performance measures.**

***Index Terms*—key frames, meta data, shadow removal, color space, bag-of-visual-words, KNN classifier**

I. Introduction

The basic video analysis operations such as object& shadow detection, classification, tracking require scanning the entire video which is a time taking process. Instead of searching the entire video, we can use Meta data file for basic video analysis techniques with less time complexity. Table I shows Meta data file that stores key frame numbers and corresponding locations of the object within that frame. For example first row indicates that in frame number 1; object was located with coordinates (x, y) (a, b) (p, q) and (r, s) and so on.

Shadows are unwanted regions causing problems in object detection and classification. In the proposed system, shadow identification and removal is done for every key frame stored in the metadata and the output frame is given to detection and classification phases. The paper is organized as follows: Section 2 presents literature survey on various existing methods for all the three phases. Methodology of proposed system is described in Section 3. Section 4 includes results obtained from our proposed system and analysis w.r.t performance measures. Conclusions and future work are given in Section 5.

TABLE I. Meta Data Format

Frame No	Coordinates
1	(x, y) (a, b) (p, q) (r, s)
2	(c, d) (m, n) (t, u) (i, j)
...	....

II. Literature Survey

A. *Survey Survey on Key Frame Extraction and Background Modeling*

Key frame algorithms basically can be categorized as given in [1]. (i) Shot boundary based methods extract frames from a fixed position in a lens as key frames, which mainly chooses the first frame, middle frame and the last frame. (ii) Visual content based approaches detects key frames using low-level features such as edge, color and texture changes (iii) Clustering based approaches calculate the distance from each frame to several existing clustering center and if the distance is less than the threshold value, then classified in to the smallest distance cluster, or creates a new cluster. (iv) Motion analysis based approach uses movement information of objects in the video or camera to conclude upon key frames. (v) Compressed video stream extraction based approaches uses features of the compressed video data. Ref. [2] proposed shot boundary based key frame extraction method for segmenting user generated videos on video sharing websites using visual content. Ref. [3] proposed image epitome based framework for extracting key frame from consumer video that uses image epitome to measure dissimilarity between frames of the input video and does not require shot detection, segmentation. Ref. [4] proposed key frame extraction based on intra-frame and inter-frame motion histogram analysis where

Manuscript received March 8, 2015; revised September 1, 2015.

doi: 10.18178/joig.3.2.90-95

entropy values of the consecutive frames are used to conclude on motion objects. Ref. [1] uses heterogeneity image patch (HIP) index of a frame in a video sequence to judge on key frame. Ref. [5] proposed a hybrid approach that uses both shot boundary and visual content based methods for key frame extraction. We proposed statistical measure approach for key frame extraction.

Background modeling approaches can be classified as intensity based background subtraction methods, edge based methods, texture based methods, a combination of color and edge based methods, combination of color and gradient based methods and Hierarchical approaches. Ref. [6] proposed a method for modeling the background that uses per pixel, time adaptive, Gaussian Mixtures in the combined input space of depth and luminance –invariant color. In this method Foreground Objects are treated as Novel regions and background as Non-Novel Regions. Ref. [7] proposed a background subtraction technique based on modified adaptive GMM for detecting moving objects. A multiple feature integration algorithm for background modeling and subtraction is proposed in [8], where the background is modeled with a generative method and background and foreground are classified by a discriminative technique. Our proposed method uses Gaussian white noise for background modeling. Convex non overlapped blobs are identified by using LoG (Laplacian of Gaussians).

B. *Survey on Shadow Detection and Removal*

Shadow detection methods can be classified as single region classification and pair-wise region relationship classification. Ref. [9] proposed pair-wise relationship method for shadow detection. Ref. [10] uses morphological operations for identifying and removing the shadow. Ref. [11] proposed a method for detecting shadows while preserving texture appearance. Ref. [12] developed tools for shadow modification in images where a shadowed region is characterized by soft boundaries with varying sharpness along the shadow edges. Ref. [13] proposed an automatic method for producing a high-quality shadow-free image from a single input image. Ref. [14] presents an approach to shadow removal that preserves texture consistency between the original shadow and lit area. Existing automatic approaches uses lab color space for shadow detection and removal resulting to inefficient removal. In the proposed user aided approach, we used four color spaces for better removal of shadows. The proposed algorithm addresses limitations in uneven shadow boundary processing and umbra recovery. This approach initially takes four channels from four color spaces, V channels from two color spaces HSV and LUV and Y, C_b channels from the Y C_b C_r.

C. *Survey on Object Detection and Classification*

There are two main approaches for object detection and classification: the single model approach and the parts-based model approach [15]. In the single model approach, an object is assumed to be rigid and the descriptor (edge based) will cover the whole object where as in parts-based approach, each part is detected separately and we say that an object is found, if several of its parts are found. Ref. [16] proposed parts-based model for object detection. Ref. [17] uses a scanning-window grid approach for object detection. Ref. [18] proposed a method where Objects are recognized and classified based on pattern recognizers. Ref. [19] proposed a method to extract local features based on interest point which is used to detect key-points within an image. This method used Speed-up Robust Feature (SURF) method as interest point detector. Ref. [20] explored the different possible parameters of the Bag of Words (BoW) approach, namely inverted file and min-hash for recognition. Our proposed method for Object detection and classification uses canny edge detection and Harris corner point detection for detecting the objects, Bag of words with SIFT descriptors for extraction of features and KNN classifier for classification.

III. Proposed System

The overall flow diagram of our proposed system is shown in Fig. 1. We considered recorded video from a single stationary camera.

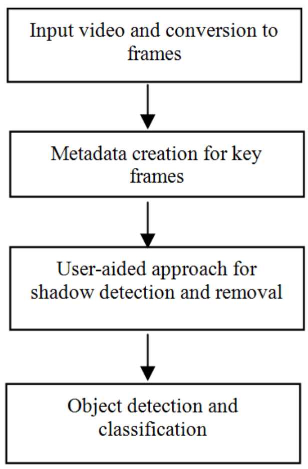


Figure 1. Overall flow diagram for the proposed method.

Algorithm: Meta data creation

Input: video

Output: Meta data

Step 1: Take input video and convert it into frames. 25-30 fps (frames per sec) is preferable.

Step 2: Key frame Extraction is done by calculating mean and standard deviation as given in (1.1) and (1.2).

$$\mu = \frac{1}{n}\sum_{i=1}^{N} x \tag{1.1}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2} \tag{1.2}$$

Frames having the dissimilarity above the threshold as given in (1.3) indicate key frames having interest objects.

$$\text{Threshold value} = \mu + \sigma * 2 \tag{1.3}$$

Initialize an array with null value and assign the key frame number, whenever the frame passes the threshold check; as shown below:

if (dissimilarity between the frames>Threshold)
temp= [temp; k];
end.

temp array gives us the key frame numbers that under gone for blob detection.

Step 3: Background subtraction was done by using Gaussian white noise as given in (1.4) in which values at any time are statistically independent and identically independent (in the simple words they are correlated).

$$P(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)}{2\sigma^2}} \tag{1.4}$$

Step 4: Locate the objects by using nonzero collections method as given in (1.5) and blob detection method as given in (1.6) to identify the object location with its coordinates.

$$a[i][j] = \sum_{i=1}^{n}\sum_{j=1}^{n} if\left(I(x, y)\right) \tag{1.5}$$

Blob detection is based on the LoG. For a given image f(x, y) is convolved by using (1.6).

$$g(x, y, t) = \frac{1}{2\pi t^2} e^{-\frac{x^2+y^2}{2t^2}} \tag{1.6}$$

For time t; scalar presentation is given as (1.7):

$$L(x, y; t) = g(x, y, t) * f(x, y) \tag{1.7}$$

Apply Laplacian operator as given in (1.8).

$$\nabla^2 L = L_{xx} + L_{yy} \tag{1.8}$$

Blob detector with automatic scale selection can be normalized as given in (2).

$$\nabla^2_{norm} L(x, y, ; t) = t\left(L_{xx} + L_{yy}\right) \tag{2}$$

Step 5: Place the key frame numbers and object located coordinates in a separate file which we called as Meta data.

for i=1 to n
m[i][1]=key frame number
m[i][2]= [x1 y1]
m[i][3]= [x2 y2]
m[i][4]= [x3 y3]
m[i][5]= [x4 y4]

Algorithm: Shadow detection and removal

1. Read key frames for shadow detection and removal
2. Initial shadow boundary detection.

User will specify a rough sketch on the shadow area. This rough sketch will extend up to entire shadow region using Sparse Field Active contours technique. Active contour methods for image segmentation allow a contour to deform iteratively to partition an image into regions. Equation (3) is used for indentifying the initial shadow point. Equation (4) is used to extend the initial shadow boundary point to four color channels of the key frame.

$$\varphi(x) = x^{-\lambda} \, (\lambda > 0) \tag{3}$$

where x= pixel intensity.

λ = Constant value (default value 5. If we increase λ value, function reaches the closer value of 0).

$$F = \left(\sum_{l=1}^{4} c_l \varphi(\sigma_l)\right) \Big/ \left(\sum_{l=1}^{4} c_l \varphi(\sigma_l)\right) \tag{4}$$

where c_l = color channel, σ_l = Standard deviation.

3. Remove noise artefacts

To remove noise artefacts we apply bilateral filtering to the fused image as given in (5).

$$I^{filtered}(x) = \frac{1}{W_p}\sum I(x_i)\, f_r\left(\| I(x_i) - I(x) \|\right) g_s\left(\| x_i - x \|\right) \tag{5}$$

The normalization term using (6) ensures that the filter preserves image energy.

$$W_p = \sum_{x_i \in \Omega} f_r\left(\| I(x_i) - I(x) \|\right) g_s\left(\| x_i - x \|\right) \tag{6}$$

where:

$I^{filtered}$ is the filtered image;
I is the original input image to be filtered;
x is the coordinates of the current pixel to be filtered;
Ω is the window centered in x;
f_r is the range kernel for smoothing differences in intensities.
g_s is the spatial kernel for smoothing differences in coordinates.
W_p is the weight

4. Identify uneven shadow boundaries

Interval variable sampling is used to identify the uneven shadow boundaries as given in (7).

$$\begin{cases} \bar{Q} = \lfloor Q_m / \xi \rfloor (m \le N, m \in N) \\ D_n = \bar{Q}_{n+1} - \bar{Q}_n (n \le N - 1, n \in N) \end{cases} \tag{7}$$

where:

N is the number of boundary points
m and n are indexes of boundary points
$\bar{Q}$ is the quantized sum
Q is the normalized sum

5. Illumination variation estimation

Illumination variation estimation can be performed to identify the change of pixel intensities along the shadow area using (8).

$$S_i(x) = \begin{cases} k & x < x_1 \\ f(x) & x_1 < x \le x_2 \\ 0 & x > x_2 \end{cases} \tag{8}$$

where x is pixel location
x_1 and x_2 are starting and end of the penumbra (shadow area)
K is a negative scale constant

6. Reconstruct the lost portions

In painting technique is used to reconstruct the lost or damaged portions of the image by replacing the pixels with the surrounding pixel intensities. Image in painting is the process of reconstructing distorted are lost portions of the digital image are video using (9).

$$I_{i,j}^{n+1} = I_{i,j}^{n} + \Delta t I_{t,i,j}^{n} \forall (i, j) \in \Omega \tag{9}$$

where $I_{i,j}^{n+1}$ = n+1th iteration on image pixels

$I_{i,j}^{n}$ = n th iteration on image pixels

$\Delta t\, I_{i,j}^{n}$ =increment to be added to the nth iteration

Ω=image pixels group

7. Comparison of Original Video with the Reproduced video

The original video is compared with the reproduced video to visualize the results and performance of the proposed method**.** As we applied bilateral filtering for the input fused images the noise present in the key frames can be reduced to greater extent. This can be identified by the PSNR values of the input key frames with the PSNR values of the shadow removed frames using (10) and (11).

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{10}$$

where:

MAX_i=maximum possible pixel value of the image.

MSE=Mean Squared Error can be calculated using (11).

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I(i,j) - K(i,j)\right]^2 \tag{11}$$

where:

I(i,j)=original image

K(I,j)=noisy image

Algorithm: Object detection and classification

Step 1: C**anny edge detection**

Smoothing:

Image is first smoothed by applying Gaussian filter. The kernel of a Gaussian filter with a standard deviation of σ = 1.4 using (12).

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{12}$$

where *N* is the total number of dataset values.

μ is the mean value of given data.

$X_i = x_1, x_2, x_3$…………

Finding gradients: The edges should be marked where the gradients of the image has large magnitudes.

|G| = G(x) + G(y)

|G| = |Gx| + |Gy|

where *Gx* and *Gy* are the gradients in the x- and y-directions respectively.

Non-maximum suppression:

Max |G| = |Gx| + |Gy|

Only local maxima should be marked as edges.

Harris corner point detection

Here, we calculate the corner points of the objects using (13):

C = (Ix.*Iy - Ixy.^2)./(Ix + Iy + eps) (13)

where eps=5

C= corner points of the binary image

x= row coordinates

y= column coordinates

Step 2: **BOW: (bag of visual words)**

Bag of visual words is constructed by taking sample images of human faces, airplane, car, motor bike etc each having up to 50 different models. KNN classifier is trained using these 50 models. For testing we used 100 samples of each of the human faces, aero plane, car and motor bike.

Three stages:

1. Represent each training image by a vector. For representation of features, SIFT descriptor is used. Features such as nose, ears are considered for SIFT descriptor as given in (14).

$$D(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \cdot I(x, y) \tag{14}$$

2. Train a classifier to discriminate vectors corresponding to positive and negative training images.
3. KNN algorithm

Here we measure distance between centroid and every other pixel in the blob, such that we can find the smallest distances between the pixels that can give better classification. Apply the trained classifier to the test image as given in (15).

a2 = (ones(ncentres, 1)*sum((x.^2)', 1))'+.. ones(ndata, 1) * sum((y.^2)',1)...2.*(x*(y')) (15)

IV. Performance Analysis

Input video from the website [21] is used as dataset. Performance is calculated with the following two methods for creating Meta data:

1. Collecting non zero pixels
2. Blob detection

In method 1, procedure stores the values of all pixel values with its gray values, looking for non zero values next to it. Calculate the height, width, minimum and maximum location of pixels by indexing for loop on image gray value matrix to retrieve the object location. As this process runs 255*255 times every time, even if there are no objects (no white colored images), becomes main disadvantage and slower. Considering time complexity, method 2 (blob detection) is used to find the object location by its centroid, there by storing the coordinate’s w.r.t its frame numbers in the sequence of frames as shown in snapshot Table II.

PSNR metric is used to compare the performance of user aided and manual approaches. Peak Signal to Noise Ratio can be computed using (10) and (11) by first calculating MSE and then dividing it by the maximum range of the data type. PSNR is most commonly used to measure the quality of reconstruction of images.

SIFT can find more number of matching points than SURF, as such performance of identifying objects can be increased. In existing systems KNN algorithm is used for classification where as the proposed method takes centroid values such that we can find the smallest distances between the pixels improving the performance of classification.

On an average, method 1 takes minimum of 3 seconds and maximum of 10-15 seconds and method 2 took average of 5 sec to do the job as shown in Table III.

TABLE II. META DATA IN EXCEL FILE

1 pos 1	2 pos 2	3 pos 3	4 pos 4	5 pos 5
30	[267,25]	[267,85]	[357,25]	[387,175]
75	[325,85]	[325,145]	[415,85]	[445,235]
79	[318,48]	[318,108]	[408,48]	[438,198]
172	[227,186]	[227,246]	[317,186]	[347,336]
202	[28,105]	[28,165]	[118,105]	[148,255]
217	[-2,242]	[-2,302]	[88,242]	[118,392]
298	[33,263]	[33,323]	[123,263]	[153,413]
367	[199,165]	[199,225]	[289,165]	[319,315]
533	[414,122]	[414,182]	[504,122]	[534,272]
675	[233,130]	[233,190]	[323,130]	[353,280]
697	[162,178]	[162,238]	[252,178]	[282,328]
714	[76,82]	[76,142]	[166,82]	[196,232]
749	[191,210]	[191,270]	[281,210]	[311,360]
769	[222,167]	[222,227]	[312,167]	[342,317]
890	[28,86]	[28,146]	[118,86]	[148,236]
952	[219,304]	[219,364]	[309,304]	[339,454]
957	[233,296]	[233,356]	[323,296]	[353,446]
981	[279,298]	[279,358]	[369,298]	[399,448]
982	[379,246]	[379,306]	[469,246]	[499,396]
998	[297,232]	[297,292]	[387,232]	[417,382]

TABLE III. RESULT SET COMPARISON

Frames	Time taken to identify the object	
	Using Method 1	Using Method 2
10 Frames	35 Sec	20 Sec
1000 Frames	3900 Sec	2000 Sec

PSNR values of automatic and proposed approach are shown in Table IV for shadow detection and removal.

TABLE IV. COMPARISON OF RESULTS BETWEEN AUTOMATIC AND MANUAL APPROACHES

Test case number	**PSNR value comparison**	
	[22] Automatic Approach	**Proposed User Aided Approach**
1	17.8891	19.8811
2	9.7919	13.3614
3	14.4097	16.8642
4	10.3704	13.7491
5	3.6141	4.8698
6	7.7982	15.8700
7	18.3704	19.7491
8	23.6541	24.8008
9	22.7682	23.1260

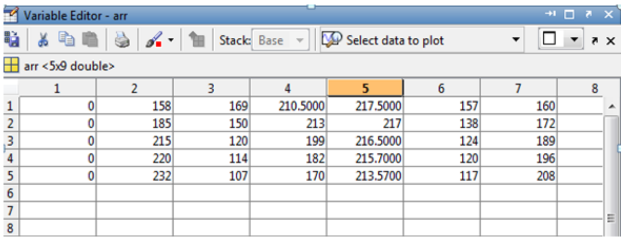
These results clearly illustrates that existing automatic approach [22] does not fully remove the shadows while our proposed User aided approach that uses four color spaces produces better shadow removal results. Results were compared for object detection with different measures SIFT, SURF as listed in Table V.

TABLE V. RESULT COMPARISON

	SURF	SIFT
Feature Matching points	20	48
Time Taken to detect	0.4 sec	1.3 sec
Performance	Work faster	Gives accurate results

On the basis of the input data, code words obtained for the bag of words can be stored. Features from the Harris can be stored as shown in snapshot Table VI.

TABLE VI. SIFT FEATURES


Variable Editor - arr

arr <5x9 double>

	1	2	3	4	5	6	7	8
1	0	158	169	210.5000	217.5000	157	160	
2	0	185	150	213	217	138	172	
3	0	215	120	199	216.5000	124	189	
4	0	220	114	182	215.7000	120	196	
5	0	232	107	170	213.5700	117	208	
6								
7								
8								

V. CONCLUSIONS AND FUTURE WORK

Our proposed approach used Gaussian white noise instead of Adaptive GMM for background modeling. Key frames are identified using threshold on the basis of mean and standard deviation. Convex blobs are identified on subtracted images to notify the object locations. Main advantage of proposed system is to reduce the time complexity of video analysis system. A user-friendly texture-preserving shadow and ghost removal method that overcomes some common limitations from the past work is also presented. Specifically, our approach retains shadowed texture and performs well on highly-uneven shadow boundaries, non-uniform ghost illumination, and non-white lighting. This user aided approach works in four color spaces namely V channels from two color spaces HSV and LUV and Y, C_b channels from the Y C_b C_r was compared with the automatic shadow removal. PSNR values of automatic and proposed approach are calculated and user aided approach produced better results. Canny edge detection is used that allows collecting evidence about possible object locations. Corner points are detected using Harris and the features of the objects can be obtained by using bag of visual words and are trained using SIFT descriptor so that the feature matching points and the performance of the identification of the objects can be improved.

As a future work, results have to be compared with other existing works such as Chinh T. Dang, Ling Shao, Pascal Kelm, Michael Harville. In this system we did not consider overlapped blobs. Hence our future work concentrates on identifying the blobs which may overlap with each other. The proposed user aided approach should be compared with other automatic approaches such as [10]. Also instead of selecting neighbor pixels, image regions which are of the same class, surface type with shadowed part, could be selected. Our proposed system works efficiently for one stationary camera and has to be tested for multiple and PTZ cameras. It can be even extended to identify all the objects present in the video instead of only moving objects from key frames.

REFERENCES

[1] C. T. Dang and H. Radha, "Heterogeneity image patch index and its application to consumer video summarization," *IEEE Trans Image Process*, vol. 23, no. 6, pp. 2704-2718, 2014.

[2] P. Kelm, S. Schmiedeke, and T. Sikora, "Feature-Based video key frame extraction for low quality video sequences," in *Proc. 10th IEEE Workshop on Image Analysis for Multimedia Interactive Services*, 2009.

[3] C. T. Dang, M. Kumar, and H. Radha, "Key frame extraction from consumer videos using epitome," in *Proc. 19th IEEE International Conference on Image Processing (ICIP)*, 2012.

[4] L. Shao and L. Ji, “Motion histogram analysis based key frame extraction for human action/activity representation,” in *Proc. IEEE Canadian Conference on Computer and Robot Vision*, 2009.
[5] M. Huang, *et al.*, “An integrated scheme for video key frame extraction,” in *Proc. 2nd International Symposium on Computer, Communication, Control and Automation*, 2013.
[6] M. Harville, G. Gaile, and J. Woodfill, “Foreground segmentation using adaptive mixture models in color and depth,” in *Proc. Detection and Recognition of Events in Video*, 2001.
[7] K. Kumar and S. Agarwal, “An efficient hierarchical approach for background subtraction and shadow removal using adaptive GMM and color discrimination,” *International Journal of Computer Applications*, vol. 75, no. 12, pp. 1-7, 2013.
[8] B. Han and L. Davis, “Density-Based multifeature background subtraction with support vector machine,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1017-1023, 2012.
[9] R. Q. Guo, Q. Y. Dai, and D. Hoiem, “Single-Image shadow detection and removal using paired regions,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
[10] S. A. Lakhotiya and M. D. Ingol, “Robust shadow detection and optimum removal of shadow in video sequences,” *International Journal of Advanced Engineering Research and Studies*, pp. 25-28, 2013.
[11] T. P. Wu and C. K. Tang, “A Bayesian approach for shadow extraction from a single image,” in *Proc. IEEE Int. Conf. Computer Vision*, 2005, pp. 480-487.
[12] A. Mohan, J. Tumblin, and P. Choudhury, “Editing soft shadows in a digital photograph,” *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 23-31, 2007.
[13] Y. Shor and D. Lischinski, “The shadow meets the mask: Pyramid-based shadow removal,” *Computer Graphics Forum*, vol. 27, no. 2, pp. 577-586, 2008.
[14] F. Liu and M. Gleicher, “Texture-Consistent shadow removal,” in *Proc. European Conf. on Computer Vision (ECCV)*, 2008, pp. 437-450.
[15] D. K. Prasad, “Survey of the problem of object detection in real images,” *International Journal of Image Processing (IJIP)*, vol. 6, no. 6, pp. 441-466, 2012.
[16] C. H. Nguyen, “Feature identification and tracking for an autonomous ground vehicle,” M.S. thesis, Virginia Polytechnic Institute and State University, 2013.
[17] M. Stigson, “Object tracking using learning detection in thermal infrared video,” *Computer Vision and Robotics*, 2013
[18] D. S. Prabha and J. S. Kumar, “Three dimensional object detection and classification methods: A study,” *IJERST*, vol. 2, no. 2, pp. 1-12, 2013.
[19] A. M. Ahsan and D. B. Mohamad, “Features extraction for object detection based on interest point,” *Telkomnika Indonesian Journal of Electrical Engineering*, vol. 11, no. 5, 2013
[20] M. Aly, M. Munich, and P. Perona, “Bag of words for large scale object recognition properties and benchmark,” in *Proc. Sixth International Conference on Computer Vision Theory and Applications*, 2013.
[21] CCTV camera files. [Online]. Available: http://www.camerasecuritynow.com/cctv-cameras/avi-files.asp
[22] A. H. Suny and N. H. Mithila, “A shadow detection and removal from a single image using LAB color space,” *International Journal of Computer Science Issues*, vol. 10, no. 4, pp. 270-273, July 2013.

Dr. S. Vasavi is working as a Professor in Computer Science & Engineering Department, VR Siddhartha Engineering College, Vijayawada, India. With 19 years of experience. She currently holds funded Research projects from University Grants Commission (UGC) and Advanced Data Research Institute (ADRIN, ISRO). Her research areas are Data mining and Image Classification. She is a life member of Computer Society of India. She published 35 papers in various International conferences and journals.

Dr. V. Srinivasa Rao is working as a Professor & HOD in Computer Science & Engineering Department VR Siddhartha Engineering College, Vijayawada, India with 25 years of experience. He currently holds funded Research from University Grants Commission (UGC) and Advanced Data Research Institute (ADRIN, ISRO). His research areas are bio-Informatics and Image Processing. He is a life member of Computer Society of India He published more than 35 papers in various conferences and journals.