IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555
Vol. 3, No.5, October 2013

Homography based Multiple Moving Objects Using Multi-Dimensional Event Feeds

Pushpa, D.
Asst. Prof: Dept of Information Science & Engg.
Maharaja Institute of Technology
Mysore, Karnataka, India
E-Mail: pushpad15@gmail.com

Dr. H.S. Sheshadri
Prof.: Dept. of Electronics & Communication Engg.
PES College of Engg.
Mandya, Karnataka, India
E-Mail:hssheshadri@gmail.com

Abstract— **Object detection and tracking is one of the active domain of research currently seen in the area of computer vision. The various challenges like partial or complete occlusion, object mobility, unprecedented scene changes, unpredictable appearance and disappearance of an object posses a greater deal of issues in designing object detection and tracking. Therefore, the proposed system exhibits a novel framework that is performed using homography based techniques where frame sequences of multiple dimensionality is considered. All the previous work has focused using 3D views whereas the current work is done considering 6 views of similar event. Finally, the accomplished results are compared with all the major studies done in past with standard parameters for comparative performance evaluation to show the proposed model exponentially increases the multiple moving object detection and tracking mechanism.**

Keyword-; Object Detection and Tracking, Homography, Multiple Objects

I. INTRODUCTION

In the last few years, object detection and localization have become very popular areas of research in computer vision. Although most of the categorization algorithms tend to use local features [1], there is much more variety on the classification methods. In some cases, the generative models show significant robustness with respect to partial occlusion and viewpoint changes and can tolerate considerable intra-class variation of object appearance. One good way to understand homographies is to put them into the context of other geometric transformations [2]. The homography transformation has 8 degrees of freedom and there are other simpler transformations that still use the 3x3 matrix but contain specific constraints to reduce the number of degrees of freedom. There are many situations in computer vision where estimating a homography may be required as illustrated below [2].

- **Camera calibration**: Camera calibration is the process of determining the intrinsic and extrinsic parameters of the camera setup. Calibration is often the primary step of many vision applications as it allows systems to determine a relationship between what appears on an image and where it is located in the world.
- **3D reconstruction**: 3D reconstruction is a problem in computer vision where the goal is to reconstruct scene structures and camera positions from images of the scene. One domain where this is extremely useful is in medical imaging, where multiple images of a body part, such as the brain, can be used to create a 3D model of the part being analyzed. Solving for homographies is a key step in 3D reconstruction as it is often required to obtain mappings between images of the scene.
- **Visual metrology**: The goal in visual metrology is to estimate distances between and sizes of objects from images of those objects. Homography estimation is crucial in this domain as it allows multiple images of a plane to be transformed to a common plane where measurements can be acquired.
- **Stereo vision**: Stereo vision is the process in visual perception of sensing depth from multiple views of a scene. These multiple views are commonly taken from different camera positions, but can also be acquired from a fixed camera using photometric stereo. Stereo is a widely researched problem in computer vision and numerous algorithms have been proposed to solve for the stereo properties of a scene represented by images.
- **Scene understanding**: Scene understanding can be considered in the superset of many of the situations already discussed above. Image mosaicing can also be considered an application of scene understanding, as the goal is to reconstruct a full view of a scene from multiple images where each individual image only captures part of the scene.

In the proposed research paper, we highlight an efficient framework for homography based object tracking system with comparative analysis on all the prominent work done in the past in the same field. In section 2 we give an overview of related work which identifies all the major research work being done in this area. Section 3 highlights about the geometrical aspects of homography system. Existing homography techniques are discussed in Section 4. Proposed system is discussed in Section 5 followed by implementation in Section 6. Section 7 discusses about result discussions followed by performance comparison in Section 8 and finally in section 9 we make some concluding remarks.

II. Related Work

Jon Arrospide et al [3] have presented a full system for vehicle detection and tracking in highway environments given by an on-board camera has been presented. The system first detects vehicles on the basis of successive image rectification using plane-to-plane homographies. The framework presented for homography estimation and image alignment has been proven to be robust to errors in the measurement process, thus providing reliable and accurate vehicle detection. On the other hand, a method for tracking of instantaneous vehicle detections has been presented on the basis of an adaptive multidimensional particle filter. This filter is able not only to provide robust tracking of the vehicles in the image, but also to efficiently handle entry and exit management.

Guoqiang Hu et al [4] has demonstrated, an adaptive homography-based visual servo tracking controller is developed for the camera-in-hand problem using a quaternion formulation to represent rotation tracking error.

Ognian Boumbarov et al [5] presented a framework for determining the orientation of human faces with a fixed monocular camera which can be used for the purposes of the gaze tracking afterwards.

Vilas K. Chitrakaran et al [6] has presented the design of a vision-based controller for an under actuated, unmanned aerial vehicle (UAV) equipped with a pan-tilt camera unit (PTCU) to achieve the objective of following a leader vehicle autonomously.

Ran Eshel and Yael Moses [7] has suggested a method for simultaneously tracking all people in a dense crowd using a set of cameras with overlapping fields of view. To overcome occlusions, the cameras are placed at a high elevation and only people's heads are tracked.

Takeaki Iiyoshi and Wataru Mitsuhashi [8] has deals with three types of techniques for performing an image mosaicing: Homography estimation for determining geometrical relationships between the image pair, Expectation-Maximization algorithm for removing inconsistent overlapping region, and Graph Cuts for seamless stitching and background estimation. Experimental results indicate that the proposed technique is effective to synthesize a panoramic image from a series of narrow-field-of-view snapshots.

Farid Arvani et al [9] has proposed a new approach is proposed to estimate the depth of the target using active monocular stereo. The proposed method employs a tapped delay line (TDL) neural network to approximate the depth of the target.

Tiago Goncalves eta [10] has proposed the Euclidean homography matrix as visual feature in an image-based visual servoing scheme in order to control an aircraft along the approach and landing phase. With a trajectory defined in the image space by a sequence of equidistant key images along the glidepath, an interpolation in the homography space is also proposed in order to reduce the database size and ensure the required smoothness of the control task.

Chuantao Zang and Koichi Hashimoto [11] has presented their novel implementations of a GPU based Efficient Second-order Minimization (GPUESM) algorithm. By utilizing the tremendous parallel processing capability of a GPU, they have obtained significant acceleration over its CPU counterpart. Currently their GPUESM algorithm can process a 360×360 pixels tracking area at 145 fps on a NVIDIA GTX295 board and Intel Core i7 920, approximately 30 times faster than a CPU implementation.

Chuantao Zang and Koichi Hashimoto [12] has presented an improved automatic visual inspection system. In this system, homography based visual servo is used to accurately locate the camera position and attitude so that a template matching inspection can be realized.

Sachin Bangadkar et al [13] has proposed algorithm was executed on an IBM PC platform with 1GB RAM and Intel® Core™2 processor. The input video was of VGA resolution (640 X 480). The processing speed for calculating the initial mapping matrix was approximately 25msec. Once the mapping matrix is calibrated, the remaining frames were processed at a frame rate of about 15-17 FPS.

Evaggelos Spyrou, and Dimitris K. Iakovidis [14] has propose a method for the estimation of the orientation of the capsule based only on information extracted from WCE video frames. The proposed method involves image feature extraction, detection of feature correspondences between consecutive video frames, homography estimation, and homography decomposition into rotation and translation components.

Murray Evans et al [15] has proposes an approach to predicting where these ghost detections will occur, and provides a mechanism for suppressing their appearance.

Jan Herling et al [16] has presented PixMix, their pixel based approach to image inpainting and real-time Diminished Reality. Further, a real-time capable object selection and tracking algorithm has been introduced. Their inpainting approach allows for easy weighting between the spatial and the appearance term of the cost function in order to provide optimal inpainting results. The overall results showed fewer artifacts than other approaches, providing high-quality image inpainting in real-time.

Christopher Mei et al [17] have studied an alternative dubbed efficient second-order minimization (ESM) that provides second-order convergence for the cost of a first-order approach. They show how this algorithm can be applied to 3D tracking and provide variants with better computational complexities. These results have applications in motion estimation, structure from motion, visual serving and SLAM. The tracking algorithm was validated using an unidirectional sensor mounted on a mobile robot.

Hence, it can be seen that there were multiple work being carried out for the purpose of homography based framework. Majority of the approaches were efficient in detection process and tracking using 3D features where less focus on tracking features with increased dimensionality exists in homography which is majorly considered for preciseness of the backbone of visual tracking features in real-time application. Hence, this paper will propose a solution that addresses this issue.

III. HOMOGRAPHY

A homography is an invertible transformation from a projective space (for example, the real projective plane) to itself that maps straight lines to straight lines. Homography estimation is used for 3D analysis [18], [19], [20], [21], [22], [23], [24], [25], [26], mosaicing [27], camera calibration [28], [29], and more. The induced homography between a pair of views depends on the intrinsic and extrinsic camera parameters and on the 3D plane parameters [30]. While camera parameters vary across different views, the plane geometry remains the same. In this paper, we show how we can exploit this fact to derive multi-view linear subspace constraints on the relative homographies of multiple (2) planes, and for restricted cases of camera motion—on the collection of homographies of a single plane. First, we derive the basic homography notations which will be used later in the paper. Let $Q_1=(X_1, Y_1, Z_1)^T$ and $Q_2=(X_2, Y_2, Z_2)^T$ denote the 3D coordinates of a scene point with respect to two different camera views. Let $q_1=(x_1,y_1,1)^T$ and $q_2=(x_2,y_2,1)^T$ denote the homogeneous coordinates of its projection point in the two images. We can write

$$q_1 \cong C_1Q_1 \text{ and } q_2 \cong C_2Q_2. \quad (1)$$

where $\cong$ denotes equality up to a scale factor. C_1 and C_2 are $3x$ 3 matrices [30] which capture the camera's internal parameters and the projection.

Let π be a planar surface with a plane normal n_1, then $n_2^TQ_1= 1$ for all points $Q_2 \in \pi$ ($n_1=n_{1\pi} / d_\pi$), where $n_{1\pi}$ is a unit vector in the direction of the plane normal and d_π is the distance of the plane from the first camera center). The transformation between the 3D coordinates of a scene point $Q \in \pi$, in the two views, can be expressed by

$$Q_2=GQ_1 \quad (2)$$

$$\text{Where } G=R+t_1n_1^T \quad (3)$$

R is the rotation matrix capturing the change of orientation between the two cameras views, and t_1 is the translation between the two camera views. Therefore, the induced transformation between the corresponding image points is

$$q_2 \cong Aq_1 \quad (4)$$

$$\text{where } A=C_1(R+t_1n_1^T)C^{-1} \quad (5)$$

is the induced homography between the two views of the plane π. From (4), it is clear that, when A is computed from image point correspondences, it can be estimated only up to a scale factor.

IV. EXISTING HOMOGRAPHY TECHNIQUES

In video surveillance domain object detection and tracking is of central importance for any modern video surveillance and behavioral analysis system. Many techniques have been proposed for homography based object detection and tracking. The below are some of the techniques used.

- **3D Euclidean reconstruction**: The process uses both the background and moving objects in a video sequence. Initially for detecting the objects camera projection matrix for the frames are considered. Camera projection consists of camera rotation and translation frame. This method requires interframe homographies corresponding to a plane and method used is plane base self calibaration. In this method iteratively estimates intrinsic and extrinsic parameters and the ground plane in which camera projection considered to be of matrices of canonical forms. The newtons method are used for camera points refinement like rotation, camera center and point co-ordinates to minimize the total projection errors. However for estimating the 3D object, the author considers the static object while it is being observed by hypothetical moving virtual camera and the camera which shots the actual sequence is called real camera. The motion of virtual camera is a combination of both real camera and moving object and at next standard Structure from Motion (Sfm) methods are applied to obtain object shape. Based on these techniques author derives a set of equations by lemma method for object tracking. In this method the 3D instantaneous translation of the moving object must be parallel to the ground plane [31].
- **A Nonlinear Estimation Strategy**: The process uses to identify the Euclidean structure and velocity of a moving object using a monocular calibrated moving camera. The proposed algorithm relies on the availability of a reference image of the target object, captured at a known orientation relative to the camera. For the general case where both the camera and the object are in motion relative to an inertial frame, a single geometric length on the object is assumed to be known, and sensors mounted on the camera are assumed to provide camera velocity information. The proposed work addresses the problem using nonlinear system analysis tools. Equations for motion kinematics are first developed in terms of Euclidean and image-space information. To facilitate the development of an estimator for Euclidean coordinates of the feature points on the object i.e., the vectors relative to the object frame and relative to the camera frame for all feature points on the object are considered. The integral feedback estimation method is employed to identify the linear and angular velocity of object and next estimated velocities facilitate the development of a measurable error system that can be used to formulate a nonlinear least squares adaptive update law for Euclidean structure estimation.[32]
- **Wireless Capsule Endoscopy (WCE)**: WCE is a developing imaging technology for screening the gastrointestinal system. In this process method for capsule orientation estimation based only on image features, without the requirement of any external equipment is presented. The technique adopted in this

process is based on set of points that are extracted from the images i.e. set of salient points extracted from each video frames and next choose to extract the well-known and widely adopted SURF (Speeded-Up Robust Features) features, and use them in order to capture and represent the visual properties of video frames. The SURF feature extraction algorithm proceeds in two distinct steps. The first step detects local interest points, whereas the second step extracts a descriptor from the area surrounding each interest point and the first order Haar wavelet responses are computed with the use of integral images, resulting in a 64-dimensional feature vector. To estimate the visual properties is to estimate a transform between two consecutive frames and to detect similar features uses RANSAC (RANdom SAmple Consensus) algorithm. The estimation of a homography using RANSAC has been applied very effectively in tasks such as stereoscopic camera calibration. Once the homography is calculated it is decomposed into rotation matrix and a translation matrix to get the required result.[33]

- **Detection and Tracking**: The process multi moving object detection and tracking under moving camera is considered. Moving objects are detected by homography-based detection and author uses online-boosting algorithm to track the object. Initially author extracts features of images after that author applies Kanade-Lucas-Tomasi (KLT) feature tracker between two frames, and homography is obtained by RANSAC scheme. From homography compares all pixels between two frames and gets the residual pixels which appears in border of moving objects and refines pixels by morphological process. To track the objects which are detected author uses online-boosting algorithm to deal with the appearance of changes in objects tracked. Tracker is initialized or terminated when an observation model has higher or lower value than some threshold. The Gating function is used to measure the similarity between sizes of target and detection and also to measure distance between objects.[34]
- **Ground Plane Detection**: In this process reliable estimation of the homography between ground planes in successive images is presented. For estimation of the ground plane a single onboard camera is used. The author uses temporal coherence of the inter-frame plane-to-plane **homography** to construct a probabilistic prediction framework based on Kalman filtering for the computation of the homography. The method first calculates the planar homography directly from feature correspondences rather than computing ego-motion which involves a linear data estimation framework that provides a time-filtered estimation of the homography. In this method lane detectors are used to localize the regions. In this work, the estimation of the homography is modelled as a linear process and controlled by means of a Kalman filter. In effect, the homography between ground planes from frame to frame depends only on the position of the camera relative to the ground plane and the displacement and rotation of the camera. In order to evaluate the adequacy of the data to a Gaussian distribution a Kolmogorov–Smirnov test is performed over the elements of the homography matrix.[35]

V. PROPOSED MODEL

The proposed system introduces an efficient framework on homography based tracking of multiple moving objects. Our previous work [36] has started with focusing the issues of object detection on dense crowd taking small set of video dataset low level feature extraction. With presence of minor false positives in this work, the enhancement of multiple moving object is rendered in our second work [37] by using unscented Kalman filter. This work has highly sensitive to detect even a minor movement of the multiple moving objects with considerable larger size of dataset of frame sequences. However, the work doesn't consider usage of 3D views of the events in crowd. Hence, the proposed system considers usage of a pure homography based techniques where a large dataset of 1GB of frame sequences are considered to perform multiple moving object detection.

Typically, homographies are estimated between images by finding feature correspondences in those images. The most commonly used algorithms make use of point feature correspondences, though other features can be used as well, such as lines or conics. A 2D point (x; y) in an image can be represented as a 3D vector $x = (x_1, x_2, x_3)$ where $x = (x_1/x_3)$ and $y =(x_2/x_3)$. This is called the homogeneous representation of a point and it lies on the projective plane P^2. A homography is an invertible mapping of points and lines on the projective plane P^2. Other terms for this transformation include collineation, projectivity, and planar projective transformation. Hartley and Zisserman [19] provide the specific definition that a homography is an invertible mapping from P^2 to itself such that three points lie on the same line if and only if their mapped points are also collinear. They also give an algebraic definition by proving the following theorem: A mapping from $P^2 \rightarrow P^2$ is a projectivity if and only if there exists a non-singular 3*x*3 matrix H such that for any point in P2 represented by vector x it is true that its mapped point equals H_x. This tells us that in order to calculate the homography that maps each x_i to its corresponding x_0^i it is sufficient to calculate the 3*x*3 homography matrix, H.

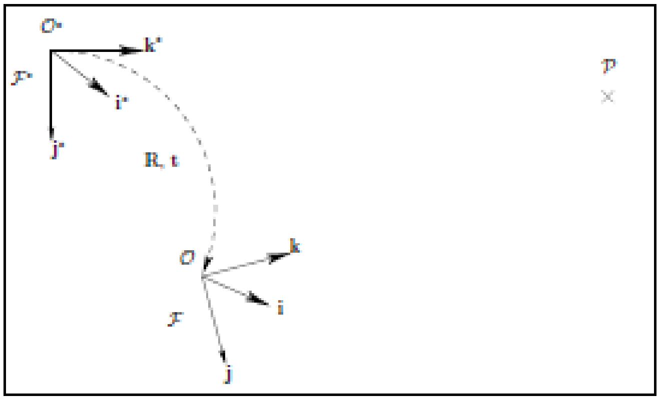

Figure 1 Geometry of Homography

Figure 1 illustrates the 3D space that is basically an Euclidean space with standard orthonormal frame. Let P be a 3D point and F = (O , i , j , k) be the "reference frame" of the Cartesian space. In the frame F , *P* has the coordinates X = [X Y Z]^T ∈ $\mathbf{R}^3$. Let R be the rotation matrix and t be the translation vector between F and a second frame F = (O, i, j, k) called the "current frame". R is a (3*x*3) matrix: R ∈ SO(3) the Special Orthogonal group (RR^T = I). t is a (3*x*1) vector: t ∈ R^3. In the frame *F*, *P* has the coordinates *X* = [X Y Z]^T ∈ R^3 verifying: X = R*X* + t. We define the "reference metric image", I $_m$, the plane perpendicular to the axis k situated at 1_m from the projection center O $_n$ The point *P* is projected on the I $_m$ using the perspective model in a projective point m ∈ P^2 that has the homogenous metric coordinates m = [x y 1]^T where: m = Z ^{1}X .

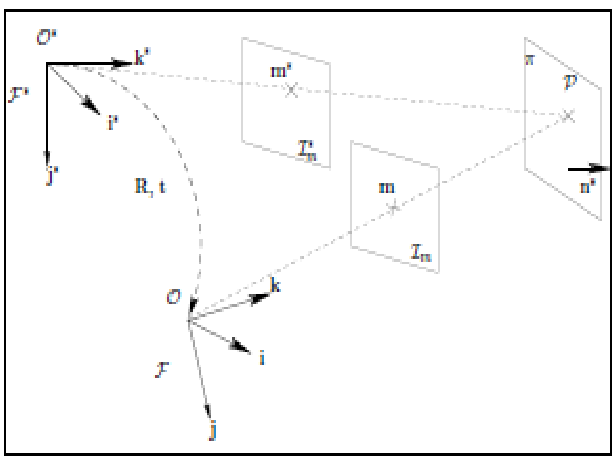

Figure 2 3D space with homogenous metric coordinates

Similarly, we define the "current metric image", I_m, the plane perpendicular to the axis *k* situated at 1_m from the projection center *O*. The point *P* is projected on the I_m using the perspective model in a projective point m∈ P^2 that has the homogenous metric coordinates m = [x y 1]^T where: m = Z ^{1}*X*. Therefore, m and m are "2 corresponding points" since they are the projection of the same 3D point *P*.

Suppose that we have P ∈ π, where π is a plane with a normal vector n (expressed in F) and its distance to O is d . If we suppose that n verifies: ||n || = √(n ^{T}n) = d ^{1}, the equation of the plane π becomes: n ^{T}*X* = 1.

Using the previous equations, we can establish a relationship between two corresponding points m and m :

$$\frac{Z}{Z^*} m = H_o m^* \qquad (6)$$

Where H_o=R + tn^{*T}.

Knowing m and a matrix H proportional to H_o: H =αH(α ∈ R), it is possible to compute m without knowing the ratio Z/Z^*.

It suffices to multiply m by H and then to normalize the third entry of m to 1. If, we write the matrix H as: H =[h_1 h_2 h_3]^T, the normalization can be written:

$$m = [\frac{h_1 m^*}{h_3 m^*} \frac{h_2 m^*}{h_3 m^*} .1]^T \alpha H m^* \qquad (7)$$

The matrix H is called "homography". It matches the points of the plane π between the two metric images I $_m$ and I_m. Given *N* non-collinear corresponding points (m $_i$, m_i) where i ∈ {1, 2, ...,N} and N ⩾ 4, it is possible to compute the homography matrix H. H is defined up to a scale factor. It has 9 entries (a (3*x*3) matrix) and only 8 degrees of freedom (3 for the rotation, 3 for the translation and 2 for the normal vector to the plane). The scale factor can be chosen with different manners. We choose the scale factor of H such that its determinant is equal to 1. Then, we have H ∈ SL(3) the Special Linear group. This choice is well justified since det(H) ⩽ 0 happens when the projection center O crosses the plane π, and we suppose that O and O are always in the same side of the plane π.

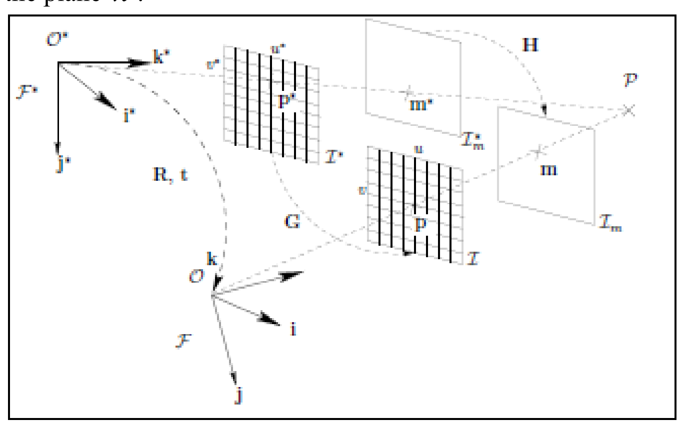

Figure 3 Complete geometry of proposed homography

In the "pinhole model", the camera performs a perspective projection. In the image I , the 3D point *P* is projected in p ∈ P^2: p = [u v 1]^T = Km and in the image I, it is projected in p∈P^2: p = [u v 1]^T = Km. The matrix K is a (3*x*3) upper triangular matrix that depends on the camera intrinsic parameters:

$$k = \begin{bmatrix} f & f_s & u_o \\ 0 & f_r & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where *f* is the focal in pixels, *s* is the skew (describing the orthogonality of the sensor axes), *r* describes the ratio of the pixels dimensions, [u_0 v_0 1] are the coordinates of the image of the principal point $m_o = [0\ 0\ 1]^T$ in pixels. Therefore, the two corresponding image points p and p verifies: p $\propto$ Gp . Where the image homography matrix G verifies:

$$G = KHK^{\ 1} \quad (8)$$

Given N non-collinear corresponding points (p $_i$, p_i) where i∈ {1, 2, ...,N} and N ≥ 4, it is possible to compute the image homography matrix G up to a scale factor. It is clear that G∈SL(3) if H∈SL(3).

VI. IMPLEMENTATION

The proposed system considers database from CrowdPET2009[38] where the dataset are multiple visual sensor sequences containing different crowd activities. All the frames are in compressed format of JPEG image sequences with approximately 90% of visual quality. Although there was an aggressive attempt of homography based techniques on 3D model of frame capturing, it was majorly found that tracking of multiple object moving on the crowd was a very challenging task. Therefore, we attempt to explore the possibility of capturing the frame-feeds from more than multiple cameras. Currently, the database in use has captured the event-feed from more than 6 visual sensors with a sophisticated and effective camera calibration. Therefore, performing this experiment in real-time will require the similar visual sensors that operation and synchronizes at same time instant, that posses initial challenges even before the real experiment starts. It is to be noted that while effort has been made to make sure the frames from different views are synchronized but due to technical issues there might be slight delays and frame drops in some cases. This dataset contains a set of sequences from different views Background images are provided of the monitored surveillance system. The scene is never completely empty of objects (people). However, it is envisaged that the data is useful for training some systems. The experiment has been conducted in standard 32-bit Windows operating system with 2.84 GHz Intel core-i3 processor with 4GB RAM size. Matlab is considered as programming tool to implement the thought of proposed homography based tracking. The algorithm description is as below:

Algorithm-1: Homography based tracking of multiple moving object
Input: CrowdPET2009 dataset
Output: Tracking of multiple moving objects.
START
1. Select a set of unit view of dataset of M categories.
2. For j=1 to N do
3. Consider reading the total dataset frame (n<M)
4. Sort the elements in category vectors
5. Perform object selection [i)Use Region of Interest, ii) Feature point selection]
6. Perform Homogenous transformation on 3*x*3 blocks
7. Estimate centroid of Reference frame
8. Generate translational matrix & average RMS
9. Generate Scaling Matrix
10. Find corresponding points matching.
11. Perform Bicubic interpolation
12. Initialize Homography matrix
13. Use Linear Transformation for similarity matching
14. Perform outlier detection
15. Optimize tracking using DLS
16. Apply SVD on H-matrix
17. Generate Hessian matrix for tracking points.
18. If feature point is inside ROI
19. Perform detection of landmark points
20. else repeat Step-14.
21. Perform Bounding Box.
22. Go to step-14.
23. Perform tracking on current plane.
24. Repeat step-6 to 23 for entire value of M.

END

The schematic diagram of the proposed system can be seen below in Fig 4.

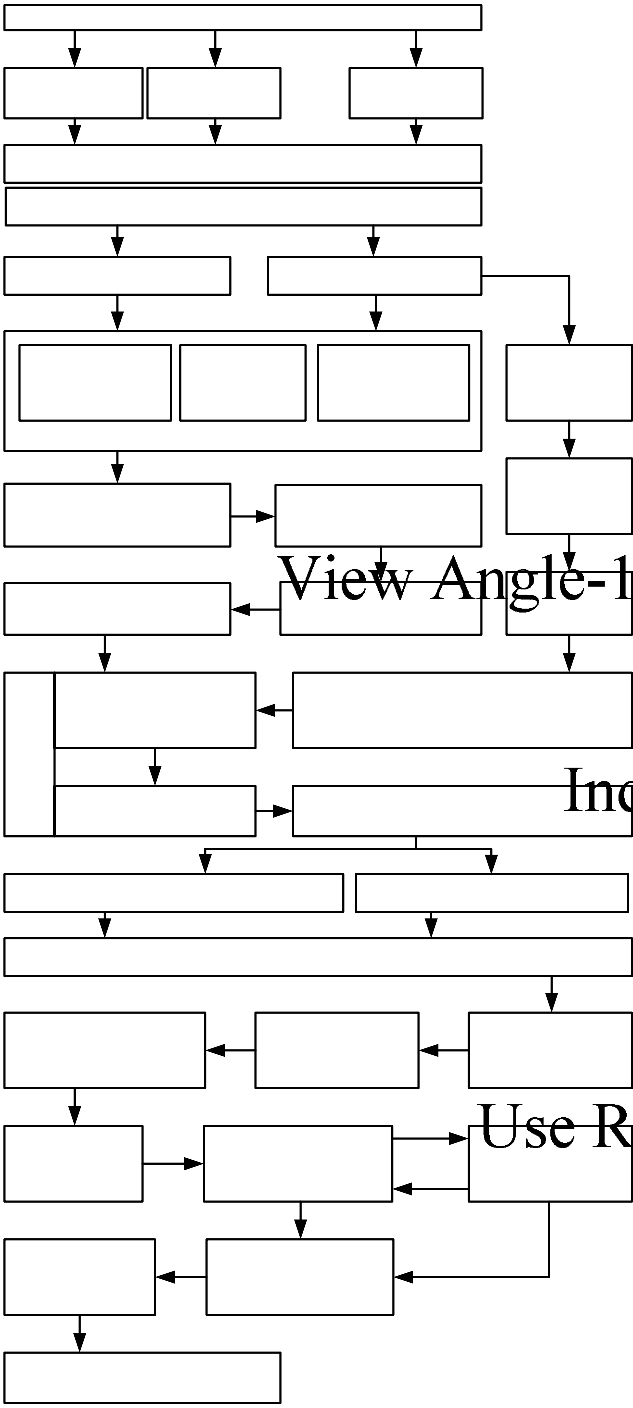

Figure 4 Schematic Diagram of Proposed Model of Homography

The schema (Fig. 4) takes the feed of frames recorded from multiple views and multiple time instants that is followed by initializing start and end point of the frame to ease down computation time. The proposed system however perform homography based tracking for multiple moving objects based on two process, e.g. i) Region of Interest and ii) Feature Points. It is to be noted that both the process are used for initiating user to select their object of interest and perform the check for proposed algorithm efficiency for tracking. The system uses considering current frame, 3x3 matrix of the current frame, and estimation of region. Usage of distinct feature points leads to generation of key-points on the selected object in the current frame where landmark descriptor is used to perform background registration. Finally, the next frame is selected where the corresponding points are matched and a matrix is generated that stores the best match pairs of the corresponding points in both the frames.

On the other hand, using region of interest centroid of the reference frame is initially extracted for the purpose of generating translational matrix and coordinates. The difference between two consecutive frames is estimated using root mean square value (RMS). For better accuracy in preliminary stage of h-matrix creation, it is important that scaling factor be introduced. Therefore, the next phase is focused on generation of scaling matrices that is again applied for corresponding matching points as well as simultaneously, Bicubic interpolation is applied on the geometric translation of the image (frame). Therefore, redundancies are removed to a greater extent between the errors in corresponding point between two frames. Homography computation can now be initiated after this step. The next steps performs two processes simultaneously e.g. i) outlier detection and ii) damped least-square method (DLS). The outlier detection algorithm is discussed below:

Algorithm-2: Homography based Outlier Detection
Input: feature extracted from 2 frames
Output: error (outlier) detection
START
1. Initialize iter=0
2. Initialize best_match= nil
3. Initialize best_matching_set=nil
4. Initialize best_error=infinity
5. While iterations < k //K=Number of iter performed yet.
6. Maybe_inliers= n randomly selected values from data
7. Maybe_model= model parameters fitted to maybe_inliers
8. Matching_set:= maybe_inliers
9. For every point in data not in maybe_inliers
10. If point fits maybe_model with an error smaller than t
add point to matching_set
11. If the number of elements in matching_set is > d
[d=the number of close data values required to assert that a model fits well to data]
12. this_model= model parameters fitted to all points in matching_set
13. this_error= a measure of how well this_model fits these points
14. If this_error < best_error
15. Store it until a better one is found
16. best_model= this_model
17. best_matching_set= matching_set
18. best_outliers= this_error
19. increment iterations
20. return best_model, best_matching_set, best_outliers
END

The DLS interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent. The prime role of DLS is therefore to perform optimization of tracking algorithm to most complicates sequences of frames. H-matrix is computed that is followed by isolation of inliers. A temporary buffer is created that stores every frequent values of outliers and inliers vector and hence the computation is optimized for better tracking by normalizing the corresponding points to average RMS. The next phase of computation generates A matrix where singular value decomposition is applied for addressing the issues of rotation and scaling in the moving objects. This process leads to Hessian matrix creation for tracking points that is repeatedly normalized depending on the window size of each frames. This evaluation finally yields a tracking features which undergoes a check for their similarity match with the previous corresponding points on previous current frame. If the tracking points matches, a bounding box is generated to encapsulate the object and tracking counter is set for new sequences of frames.

VII. RESULT DISCUSSION

The execution of the proposed framework leads to following results.

Test-Environment-1: This environment is selected for nice and clean view of the object. The main purpose is to check the initial efficiency of proposed homography based tracking system as shown in Fig. 5.

Figure 5 Results of Homography based multiple moving object detection

Test-Environment-2: In this test-environment, the difficult level for visual captures are slightly increased. The prime purpose of this test is to check the false positives in the detection and tracking of multiple objects. From Fig 6, it can be seen that false positives are highly reduced giving more sustainance and reliability of the bounding box.

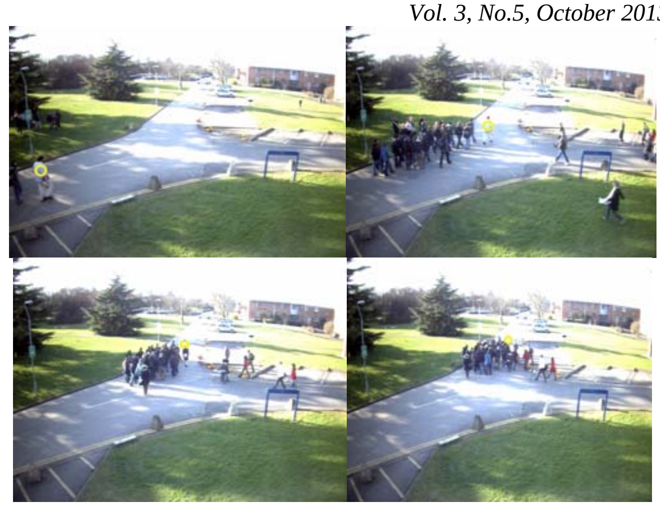

Figure 6 Results of Homography tracking for false positives

Test-Environment-3: The level of difficulty for this test-environment is higher as the object (of interest) totally disappear from the point of view. This test is conducted to check if the machine can perform detection and effective tracking when some of the corresponding point may highly differ in some intermediate frames and reappear again (Fig.7).

Figure 7 Results of Homography for object disappearance and reappearance

One of the interesting findings of the proposed system is that it addresses issues like false positives, partial or complete occlusion, and sudden disappearance and unpredictable reappearance in uncertain time interval. The robustness of the model is checked with large size of sequence frame data (1 GB) with more than 3000 frames underwent testing. In very few cases, the system gave false positive that is due to high variation of illumination in multiple homographic views (6 views of one event in total), however, the system shows fast computation when it comes to instant tracking. Another best part of the proposed result is extremely less storage is require to create a buffer for storing the corresponding points for tracking.

VIII. PERFORMANCE COMPARISON

For the purpose of comparative analysis, various prior significant researches works in the area of homography based object detection and tracking has been considered. The proposed system has also being compared and experimented with all the major significant previous research work like:

- **Prior-Work-1**: Chang Yuan and Gerard Medioni (2006) [31] have presented a novel method to obtain a 3D Euclidean reconstruction of both the background and moving objects in a video sequence.
    - *Comparison to our Approach*: The above discussed approach has consideration of 3D views where the scaling parameter is infrequently invariant. The input of the above discussed work is extremely less compared to us even being focused on real time event captures. No solution addressing occlusion, infrequent movement of objects are discussed in the work. Our approaches has deployed a large dataset to check if the proposed framework on homography is reliable with dimension being increased from 3-6 scale, whereas the work of Chang Yuan has focused the video feed of 3 dimensional.
- **Prior-Work 2**: Chitrakaran e.t. al. (2006) [32] have presented the development of a vision-based estimator for simultaneous determination of velocity and structure of an object (i.e., the Euclidean position of its feature points) for the general case where both the object and the camera are moving relative to an inertial frame of reference
    - *Comparison to our Approach*: The reason of selection of this work for comparative analysis is that the work has considered capturing the event feed from moving vehicle where the challenges are almost equivalent to our proposed model. But, the resultants of the work are quite dependent on the velocity of the vehicle with no discussion on dimensionality, false positive, and occlusion. The system doesn't seem to discuss the reliability of the homography based tracking for the cases that we have considered (e.g. infrequent mobility of multiple moving objects)
- **Prior-Work 3:** Spyrou and Iakovidis (2012) [33] have proposed a method for capsule orientation estimation based only on image features, without the requirement of any external equipment. The proposed method involves two steps: a) salient point detection and extraction of image features from these points, and b) detection of feature correspondences between consecutive frames and homography estimation**.**
    - Comparison to our Approach: The reason for selecting this work for comparative analysis is because the author has used the similar method like us to detect the similar features and find the correspondences between video frames above discussed approach has not considered complex background images. The method has provided some interesting result being performed on medical video datasets. However, the inliers detection performed in our model is highly sophisticated and the tracking being optimized at the end.
- **Prior-Work 4**: Kim and Kweon (2011) [34] has proposed a model that deals with multi moving object detection and tracking under moving camera. The algorithm detects and tracks multi moving objects without background modeling.
    - *Comparison to our Approach*: The work has used an Kanade-Lucas-Tomasi (KLT) feature tracker between two frames. Hence the detection and tracking rate is found to be 89.6%. However, consideration of multiple challenges that we are addressing in our proposed system are not explored in this work.
- **Prior-Work 5:** Arrospide e.t. al (2010) [35] have presented a robust method for ground plane detection in vision-based systems with a non-stationary camera. The proposed method is based on the reliable estimation of the homography between ground planes in successive images. This homography is computed using a feature matching approach, which in contrast to classical approaches to on-board motion estimation does not require explicit ego-motion calculation.
    - *Comparison to our Approach*: The reason for selecting this work is because of the simplistic approach of homography discussed by the author. The author uses Kalman filter for object tracking and region of interest, the method which is almost similar to our proposed system. However, the consideration of highly challenging test-environment is missing in the work.

The empirical effectiveness of our proposed algorithm can be derived by estimating the accuracy in tracking of the objects from the input frame sequence in the application designed. The proposed performance comparative analysis is based on the fact an efficient object tracking can be only reliable if it give better results with respect to multiple moving objects detection in 6 dimensional homographic feeds of the input. The following parameters are considered for this purpose:

- Actually Recognized Object Region (AROR) consisting of Object.
- False Identified Object Region (FIOR) which do not contain any object.
- Missing Object (MO) which ignores some object statistics.

The approach manually estimates the Original Object region (OOR) where the Object Identification Rate (OIR) can be evaluated as:

$$OIR = AROR / OOR$$

And Error in Object Identification rate (EOIR) as given by,

$$EOIR = FIOR / (AROR + FIOR)$$

Non-Object Identification Rate is given as

$$(NOIR) = MO / AROR$$

The comparative analysis is done by estimating the above empirical parameters for the proposed work with all the significant research work specified in this paper. Same sets of the input type and considerations are made towards the analysis done in 50 iterations.

Table 1. Comparative Performance analysis Parameters with 5 prior research work and proposed system.

Prior Work	OOR	AROR	FIOR	MO
Chang Yuan and Gerard Medioni (2006) [31]	500	393	87	79
Chitrakaran e.t. al. (2006) [32]	500	349	50	35
Spyrou and Iakovidis (2012) [33]	500	251	94	94
Kim and Kweon (2011) [34]	500	458	39	55
Arrospide e.t. al (2010) [35]	500	350	55	70
Proposed Work	500	485	38	37

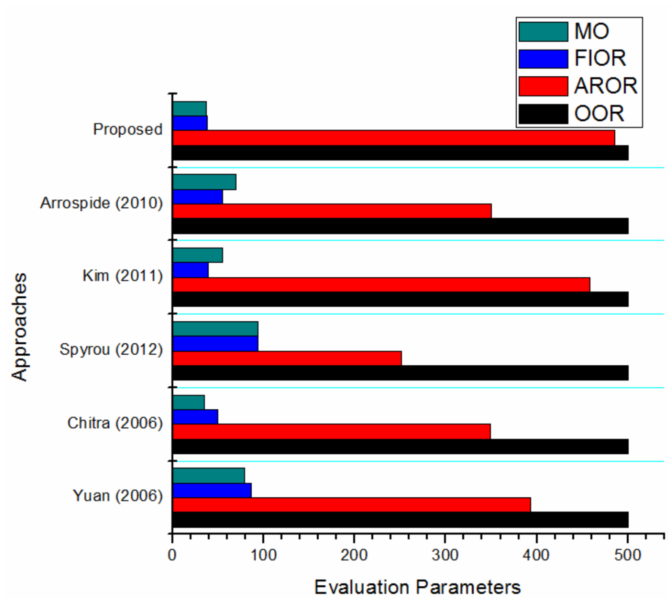

Figure 8 Comparative performance analysis-1

Table 2. Estimation of OIR, EOIR, and NOIR for all the 5 prior research work with proposed system.

Method	OIR	EOIR	NOIR
Chang Yuan and Gerard Medioni (2006)[31]	78.6	18.12	20.01
Chitrakaran e.t. al. (2006) [32]	69.8	12.53	10.00
Spyrou and Iakovidis (2012) [33]	50.2	27.72	37.45
Kim and Kweon (2011) [34]	91.6	7.8	12.00
Arrospide e.t. al (2010) [35]	70.6	13.5	20
Proposed Approach	**97.0**	**7.0**	7.6

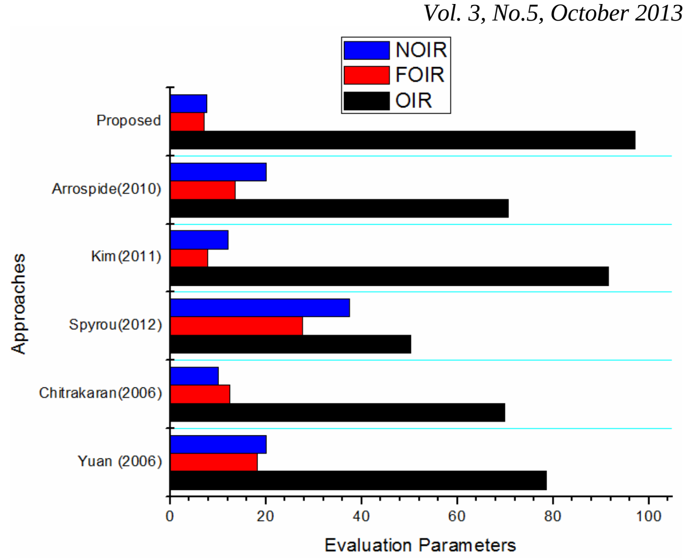

Figure 9 Comparative performance analysis-2

From the above table, it can be easily identified that proposed system has better Object Identification Rate (97%) as well as less Error in Object Identification rate (7%) as compared with the prior research work stated.

IX. CONCLUSION

The proposed system introduces a novel framework that can perform homography based multiple moving object detection by considering 6 dimensional views of datasets. The proposed system uses simple and cost effective algorithm that is considerable less computationally complex as compared to our previous techniques using either Lucas Kanade technique or unscented Kalman filter. The proposed system has a used a unique technique where both inliers and outliers are detected and are considered for increasing the precision of tracking system. The corresponding points for each frames are considered for evaluation to see that they precisely can compute the similarity matrixes followed by computation of homography matrix. The system uses negligible storage area illustrating that the system can be directly put into real time practices. Finally, 5 most prominent studies were considered that has previously addresses object detection using homography based approach. When comparison of performance analysis is done for proposed system to all the 5 considered prior work, it can be found that the proposed system exponentially increase the throughput by 97%.

REFERENCES

[1] Murphy, K., Torralba, A., Eaton, D., and Freeman, W., Object detection and localization using local and global Features, Toward Category-Level Object Recognition, Lecture Notes in Computer Science, Volume 4170, pp 382-400, 2006

[2] Dubrofsky, E., Homography Estimation, Masters Thesis, Carleton University, 2007

[3] Arrospide. J., Salgado, L., Nieto, M., Vehicle detection and tracking using homography-based plane rectification and particle filtering, IEEE

Intelligent Vehicles Symposium University of California, San Diego, CA, USA,pp. 21-24, 2010

[4] Hu, G., Gans, N., Coy-Fitz, N, and Dixon, W., Adaptive Homography-Based Visual Servo Tracking Control via a Quaternion Formulation, IEEE Transactions on Control Systems Technology, Vol. 18, no. 1, January 2010

[5] Boumbarov, O., Panev, S., Paliy, I., Petrov, P., and Dimitrov, L., Homography-Based Face rientation Determination from a Fixed Monocular Camera, The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 15-17, Prague, Czech Republic, September 2011,

[6] Chitrakaran, V.K., Dawson, D.M., Kannan, H., and Feemster, M., Vision Assisted Autonomous Path Following for Unmanned Aerial Vehicles, IEEE-2006

[7] Eshel, R., Moses, Y., Homography Based Multiple Camera Detection and Tracking of People in a Dense Crowd, IEEE-2008

[8] Iiyoshi, T., and Mitsuhashi, W., Homography-based Image Mosaicing for Automatically Removing Partial Foreground Objects, IEEE-2008

[9] Arvani, F., George, K. I., Fisher, A., Gosine, R.G., Range estimation using TDL Neural Networks and Application to Image-Based Visual Servoing, IEEE Conference on Soft Computing in Industrial Applications, Muroran, JAPAN, June 25-27, 2008.

[10] Goncalves, T., Azinheira, J., and Rives, P., Homography-based Visual Servoing of an Aircraft for Automatic Approach and Landing, IEEE International Conference on Robotics and Automation Anchorage Convention District, Anchorage, Alaska, USA, May 3-8, 2010,

[11] Zang, C., Hashimoto, K., Using GPUs to improve system performance in Visual Servo systems, IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, October 18-22, 2010.

[12] Zang, C., Hashimoto, K., GPU Acceleration in a Visual Servo System, Graduate School of Information Science, 2012

[13] Bangadkar, S., Dhane, P., Nair, S., and Kutty, K., Mapping Matrix for Perspective Correction from fish eye Distorted Images, IEEE-International Conference on Recent Trends in Information Technology, ICRTIT, MIT, Anna University, Chennai. June 3-5, 2011

[14] Spyrou, E., and Iakovidis, D.K., Retrieving landmark and non-landmark images from community photo collections. ACM Multimedia 2010: 153-162

[15] Evans, M., Li, L., Ferryman, J., Suppression of Detection Ghosts in Homography Based Pedestrian Detection, IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, 2012

[16] Herling, J., and Broll, W., PixMix: A Real-Time Approach to High-Quality Diminished Reality, IEEE International Symposium on Mixed and Augmented Reality, 2012

[17] Mei, C., Malis, E., and Rives, E, Efficient homography-based tracking and 3D reconstruction for single viewpoint sensors, IEEE-2008

[18] Luong, Q.T., and Faugeras, O., Determining the Fundamental Matrix with Planes, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 489–494, June 1998.

[19] Hartley, R.I. and Zisserman, A., Multiple View Geometry in Computer Vision. Cambridge Univ. Press, ISBN: 0521623049, 2000.

[20] Shashua, A. , Projective Structure from Uncalibrated Images: Structure from Motion and Recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, pp. 778–790, 1994.

[21] Szeliski, R., Torr, P.H.S., Geometrically Constrained Structure from Motion: Points on Planes, Proc. European Workshop 3D Structure from Multiple Images of Large-Scale Environments, pp. 171–186, 1998.

[22] Irani, M. , Anandan, P., and Weinshall, D. , From Reference Frames to Reference Planes: Multi-View Parallax Geometry Application, Proc. European Conf. Computer Vision, pp. 829–845, 1998.

[23] Fitzgibbon, A.W., and Zisserman, A., Automatic Camera Recovery for Closed or Open Image Sequences, Proc. European Conf. Computer Vision, pp. 310–326, 1998.

[24] Kanatani, K., Optimal Homography Computation with a Reliability Measure, Proc. IAPR Workshop Machine Vision, pp. 17–19, Nov. 1998.

[25] Pritchett, P., and Zisserman, A., Matching and Reconstruction from Widely Separated Views In 3D Structure from Multiple Images of Large-Scale Environments, Lecture Notes in Computer Science 1506, Springer-Verlag, pp. 219–224, 1998.

[26] Ma, J., and Ahuja, N., Dense Shape and Motion from Region Correspondences by Factorization, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 219–224, 1998.

[27] Irani, M. , Anandan, P., and Hsu, S. ., Mosaic Based Representations of Video Sequences and Their Applications, Proc. Int'l Conf. Computer Vision, pp. 605–611, Nov. 1995

[28] Triggs, B., Autocalibration from Planar Scenes, Proc. European Conf. Computer Vision, pp. 89–105, 1998.

[29] Zhang, Z., Flexible Camera Calibration By Viewing a Plane From Unknown Orientations," Proc. Int'l Conf. Computer Vision, Sept. 1999.

[30] Hartley, R.I., and Zisserman, A., Multiple View Geometry in Computer Vision. Cambridge Univ. Press, ISBN: 0521623049, 2000

[31] Yuan, C., and Medioni, G., 3D Reconstruction of Background and Objects Moving on Ground Plane Viewed from a Moving Camera, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006

[32] Chitrakaran, V. K., Dawson, D. M., Chen, J., and Kannan, H., Velocity and Structure Estimation of a Moving Object Using a Moving Monocular Camera, Proceedings of the 2006 American Control Conference Minneapolis, Minnesota, USA, June 14-16, 2006

[33] Spyrou,E., and Iakovidis, D.K., Homography-Based Orientation Estimation for Capsule Endoscope Tracking, IEEE-2012

[34] Kim, W.J., and Kweon, I.S., Moving Object Detection and Tracking from Moving Camera, 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2), 2011

[35] Arrospide, J., Salgado, L., Mohedano, N.R., Homography-based ground plane detection using a single on-board camera, IET Intelligent Transport Systems, 2009

[36] Pushpa, D., Sheshadri, H.S., Precise multiple object identification and tracking using efficient visual attributes in dense crowded scene with regions of rational movement, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2, March 2012

[37] Pushpa, D., Sheshadri, H.S., Semantic Analysis of Precise Detection Rate in Multi-Object Mobility on Natural Scene using Kalman Filter, Springer, 2013

[38] http://www.cvg.rdg.ac.uk/PETS2009/a.html