© 2016 IJSRSET | Volume 2 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Visual Tracking System for Target Representation and Localization

Sandhya Gopal Alhat, Dr. B. D. Phulpagar

[*1]Department of Computer Engineering, P.E.S. Modern College of Engineering, Savitribai Phule, Pune University, Maharashtra, India

ABSTRACT

Visual object tracking is the process of tracking the object in motion. Challenges in the object representation models and tracking algorithms have attracted researchers in the process of visual object tracking within the field of image processing and computer vision. We propose a representation of the target object depending on appearance that is based on local steering kernel descriptor (LSK) as well as color histogram data. We propose a tracking method that takes a video as input to the system. The target object to be tracked in the sequence on frames of the video input is initialized in the first frame. We compare the previous video frame as well as stored instance of object model to locate and track the target object in the next frame. By searching the frame region we try to localize the object in present frame which best resembles the input. The object model gets modified as the view of the object changes over the time hence triggering these modifications. The color histogram similarities between the target object and the surrounding background is utilized in order to subtract background. We are providing a mechanism to detect occlusion of the target object and handle it and continue tracking the given target object. For this purpose we are storing the LSK matrices of previous frame onto the stack and comparing it with the LSK matrices of the current recent frame, and using a threshold criterion. The proposed tracking method has resulted in improving the execution time and tracking accuracy and successfully tracks single and multiple target objects in motion under affine transformations and occlusion conditions.

Keywords : Color Histograms, Local Steering Kernels, Visual Object Tracking, Kalman Filter.

I. INTRODUCTION

Visual object tracking is an important task within the field of computer vision. It aims at locating a moving object or several ones in time. An algorithm analyses the video frames and outputs the location of moving target objects within the video frame. So it can be defined as the process of segmenting an object of interest from a video scene and keeping track of its motion, orientation, occlusion etc. in order to extract useful information by means of some algorithms. Its main task is to find and follow a moving object or several target objects in image sequences.

The proliferation of high-powered computers and the increasing need for automated video analysis have generated a great deal of interest in visual object tracking algorithms. The use of visual object tracking is pertinent in the tasks of automated video surveillance[10], autonomous robotic systems[17] traffic monitoring, vehicle navigation, human-computer interaction[18], healthcare[4] etc.

Tracking Techniques employ different target object representation methods based on object features, texture and shape models, or object contours, object position prediction and searches the target object in the next video frame. There are five models of object representation : appearance based, model based, contour based, feature based and hybrid. Appearance based tracking methods use visual information of object projection like color, texture, shape etc. on image plane. These methods deal with simple object transformations like translation and rotation but sensitive to illumination changes [19]. Model based methods use prior

information about object shape. These methods address problem of object tracking under illumination changes, change in object viewing angle and partial occlusion. But their computational cost is more. Also require implementation of detailed model for each type in scene [20]. Contour-based tracking methods track object by considering their outline as boundary contours. These methods enable the tracking of both rigid and non-rigid objects[21]. Feature-based methods used to describe objects[22]. This process follows different steps as recognizing and tracking the object by extracting elements, to cluster elements in higher level features and to match these extracted features between images in successive frames. These methods perform well in partial occlusion and in tracking very small objects. The major drawback of feature based methods is the correct distinction between the target object and background features. Hybrid methods for object tracking use the advantages of the above mentioned methods, by combining two or more tracking methods[23] .Usually, feature-based methods are employed first, for object detection and localization. Then, region-based techniques are used to track its parts. The problem with these methods is their high computational complexity.

Automated video surveillance deals with real time observation of people or vehicles in busy or restricted environments leading to tracking and activity analysis of the subjects in the field of view. There are three key steps in video surveillance: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior.

Visual object tracking follows the segmentation step and is more or less equivalent to the "recognition" step in the image processing. Detection of moving objects in video streams is the first relevant step of information extraction in many computer vision applications. There are basically three approaches in visual object tracking. Feature based method aim at extracting characteristics such as points, line segments from image sequences, tracking stage is then ensured by a matching procedure at every time instant. Differential methods are based on the optical flow computation, i.e. on the apparent motion in image sequences, under some regularization assumptions. The third class uses the correlation to measure inter-image displacements. Selection of a particular approach largely depends on the domain of the problem.

The development and increased availability of video technology have inspired a large amount of work on object tracking in video sequences in recent years.

Two major components can be distinguished in a typical visual tracker. Target Representation and Localization is mostly a bottom-up process which has also to cope with the changes in the appearance of the target. Filtering and Data Association is mostly a top-down process dealing with the dynamics of the tracked object, learning of scene priors, and evaluation of different hypotheses. The way the two components are combined and weighted is application dependent and plays a decisive role in the robustness and efficiency of the tracker. For example, face tracking in a crowded scene relies more on target representation than on target dynamics, while in aerial video surveillance, the target motion and the ego-motion of the camera are the more important components. In real-time applications only a small percentage of the system resources can be allocated for tracking, the rest being required for the pre processing stages or to high-level tasks such as recognition, trajectory interpretation, and reasoning. Therefore, it is desirable to keep the computational complexity of a tracker as low as possible.

The remainder of this paper is organised as follows: Section II represents the related works in the field of visual object tracking. Section III represents the architecture, description, algorithm and mathematical model of proposed method. Section IV describes the dataset. Section V presents the results and finally Section VI contains the conclusions of the proposed work.

II. RELATED WORK

O. Zoidi, A. Tefas and I. Pitas have proposed a visual object tracking framework which employs appearance based representation of the target object[1]. The tracker extracted a representation of the target object and the video frame based on local steering kernels(LSK) and color histograms(CH) at video frame $t-1$ and tried to find its location in the frame *t*, which best suit the target object. Each significant change in the appearance of the target object was stored in a stack, representing the target object model. The visual resemblance was determined with respect to the detected target object in the previous video frame and the last inserted target object instance in the object model stack.

Authors K. Zhang, X. Gao, D. Tao and X. Li have given single image super resolution (SR) approach by extending the non-local kernel regression(NLKR) model to an effective regularization term[2]. A maximum a posterior probability(MAP) estimation of a least squares minimization problem is solved by gradient decent for the desired high resolution(HR) image. The test outcomes on both the simulated and real low resolution(LR) images show that the proposed method can produce promising results with fewer artifacts along edges and more plausible details in textural regions.

A feature-based approach of visual saliency detection for natural color images has been represented by X. He, H. Jing and X. Niu[3].They have considered two features to represent the local information surrounded each pixel: structure information and color information. Local steering kernel features are extracted as structure information and color values of local region around each pixel are extracted as color information.

An e-healthcare system based on video content analysis and quality-driven content-aware wireless streaming for remote human gait tracking has been developed by H. Luo, S. Ci, D. Wu, N. Stergiou and K. Siu[4]. The system can significantly reduce the reliance on traditional marker-based gait data collection facilities, providing a low-cost high-accuracy gait tracking system. A distortion-delay framework has been designed to optimize the wireless streaming for delay-bounded retrieval of the collected video data, where key system parameters residing in different network layers are jointly optimized in a holistic way to achieve the best user-perceived video quality over wireless environments.

An automatic tracking algorithm based on the hybrid strategy uses interactions between video objects and their regions [5]. Regions are objects' areas that are homogeneous with respect to a set of features such as motion, color, and texture. Regions have been represented by their region descriptors. Each region descriptor is tracked over time as representative of the corresponding video object.

A contour-based non-rigid object tracking method along with color and texture models generated for the object and the background regions, maintains a shape prior for recovering occluded object parts during the occlusion[6]. The shape priors encode the motion of the object and are built online. The energy functional is derived using a Bayesian framework and is evaluated around the contour to suppress visual artifacts and to increase numerical stability.

A multi-features fusion tracking algorithm based on local kernels learning is proposed in [7]. Histograms of multiple features are extracted based on the sub-image patches within the target region and the features fusion weights are calculated respectively for each patch according to the discrimination of features. A fast and yet stable model updation method is elaborated. The authors Hainan Zhao and Xuan Wang have also demonstrated how edge information can be merged into the mean shift framework without having to use a joint histogram. This is used for tracking objects of varying sizes.

The author Shauhua Kevin Zhou proposed an adaptive method for visual tracking which stabilizes the tracker by embedding deterministic linear prediction into stochastic diffusion [8]. Numerical solutions have been provided using particle filters with the adaptive observation model arising from the adaptive appearance model, adaptive state transition model, and adaptive number of particles. Occlusion analysis is also embedded in the particle filter.

III. PROPOSED METHOD

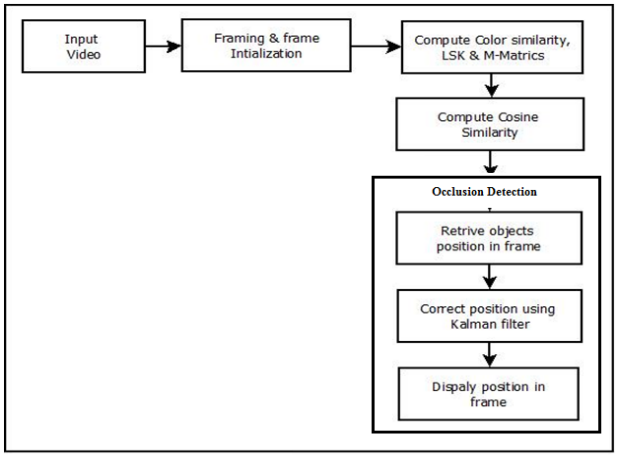


Figure 1. Architecture of the Proposed Tracking Method

The architecture diagram of the proposed tracking method is shown above in Figure1 that takes a video as an input to the system. It requires that the transformations in the target object between two consecutive frames are rather small. After taking the video as input, the video is divided into number of

frames. After that the initialization of frame is performed. Then the first frame is taken as an input to the next process. The object is selected from the first frame itself as well as the position of the object in the frame by making use of the Kalman filter. Then the color similarity of the given frame is computed using color histogram and computation of the LSK for the present as well as previous frame is performed. After the LSK values the M-Matrix calculation is performed and the highest value of the M-Matrix is retrieved.

Then for tracking the object we have compared the similarity of both the previous as well as current M-Matrix values using Cosine Similarity algorithm. At the time when the generated value of cosine similarity is greater than presented threshold we will retrieve the position of the object in the frame. After retrieving the frame, to correct the position we made use of Kalman filter and displayed the position of object in frame. Then go for next image from the framing.

If the generated value of cosine similarity is smaller than presented threshold then we rotate the frame by +/- 0.2 degree also the frame is rescaled by +/- 10%. This condition occurs at the time when the object in the frame gets rotated or turns in different direction.

A. Algorithm

Implemented algorithm:-

1. Begin
2. Get input video
3. Done framing of input video
4. Initialize frame
5. Get first image from framing
6. Set the object image
7. Find the position of object in frame by using Kalman Filter
8. Compute colour similarity
9. Compute LSK of current & previous frame
10. Compute M-Matrix
11. CSofPrevAndLastDetected=Cosine similarity of current object and previous tracking object.
12. Retrieve the highest value from M-matrix
13. If CSofPrevAndLastDetected > 0.6
14. Compute Cosine similarity of previous M- matrix value &current M-matrix value(CS) \
15. If CS> Threshold value Then retrieve the position of object in frame
16. Correct the position with Kalman filter Compute position value
17. Display position in frame
18. Get next frame from framing.
19. Else Rotate frame by +/- 0.2 degree and Rescale frame by +/- 10% and go to Step 9.
20. Else pos = Null
21. (If pos==Null) then Skip next 6 frames and Get the 7thFrame for processing & double the search region for only current seventh frame.
22. Make the correction of co-ordinate which jklman predicted or system generated by jkalman filter.
23. Pass the co-ordinate to display the object in frame.
24. Resize the search region from double to original height and width.
25. Get Next frame for processing & Go to Step (9) until end of frames.
26. End

B. Mathematical Model

Mathematical Model for the proposed method is as follows:

Let S, be the implemented system which can be represented as

$$S = \{\{I\}, \{P\}, \{O\}\}$$

Where,

I- (Input video)

P-(Functions used)

O-(Detected target object)

Where,

$$P = \{f_1, f_2, f_3, f_4\}$$

f_1 - Cosine Similarity($c(h_1, h_2)$)

f_2 - Confidence Level(B_t)

f_3 - Local Steering Function(LSK)

f_4 - Final Decision Matrix (M)

- The cosine similarity between two histograms h1, h2 - R^{256} is given by,

$$c(h_1, h_2) = \cos(\theta) = \frac{(h_1.h_2)}{||h_1||.||h_2||} \quad (1)$$

Where,

Dot (.) Defines the inner product of h1, h2, θ denotes the angle they form and ||.|| denotes the Euclidean norm. The cosine similarity takes values in the range [−1, 1].

- The confidence level is defined as

$$B_t = 100.\begin{cases} 1 - \frac{|M' - M_{max}|}{|M' - M_{min}|}, & \text{if } |M' - M_{max}| < |M' - M_{min}| \\ \frac{1}{2}, & \text{if } |M' - M_{max}| = |M' - M_{min}| \\ \frac{|M' - M_{min}|}{|M' - M_{max}|}, & \text{if } |M' - M_{max}| > |M' - M_{min}| \end{cases} \quad (2)$$

Where,

B_t is confidence level
M', M_{max}, and M_{min} as the mean, maximal, and minimal values of M_{CH} entries, respectively.

- Local Steering Kernels(LSK) vector formation:

$$K_i(p) = \frac{\sqrt{\det(C_i)}}{2\pi} \exp\left\{-\frac{(p_i - p)^T C_i (p_i - p)}{2}\right\}, i = 1, \ldots\ldots, M^2 \quad (3)$$

Where,

Image pixel p, neighboring pixel p_i, $p = [x, y]^T$ are the pixel coordinates. C_i is the Covariance matrix.

- The final decision matrix is computed by

$$M = [(1 - \lambda)M_{LSK1} + \lambda M_{LSK2}] * B_{CH} \quad (4)$$

Where $0 \leq \lambda \leq 1$ is a suitably chosen weight, M_{LSK1}, M_{LSK2} are the LSK similarity matrices for the last detected object and the last object instance, respectively, B_{CH} is the binary CH similarity matrix, and $*$ denotes the element-wise matrix multiplication.

IV. DATASET

The proposed method works on videos with varying memory sizes and video extensions like .mpg, .avi, .mjpeg, .mp4 etc.

Video	Size(MB)
EnterExitCrossingPaths1cor.mpg	2.15
OneStopMoveEnter1cor.mpg	8.76
OneStopMoveNoEnter2cor.mpg	5.73

Table 1. Dataset of Videos worked upon by the Proposed Tracking Method

V. RESULTS AND DISSCUSSION

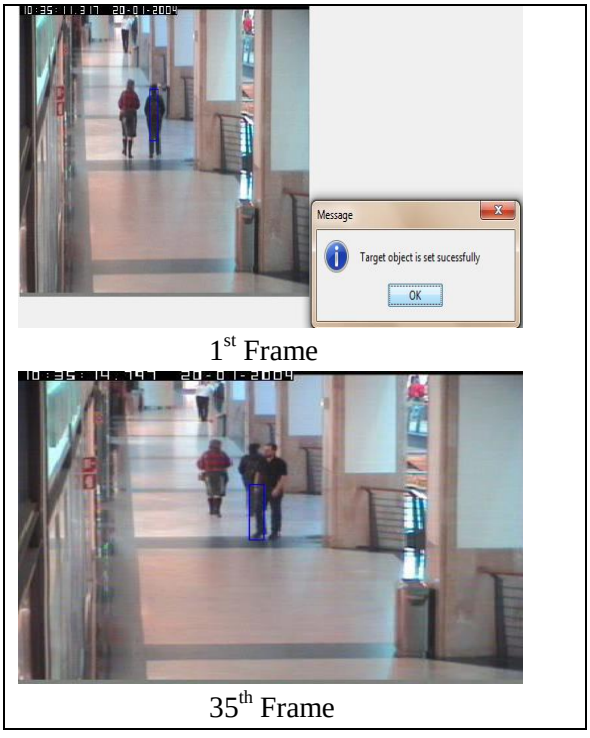
1. Home Screen

Figure 2. Home Screen

Figure 2 shows the home screen of the project from where the user can select the video for tracking the object.

2. Comparing Tracking Results of the Existing

CH-LSK Tracker and Proposed Tracking Method:



1st Frame

35th Frame

41st Frame

45th Frame

Figure 3. Blue Bounding Box : Tracking Results of Existing CH-LSK Tracker.

As shown in Figure 3. 1st frame: Target Object given as input for tracking, in 35th frame: Occlusion takes place, 41st & 45th frame : shows Tracker Deviation along with Occlusion causing object.

1st Frame

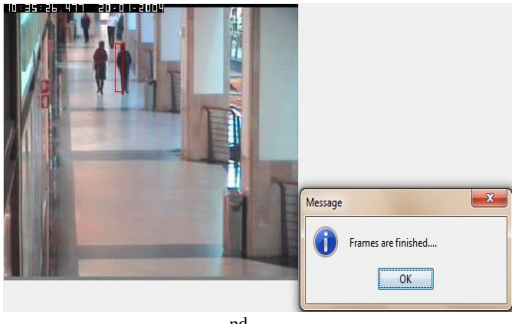
35th Frame

37th Frame

38th Frame

49th Frame

63rd Frame



152nd Frame

Figure 4. Red Bounding Box : Tracking Results of the Proposed Tracking Method.

As shown in figure 4. 1st frame: Target Object given as input for tracking, in 35th frame: Occlusion takes place, 38th frame: shows Tracker Deviation in the direction of the Occlusion causing object. But as seen in the 49th & 63rd frame: the red bounding box: proposed tracking method tries to resume tracking the given target object , 152nd frame: shows end of frames and thus the proposed tracking method successfully tracks the given target object under occlusion.

3. Accuracy Graph

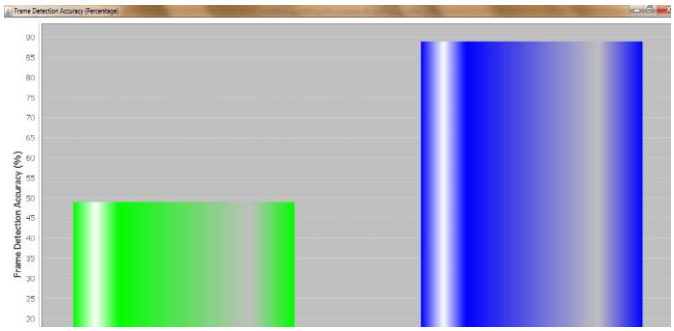


Figure. 5. Frame Detection Accuracy Graph

Above figure 5. shows the Frame Detection Accuracy graph in which the accuracy of the proposed tracking method is represented by blue colored bar and the existing system represented by green colored bar. We can clearly see that the accuracy of proposed method is better than the accuracy of the present or existing CH_LSK tracker.

4. Comparative Results for Execution Time & Frame Detection Accuracy

	Existing CH-LSK Tracker	**Proposed Tracking Method**
Execution Time in Milliseconds	104711	74700
Execution Time in Minutes	1.475	1.245
Frame Detection Accuracy (%)	49	89

Table 2 : Comparative Results for Execution Time & Frame Detection Accuracy for video input EnterExitCrossingPaths1cor.mpg

Existing CH-LSK Tracker	**Proposed Tracking Method**	**Existing CH-LSK Tracker**	**Proposed Tracking Method**
23553	15888	42	82
98448	93822	44	84
99214	90612	46	86
150854	71232	58	98
115982	100499	43	83
104711	74700	49	89

Table 3 : Comparative Results for Execution Time & Frame Detection Accuracy for different videos and target object inputs.

5. Multiple Objects Tracking Result of the Proposed Method:

2nd Frame

37th Frame

47th Frame

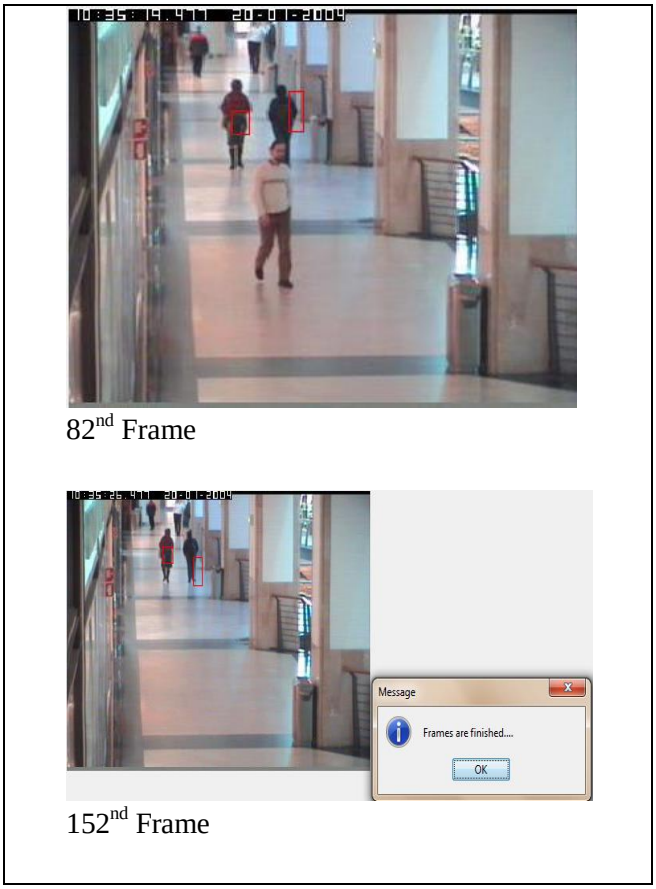


Figure 5. Red Bounding Box : Multiple Objects Tracking Result of the Proposed Method.

As shown in figure 5. 2nd frame: Proposed method tracking the two target objects inputs shown by red colored bounding boxes, in 37th & 47th frame: two target objects show occlusion, 82nd frame: the red bounding box: proposed tracking method continues tracking the two given target objects, 152nd frame: shows end of frames and thus the proposed tracking method successfully tracks the given multiple target objects under occlusion.

5. Accuracy Graph for Multiple Objects Tracking:

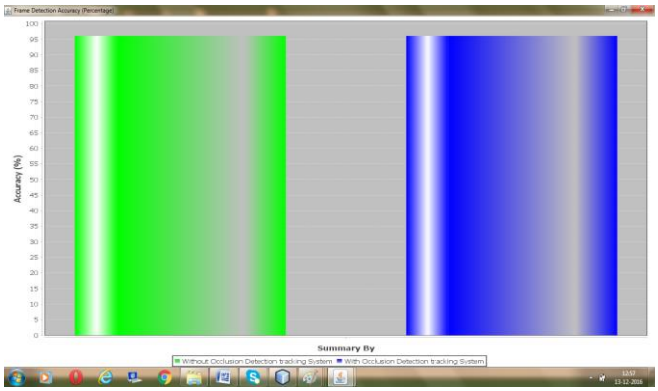


Figure 7. Frame Detection Accuracy Graph

Above figure 7. shows the Frame Detection Accuracy graph in which the accuracy of the proposed tracking method for without occlusion is represented by green colored bar and with occlusion is represented by blue colored bar. We can clearly see that the proposed tracking method is successful is tracking multiple target objects.

VI. CONCLUSION

In this paper we propose a visual object tracking method, which offers a representation of the targeted object depending on appearance which is depending on local steering kernel descriptors (LSK) and color histograms data. Unlike the existing CH-LSK tracking method, the proposed tracking method is able to track more than one target objects and has enhanced the tracking performance in terms of execution time and accuracy. The proposed tracking method handles the case of full occlusion and successfully tracks multiple target objects.

VII. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my Guide Prof. (Dr.) B. D. Phulpagar and M.E. Coordinator Prof. (Ms.) D. V. Gore, P.E.S. Modern College of Engineering, Pune-05 for their consistent guidance, valuable suggestions and support. I am also thankful to Prof. (Dr.) S. A. Itkar(H.O.D) and Prof. (Mrs.) M. K. Kulkarni for their guidance, support and motivation. I also wish to sincerely thank all those people who have directly or indirectly helped me in this work.

VIII. REFERENCES

[1]. O. Zoidi, A. Tefas and I. Pitas, "Visual Object Tracking Based on Local Steering Kernels and Color Histograms," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 5, pp. 870-882, May 2013.

[2]. K. Zhang, X. Gao, D. Tao and X. Li, "Image super-resolution via non-local steering kernel regression regularization," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 943-946.

[3]. X. He, H. Jing, Q. Han and X. Niu, "A Saliency Detection Approach to Combine LSK and Color for Color Image," Intelligent Information Hiding

and Multimedia Signal Processing (IIH-MSP), 2011 Seventh International Conference on, Dalian, 2011, pp. 129-132.

[4]. H. Luo, S. Ci, D. Wu, N. Stergiou and K. Siu, "A Remote Markerless Human Gait Tracking for E-Healthcare based on Content-aware Wireless Multimedia Communications", IEEE Wireless Commun., vol. 17, no. 1, pp. 44-50, Feb. 2010.

[5]. A. Cavallaro, O. Steiger and T. Ebrahimi, "Tracking video objects in cluttered background", IEEE Trans. Circuits Syst. Video Technol., vol. 15, pp. 575-584, Apr. 2005.

[6]. A. Yilmaz, X. Li and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras", IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 11, pp. 1531-1536, Nov. 2004.

[7]. J. Jeyakar, R. V. Babu and K. Ramakrishnan, "Robust object tracking with background-weighted local kernels", Comput. Vision Image Understanding, vol. 112, no. 3, pp. 296-309, 2008

[8]. S. Zhou, R. Chellappa and B. Moghaddam, "Visual tracking and recognition using appearance: Adaptive models in particle filters", IEEE Trans. Image Process., vol. 13, no. 11, pp. 1434-1456, Nov. 2004.

[9]. A Survey on Moving Object Tracking in Video - Barga Deori, Dalton Meiti Thounaojam, International Journal of Information Theory(IJIT), Vol3, July 2014

[10]. J. Wang, G. Bebis, and R. Miller, "Robust Video-Based Surveillance By Integrating Target Detection With Tracking," in *Proc. Conf. CVPRW OTCBVS*, June 2006.

[11]. A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in Proc. IEEE Conf. CVPR, Sep. 2006, pp. 798–805.

[12]. R. Sathya Bharathi ,"Visual Object tracking mechanism"IOSR-JCE, Vol 16 Issue 3 May-June2014.

[13]. O. Zoidi, A. Tefas , I. Pitas,"Visual Object Tracking Based on the Object's Salient features with Application in Automatic Nutrition Assistance", IEEE International Conference, 31 August 2012.

[14]. Nijad Al-Najdawi, Sara Tedmori, Eran Edirsinghe, and Helmut,"An Automated Real-Time People Tracking System Based on KLT Features Detection", IAJIT, vol .9, January 2012.

[15]. Ramsey Faragher,"Understanding the Basis of the Kalman Filter via a Simple and Intuitive Derivation,"IEEE Signal Processing Magazine Sept. 2012.

[16]. Anand Singh Jalal, Vrijendra Singh,"State of – the- Art in Visual Object Tracking", Informatica 36 (2012) 227-248.

[17]. N. Papanikolopoulos, P. Khosla, and T. Kanade,"Visual tracking of a moving target by a camera mounted on a robot: A combination of control and vision,"IEEE Trans. Robot. Autom., vol. 9, Feb.1993.

[18]. G. R. Bradski."Computer vision face tracking for use in a perceptual user interface,". in Proc. IEEE Workshop Appl. Comput. Vision, 1998.

[19]. C. Yang, R. Duraiswami, and L. Davis,"Efficient Mean- Shift Tracking via a new similarity measure,"in Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn., vol. 1, Jun. 2005, pp. 176–183.

[20]. D. Roller, K. Daniilidis, and H. H. Nagel,"Model-based object tracking in monocular image sequences of road traffic scenes,"Int. J. Comput.Vision, vol. 10, pp. 257– 281, Mar. 1993.

[21]. A. Yilmaz, X. Li, and M. Shah,"Contour-based object tracking With occlusion handling in video acquired using mobile cameras,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 11, pp. 1531–1536, Nov.2004.

[22]. L. Fan, M. Riihimaki, and I. Kunttu,"A feature-based object tracking approach for realtime image processing on mobile devices,"in Proc.17th IEEE ICIP, Sep. 2010, pp. 3921–3924.

[23]. L.-Q. Xu and P. Puig,"A hybrid blob- and appearance- based framework for multi-object tracking through complex occlusions,"in Proc. 2nd Joint IEEE Int. Workshop Visual Surveillance Perform. Evaluation Tracking Surveillance, Oct. 2005, pp. 73–80.