Yt α Dist (q, q x),
Where q = reference color histogram.
qx = color histogram of predicted location.

Particle Filter Iteration [2]

Steps:

- Initialize xt for first frame.
- Generate a particle set of N particles $\{x^m t\}m=1..N$
- Prediction for each particle using second order autoregressive dynamics.
- Compute histogram distance.
- Weigh each particle based on histogram distance.
- Select the location of target as a particle with minimum histogram distance.
- Sampling the particles for next iteration.

Initialize *a sample set* S0 = $\{\upsilon_0^j, 1)\}^J_{j=1}$ *according to prior distribution* $p(\upsilon_0)$.
For t = 1, 2, . . .
For j = 1, 2, . . . , J
Resample $S_{t-1} = \{\upsilon^{(j)}_{t-1}, w^{(j)}_{t-1}\}$ *to obtain a new sample* $(\upsilon'^{(j)}_{t-1}, 1)$.
Predict *the sample by drawing* $U^{(j)}_t$ *for* U_t *and computing* $\upsilon^{(j)}_t = F_t(\upsilon^{(j)}_{t-1}, U^{(j)}_t)$.
Compute *the transformed image* $Z^{(j)}_t$.
Update *the weight using* $w^{(j)}_t = p(Y_t|\upsilon^{(j)}_t) = p(Z^{(j)}_t|\upsilon^{(j)})$.
End
Normalize *the weight using* $w^{(j)}_t = w^{(j)}_t / £^J_{j=1} w^{(j)}_t$
End[2]

D. Kalman Filter

The Kalman filter [13] is also known as linear quadratic estimation (LQE). It is an algorithm that uses a series of measurements observed over time. Measurements can contain noise (random variations) and other inaccuracies, and it also produces estimates of unknown variables that are more precise The Kalman filter operates recursively on streams of noisy input data and produce a statistically optimal estimate of the underlying system state.

$$K_t = \hat{P_t} H^T (H\hat{P}_t^T H^T + R)^{-1} \quad (4)$$

$$\hat{x_t} = \hat{x}_{t-1} + K_t (z_t - H\hat{x_t}) \quad (5)$$

$$P_t = (I - K_t H) P_t \quad (6)$$

The matrix K_t is called the Kalman gain and is chosen such that minimizes the a posteriori error covariance P_t. The object will then be searched in a search region centered at the predicted position x_t. The size of this region varies according to the expected maximal object velocity, the object size and the reliability of the predicted position[15].

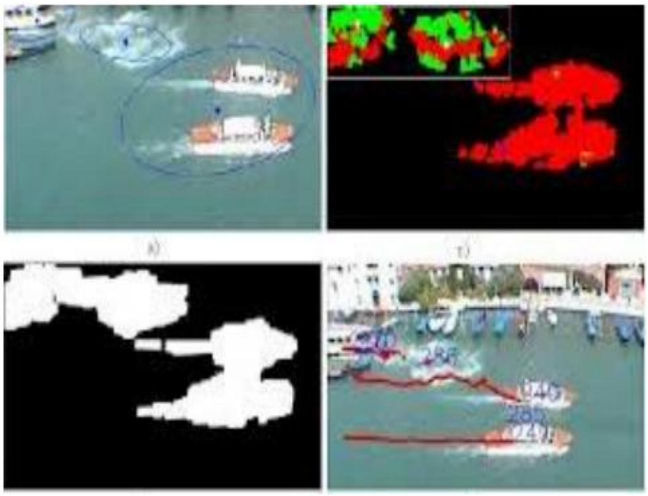

Figure 1: Kalman Filter Tracking

E. Kanade Lucas Tomasi(KLT)Tracking

The KLT algorithm is a feature detection and tracking process.

Feature tracking: Sum of Squared Difference (SSD) criteria applied to find the feature point whose window minimizes the following energy function:

$$E_t(dx,dy) = \sum [I(x+dx, y+dy, t+dt) - I(x,y,t)]^2 \quad (7)$$

Feature quality: If the quality of a tracked feature point decreases below a chosen threshold, that point is removed from consideration. To compensate, new features are identified in the same window. The feature point with the minimum criteria is retained if its SSD is below a certain threshold.

The KLT algorithm can be divided into two main parts. During the detection process, salient feature points are found and added to the already existing ones. Afterwards, in the tracking process for each feature point its corresponding motion vector is calculated. In the following, we describe each part of it in more detail.

Note that in the following Greek letters denote scalars, lowercase letters denote column vectors and uppercase letters denote matrices. We denote *I* as the current image and *J* as the immediately next image in the sequence. We write $\nabla I = \partial I / \partial (x,y)$ as the spatial image gradient of *I*, which is typically done with the Sobel or Sharr operator for robustness. Also we define *W(p)* as a small rectangular region centered at a given point *p*. Typically *W(p)* will be a 5 x 5 or 7 x 7 pixel neighbourhood. As the tracking is done with sub-pixel precision, p will have non-integer coordinates. Its neighbours are then calculated using bilinear interpolation.

Feature Point Detection
The task here is to detect new feature points in a given image *I* and add them to the already existing feature points. In order to track feature points reliably, their pixel neighbourhood should by richly structured. As a measure of 'structuredness' of the neighborhood of a pixel *p*, one can define the structure matrix *G*:

$$G = \sum_{x \in W(p)} \nabla I(x) \cdot \nabla I(x)^T$$

Its eigen values λ1, λ2 (which are guaranteed to be ≥ 0 as the matrix is positive semi-definite) deliver useful information about the neighborhood region *W*. If *W* is completely homogenous, then λ1 = λ2 = 0. In contrast, λ1 > 0, λ2 = 0 indicates that *W* contains an edge and λ1 > 0, λ2 > 0 indicates a corner. The smaller eigen value λ = min(λ1, λ2) can now be used as a measure of the cornerness of *W*, where larger values means stronger corners.

The feature detection is now composed of the following steps:
1) Calculate structure matrix *G* and cornerness λ for each pixel in the image *I*.
2) Calculate the maximum cornerness λ_{max} occurring in the image.
3) Keep all pixels that have a cornerness λ larger than a certain percentage (5% - 10%) of λ_{max}.
4) Do a non-maxima suppression within the 3 x 3 pixel neighbourhood of the remaining points to keep only
5) the local maxima.
6) From the remaining points, add as many new points to the already existing points as needed, starting with the points with the highest cornerness values. To avoid points concentrated in some area of the image, newly added points must have a specific minimum distance (e.g. 5 or 10 pixels) to the already existing points as well as to other newly added points (*Minimum-Distance-Enforcement*).

Feature Point Tracking

In the tracking step, we want to calculate for each feature point *p* in image *I* its corresponding motion vector *v* so that its tracked position in image *J* is *p* + *v*.
As 'goodness' criterion of *v* we take the SSD error function

$$\varepsilon(v) = \sum_{x \in W(p)} (J(x+v) - I(x))^2$$

It measures the image intensity deviation between a neighbourhood of the feature point position in *I* and its potential position in *J* and should be zero in the ideal case. Setting the first derivative of ε(*v*) to zero and approximating *J(x + v)* by its first order Taylor expansion around *v* = 0 results in a better estimate v1. By repeating this multiple times, we obtain an iterative update scheme for *v* which is summarized below [17]:

1. Set initial motion vector $v_1 = (0,0)^T$
2. Spatial image gradient $\nabla I = \partial I / \partial(x, y)$
3. Calc. structure matrix $G = \sum_{x \in W(p)} \nabla I(x) \cdot \nabla I(x)^T$
4. for $k = 1$ to $maxIter$
    a) Image difference $\eta(x) = I(x) - J(x + v^k)$
    b) Calc. mismatch vector $b = \sum_{x \in W(p)} \eta(x) \cdot \nabla I(x)$
    c) Calc. updated motion $v_{k+1} = v_k + G^{-1} b$
    d) if $\| v_{k+1} - v_k \| < eps$ then stop (converged)
5. Report final motion vector v

Table 1: Pseudo-code of the calculation of the motion vector *v* for a given feature point *p*. W(*p*) is a window centered at *p*. Typically the window size is set to 5 x 5 pixel, *maxIter* to 10 and *eps* to 0.03 pixel.

4. Summary of Literature Review

Paper Title	Publications	Authors	Techniques
Visual Object Tracking Based on Local Steering Kernels and Color Histograms[1]	*IEEE Transactions on Circuits & Systems for Video Technology VOL:25 NO:3 YEAR 2013*	Olga Zoidi, Anastasios Tefas, Member, IEEE, and Ioannis Pitas, Fellow	Successfully tracks target object under partial occlusion with slow transformation. Employs appearance based representation method, Color Histogram, LSK tracking technique.
Visual tracking and recognition using appearance: Adaptive models in particle filters[2].	IEEE Trans. Image Process., vol. 13, no. 11, pp. 1434–1456, Nov. 2004.	S. Zhou, R. Chellappa, B. Moghaddam	Appearance based target object representation method is used. Particle Filter Tracking technique is implemented. Single object tracking without occlusion.
Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation[3].	IEEE Signal Processing Magazine SEPT. 2012	Ramsey Faragher	The Kalman filter include smoothing noisy data and providing estimates of parameters of interest.
Tracking video objects in cluttered background[4].	IEEE Trans. Circuits Syst. Video Technol., Apr. 2005.	A. Cavallaro, O. Steiger, and T. Ebrahimi	Hybrid video object tracking, Data association, low level descriptors, object segmentation
Tracking of Multiple Objects under Partial Occlusion[5].	SPIE-Research Gate 2009.	C Bing Han, Christopher Paulson, Taoran Lu, Dapeng Wu, Jian Li.	Template matching, silhouette tracking and trajectory estimation method. Tracks moving object with high accuracy under partial occlusion.
Real-time Visual Tracking of Aircrafts[6].	Digital Image Computing: Techniques and Applications	Ajmal S. Mian.	KLT tracker. Successfully tracks aircraft in motion.

5. Conclusion

This paper presents a literature review of the target object representation methods being grouped into five categories. It also gives a detailed description of the different tracking techniques. This article also reflects that visual object tracking is subjected to varying real life situations like illumination changes, affine transformations and speedy motion of the target object. It also signifies that the existing techniques track object under partial occlusion. Tracking speedy motion of multiple objects simultaneously and under full occlusion is the area that encourages new research.

6. Acknowledgement

I express my sincere gratitude towards my guide Prof. (Mrs.) M. K. Kulkarni for her valuable guidance. I am thankful to Prof. (Dr.) B. D. Phulpagar for having provided me important suggestions and guided me whenever I approached for help. I am also thankful to Prof. (Ms.) D.V. Gore and Prof. (Dr.) S.A. Itkar for their encouragement and support.

References

[1] Visual Object Tracking Based on Local Steering Kernels and Color Histograms Olga Zoidi, Anastasios Tefas, *Member, IEEE, and Ioannis Pitas, Fellow, IEEE* TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY VOL:25 NO:3 YEAR 2013.

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

[3] Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation, Ramsey Faragher , IEEE Signal Processing Magazine SEPT. 2012.

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

[5] Tracking of Multiple Objects under Partial Occlusion.C Bing Han, Christopher Paulson, Taoran Lu, Dapeng Wu, Jian Li, SPIE-Research Gate 2009.

[6] Real-time Visual Tracking of Aircrafts. Ajmal S. Mian. Digital Image Computing: Techniques and Applications 978-0-7695-3456-/08 $25.00 © 2008 IEEE DOI 10.1109/DICTA.2008.33.

[7] 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.

[8] 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.

[9] 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.

[10] Y. Wang and O. Lee, "Active mesh—a feature seeking and tracking image sequence representation scheme," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 610–624, Sep. 1994.

[11] 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.

[12] 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.

[13] J. Wang, G. Bebis, and R. Miller, "Robust video-based surveillance by integrating target detection with tracking," in *Proc. Conf. CVPRW OTCBVS*, Jun. 2006.

[14] 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.

[15] G. Welch and G. Bishop, "An introduction to the Kalman filter," Univ.North Carolina, Chapel Hill, NC, Tech. Rep. TR95041, 2000.

[16] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 4, Oct. 2004.

[17] Realtime KLT Feature Point Tracking for High Definition Video. Hannes Fassold1 Jakub Rosner2Peter Schallauer1 Werner Bailer1. FAH-2009 GRAVISMA.

[18] Tracking of Multiple Objects under Partial Occlusion.C Bing Han, Christopher Paulson, Taoran Lu, Dapeng Wu, Jian Li, SPIE-Research Gate 2009.

[19] E. Rivlin A. Adam and I. Shimshoni. "Robust fragments-based tracking using the integral histogram,". in Proc. IEEE Conf. CVPR, 2006.

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

[21] T. Okuma. "A natural feature based 3d object tracking method for wearable augmented reality,". in Proc. AMC, 2004.

[22] D. Wu N. Stergiou H. Luo, S. Ci 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, 2010