Dynamic Hand Localization and Tracking using SURF and Kalman Algorithm

Richa Golash
EC Department
Akshaya Institute of Technology
Tumkur, India

Yogendra K. Jain
CS Department
Samrat Ashok Technological Institute
Vidisha, India

ABSTRACT
Moving Object detection and tracking its path now-a-days has become very interesting field of research. But current state of art is still facing many challenges due to natural factors of the object and environmental factors, which plays significant role in determining efficiency of visual tracking system. This paper is a part of work in the field of Dynamic Hand Recognition. It highlights the important challenges faced in locating and tracking non rigid object, hand and proposes a system to locate hand using SURF algorithm and tracking its path using Kalman filter. The proposed work will be helpful in improving the efficiency in Human Computer Interaction using hand.

Keywords
Box Filter, Detection, Kalman Filter, SURF, Tracking

1. INTRODUCTION
Recent trend is, to have natural interaction of human with machine devices. This will help human society in many ways. Hand gesture and Speech are the best method of communication, out of which hand gesture has turnout to be the most promising field and can be accepted internationally. Current state of art is that all visual tracking problem conventionally, concentrate on prediction tasks to estimate the motion of a target object. Such approaches do not take the appearance variability into consideration and thus perform well only over short period of time. Moreover analyzing non rigid object which changes geometry from frame to frame is difficult task, therefore lots of work is going in this field. This paper proposes a new methodology which combines the characteristics of SURF algorithm and Kalman Filter, to design a new approach to detect moving hand and determine its trajectory of motion.

If we analyze the movement of human hand it is found that hand moves nonlinearly and changes its velocity and direction irregularly, so it is difficult to represent hand movement using a unified motion pattern [1], [2]. Therefore hand tracking systems are facing number of challenges, such as

1. Hand motion is not uniform, it is complex in nature, which depends on person to person. Hence area of target covered in each frame varies, this increases the complexity of system.

2. Using skin Color as only feature, increase probability of detection of similar unrelated object. Therefore presence of clutter or unrelated objects similar to the target - may play a major role in tracker failure.

3. Tracking applications are also very much affected by variations in either local or global illumination conditions.

This Paper deals with a unique approach of using SURF to detect and locate moving hand to avoid target loss due to illumination variation or change in geometry and then track its trajectory using Kalman Filter. This approach will also increase system robustness against rotation and illumination variation and will decrease the rate of loss even when number of frames are increased. The proposed methodology has following four important steps to track motion of hand:

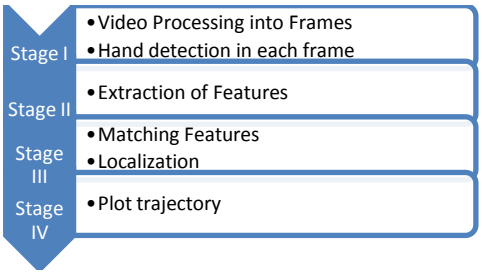


Fig1: Flow Chart of Dynamic Hand Motion Detection

2. COMPARISON OF VARIOUS TECHNIQUES USED FOR HAND DETECTION AND TRACKING:
Feature Detection and its matching is a crucial factor in determining the success rate of any tracking system. This is done by selecting robust features which are invariant to various transformations, illumination change, effect of noise and helpful to detect hand even in presence of cluttered background [3], [4]. There are various approaches to determine the features of moving object. Basically features are of two type Global and Local features. Global features have the ability to generalize an entire object with a single vector. Therefore they can be directly used in standard classification for example color, edge or texture. Local features on the other hand are computed at multiple points in the image and are consequently more robust to occlusion and clutter for example SIFT , SURF , GLOH etc. [5],[6].

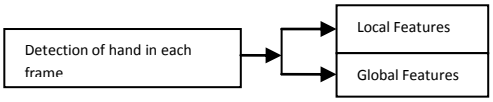


Fig 2: Types of Features

Most of the researches have used color as main feature because it is efficient where the light conditions are not uniform and are changing during the tracking but main drawback is that, the chance of selection of similar color, unrelated object also increases. Edge based detection though allows fast tracking but it is not very helpful in

case of hand motion analysis as hand is a non-rigid object also shape does not remain constant from frame to frame. Recently to increase robustness texture is also being used to detect object in multi-cue tracker system. The main drawback of using this feature is that it tend to make design research application field computationally intensive, and use classification methods that require a time expensive off-line learning phase. For these reasons this approach is not used consistently in tracking process [6], [7].

Many strategies are present to select local feature points for eg. Harris et.al. proposed a method to extract rotational and translational-invariant features by combining corner and edge detectors based on local autocorrelation functions [7]. Shi et. al . proposed a method to threshold the minimum eigen values of image gradient matrices at candidate feature points and use them as the appropriate feature points for tracking [8]. These methods generate rotational and translational invariant point features however these are variants to affine or projective transformations. Lowe et al. proposed a Scale-Invariant Feature Transform (SIFT) that is invariant to rotations, translations, scaling, affine transformations, and partially invariant to illuminations [9], [12]. Bao et.al. proposed a technique using Speed Up Robust Features (SURF), which is having similar performance as SIFT [11], but it is computationally fast due to the concept of integral image.

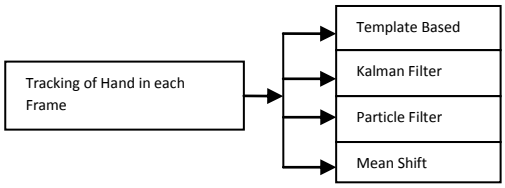


Fig 3: Types of Tracking Algorithm

The second Main Phase is in the system design is tracking it is defined as the frame-to-frame correspondence of the segmented hand regions or features towards understanding the observed hand movements. Many object tracking methods have been proposed and developed in the literature so far, an overview of visual tracking method was given by Yilmaz et. al. [13] and Sankaranarayanan et.al. [14]. Real-time performance is achieved by pre-computing "motion templates" which are the product of the spatial derivatives of the reference image to be tracked and a set of motion fields. Some approaches detect hands as image blobs, Wang et al. in each frame temporally corresponds blobs that occur in proximate locations across frames. Blob-based approaches are able to retain tracking of hands even when there are great variations from frame to frame [15]. The optimal estimation framework provided by the Kalman filter (Kalman 1960) has been widely employed in Wang et al. for turning observations (feature detection) into estimations extracted trajectory [15]. Shan et al. proposed to embed the mean shift in particle filters to track human hands, where particles with large weights from the mean shift are combined in the observation model, thereby reducing the degeneracy and requiring fewer particles than the conventional particle filter [3]. Several researchers including Huang et al. have combine other simple tracking methods with CamShift to improve the tracking performance, the approach proposed combination of CamShift algorithm with Kalman filter [16].

The paper is organized as follows. In section 3, Details of SURF algorithm along with its advantages is presented. In section 4 Kalman Filter is discussed, in section 5 Flow chart of the proposed methodology is discussed. Section 6 has Results and section 7 conclusion and future plans are discussed.

3. SURF ALGORITHM

SURF (Speeded-Up Robust Features), was originally presented by Herbert Bay et al. It is based on the idea to find corresponding points between two images of the same scene or object. The process to find corresponding points is accomplished in three steps: First is detection of 'interest points', which have high repeatability. Generally these are distinctive locations in image such as corners, blobs or T-junctions etc. Second Step is finding 'descriptors' for the interest points, which are also called as feature vectors. This is the most important step for locating the object in the next frame, hence they should be robust to noise and invariant to geometric and photometric deformation. Last and Final step is 'matching' of descriptor vector with the descriptor vector in the next frame based on distance formula for eg. Mahalanobis or Euclidean distance. The distinguish characteristic of SURF is less dimensions with fast interest point matching [10].

1st Step is Interest point detection: It is based on basic Hessian matrix and the concept of integral image because of its good performance and accuracy, it reduces computation time The integral image $I_{\sum}(x)$ at a location $x = (x; y)_T$ represents the sum of all pixels in the input image I within a rectangular region formed by the origin and x

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

Given a point x = (x; y) in an image I, the Hessian matrix H(x; σ) in x at scale σ is defined as follows

$$H(x;\sigma) = \begin{bmatrix} L_{xx}(x;\sigma) & L_{yx}(x;\sigma) \\ L_{xy}(x;\sigma) & L_{yy}(x;\sigma) \end{bmatrix}$$

Where $L_{xx}(x; \sigma)$ is the convolution of the Gaussian second order derivative $\partial_2 g(x)/ \partial x^2$ with the image I in point x, and similarly for $L_{xy}(x; \sigma)$ and $L_{yy}(x; \sigma)$. It detect blob-like structures at locations where the determinant is maximum. In order to localize interest points in the image and over scales, non-maximum suppression in a 3 x3 x 3 neighborhood is applied.

2nd step is description of Interest points: Basically it describes the distribution of the intensity content within the interest point neighborhood. For the extraction of the descriptor, first the localization of interest points is performed using scale space interpolation, then for each interest point, Haar wavelet response in both x and y direction within a circular neighborhood with 6s radius of interest points are calculated, where s stands for the scale of the level where this key point is found. Then from a sliding window of size $\pi/3$, the orientation with the largest sum of wavelet response is picked as the dominant orientation. Now a square region is constructed centered around the interest, to preserve the spatial information, the region is split up regularly into smaller 4x4 square sub-regions. For each sub-region, Haar wavelet responses at 5x5 regularly spaced sample points are computed. To increase the robustness towards geometric deformations and localization errors, the responses dx and dy are first

weighted with a Gaussian (σ = 3:3s) centered at the interest point.

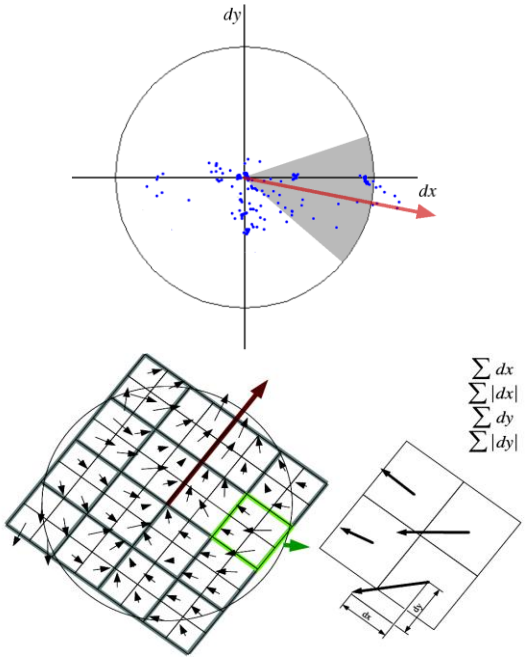


Fig 4: To build the descriptor, an oriented quadratic grid with 4x4 square sub-regions is laid over the interest point [10]

For each square, the wavelet responses are computed. These are the sums dx, |dx|, dy, and |dy|, computed relatively to the orientation of the grid. Hence, each sub-region has a four-dimensional descriptor vector v for its underlying intensity structure v = (∑dx; ∑dy; ∑|dx|; ∑|dy|). Concatenating this for all 4x4 sub-regions, this results in a descriptor vector of length 64 [10]. The wavelet responses are invariant to a bias in illumination (offset). Invariance to contrast (a scale factor) is achieved by turning the descriptor into a unit vector.

Third Stage is Matching stage. For fast indexing, the sign of the Laplacian (i.e. the trace of the Hessian matrix) for the underlying interest point is used. This sign help us in distinguishing bright blob on dark background or vice versa. Since the sign is already computed during detection phase, hence matching becomes very fast in case of SURF.

3.1 Advantage of Using Surf in Feature Detection

Many descriptors such as SIFT [9], SURF [10], and GLOH have become a core local features. In the proposed method SURF is used for matching features and locating hand in current frame because of following reasons.

1. It has in-plane rotation, scale invariant features and certain level of tolerance for view point, which makes it desirable for hand posture recognition.
2. In Surf, Key points or interest points are detected using Box filter, which is computed very quickly using Integral Image. This concept makes the calculation time of features independent of image size. This is important in our approach, as SURF can be applied to any frame size use big filter sizes.

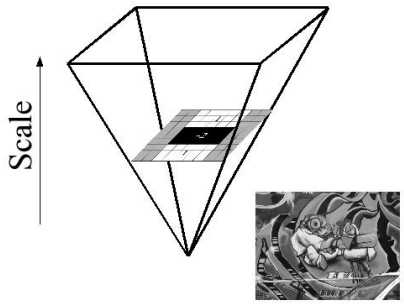


Figure 5. Advantage of using integral Image [10]

3. SURF Features are invariant to illumination change because in 2nd step of descriptor assignment for interest point's wavelet responses are calculated for each 4x4 sub region which are invariant to a bias in illumination (offset).
4. SURF handles blurring conditions better. It is less sensitive to noise because it integrates gradient information within a sub patch. Hence indexing and matching is very fast in case of SURF.

4. TRACKING USING KALMAN FILTER

In this Paper Kalman Filter is used to track Hand movement. Kalman Filter is a recursive linear filter. It optimally predict and estimate the state and also avoids noise filtering. The process involved in Kalman filtering is described in flow chart in Fig. The discrete Kalman filter is a set of mathematical equations that provides an efficient recursive mean for estimation of the state of a discrete process, in a way that minimizes the mean of the squared error. The Kalman filter estimates the process state at some time and then obtains feedback in the form of measurements. The equations for the discrete Kalman filter fall into time update equations and measurement update equations [17].

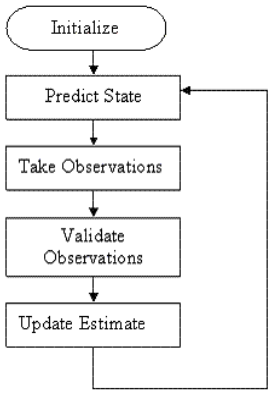


Fig 6: Flow of Kalman process of tracking

The time update equations are responsible for projecting forward (in time) the current state ($\hat{x}_k$) and error covariance estimates (P_{k-1}) to obtain the a priori estimates($\hat{x}_k^-$and $P_{\bar{k}}$) for the next time step. The measurement update equations are responsible for the feedback i.e. for incorporating a new measurement into the a priori estimates to obtain improved a posteriori estimates ($\hat{x}_k^-$ andP_k).The above process is expressed in set of mathematical linear equation [17] :

1) Time Update Equation

$$\hat{x}_k^- = A\hat{x}_{k-1}$$

$$P_{\bar{k}}=AP_{k-1}A^T + Q$$

2) Measurement Update Equation

$$K_k = P_{\bar{k}}H^T(HP_{\bar{k}}H^T + R)^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

$$P_k = (I - K_kH)P_{\bar{k}}$$

Where K_k is called Kalman gain. When measurement error covariance R approaches zero, the actual measurement z_k is trusted more and more, while the predicted measurement $H\hat{x}_k^-$ is trusted less and less [17].

5. FLOW CHART

In the Proposed Algorithm SURF is used for Feature Detection and Kalman Filter is used for tracking. SURF is a local feature descriptor used to extract scale-invariant features. It first identifies key locations in scale space by looking for candidate locations which are maxima or minima of a difference-of-Gaussian function. It extracts a feature vector on each point that models the local image region in a scale-space coordinate frame. The extracted features are invariance to local variations such as affine or 3D projections. These key points are tracked using Kalman Filter.

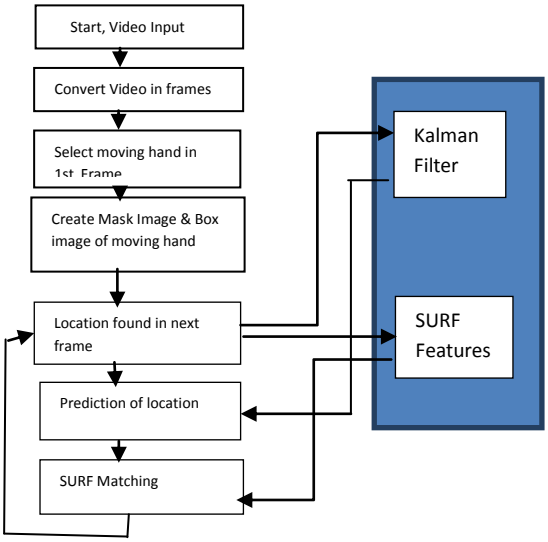


Fig 7: Flow chart of Proposed Method

6. EXPERIMENT AND RESULT

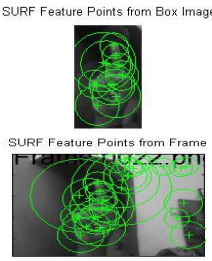
Step 1: Converting Input Video into Frames: In this Step Video is first converted into Frames

Fig 8: Video File Converted into Frames.

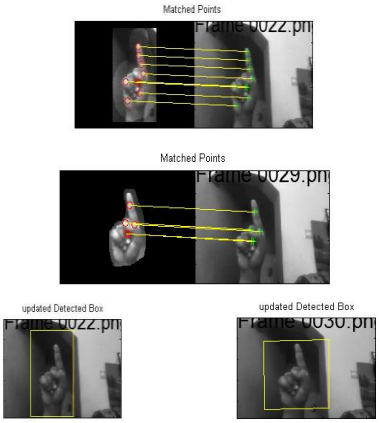
Step 2: Mask of Moving hand is created from first frame. Which is used to mask hand in each frame. In this step box image is extracted

Fig 9: Mask Image and Box Image of Moving Hand

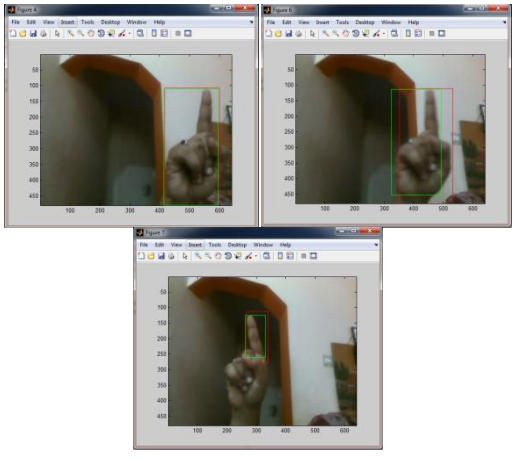
Step 3: Surf Points are extracted from box image and Current frame.



Step 4: Matching of SURF Key point and determining new location of box image i.e. moving hand. This is known as updated box image. The updated location of box image is now used to match with next frame.



Step 5: Implementing Kalman Filter determine the trajectory of motion of hand.



Step 6: The result show tracking using SURF Feature alone and tracking with SURF and then with Kalman algorithm together.

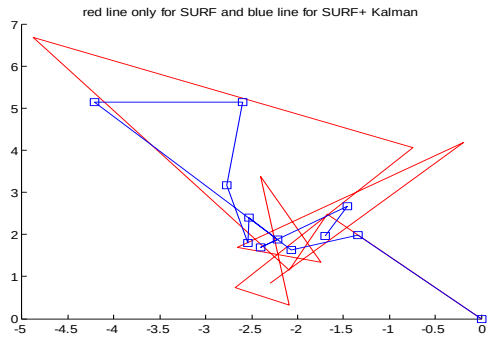


Fig 10: Plotting Trajectory using only SURF –Red line and with SURF and Kalman Algorithm- Blue line.

7. CONCLUSION

In this paper SURF is used in conjunction with Kalman Filter. The result shows that due to many advantages of SURF features and two stage process of Kalman Filter i.e. Prediction and Correction, as discussed in paper is improved. The performance of detecting non- rigid object, moving hand is improved as compared to other feature detection techniques, which can be used in several applications. The combination of these two techniques will help us to design user adaptive human computer interface using hand.

8. REFRENCES

[1] Fang Y., Wang K., Cheng J., Lu H., "A real-time hand gesture recognition method", IEEE international conference on Multimedia and Expo, pp. 995–998,2007.

[2] Rautaray S. S., Agrawal A., "Real time hand gesture recognition System for dynamic applications", International Journal of Ubi.Comp, Vol. 3, No. 1, Jan. 2012.

[3] Shan C., Tan Tieniu, Yucheng W., "Real-time hand tracking using a mean shift embedded particle filter", Pattern Recognition, Vol. 40, No. 7, pp. 1958-1970, 2007..

[4] Chaudhary A., Raheja J. L., Das K., Raheja S., "Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey", Int. J.Comput Sci. Eng Survey (IJCSES), Vol. 2, No. 1, pp. 122–133, 2011.

[5] Rautaray S. S., Agrawal A., "A Novel Human Computer Interface Based On Hand Gesture Recognition Using Computer Vision Techniques", In Proceedings of ACM IITM'10, pp.292-296, 2010.

[6] Murthy G. R. S., Jadon R. S., "A Review of Vision Based Hand Gestures Recognition", International Journal of Information Technology and Knowledge Management, Vol. 2, No. 2, pp 405-410, 2009.

[7] Harris C., Stephens M., "A Combined Corner and Edge Detector", Proc. of 4th Alvey Vision Conf., Manchester, pp.147-151, 1998.

[8] Shi J., Tomasi C., "Good features to track", Proc. of IEEE International conf. on Computer vision and Pattern Recognition CVPR, pp. 593-600, 1994.

[9] Lowe D. G., "Distinctive image features from scale-invariant key points", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91–110, 2004.

[10] Bay H., Tuytelaars T., Gool L. V., "SURF: Speeded Up Robust Features", Proc .European conf. ECCV, pp.404-417, 2006.

[11] Bao J., Song A., Guo Y., Tang H., "Dynamic hand gesture recognition based on SURF tracking", International conference on electric information and control engineering (ICEICE), pp 338–341, 2011.

[12] Tu Q., Xu Y., Zhou M., "Robust vehicle tracking based on Scale Invariant Feature Transform", Proc. IEEE Int. Conf. on Information and Automation, 2008 ICIA, pp. 86-90, 2008.

[13] Yilmaz A., Javed O., Shah M., "Object tracking: A survey", ACM Computing Surveys, Vol. 38, no. 4, 2006.

[14] Sankaranarayanan A. C., Veeraraghavan A., Chellappa R., "Object Detection, Tracking and Recognition for Multiple Smart Cameras", Proc. of the IEEE, Vol. 96, No.10, pp. 1606-1624, 2008.

[15] Wang T., Backhouse A. G., Gu I. Y. H., "Online subspace learning on Grassmann manifold for moving object tracking in video", Proc. IEEE int'l conf. on Acoustics, Speech and Signal Processing, ICASSP'2008, pp. 969-972, 2008.

[16] Huang S., Hong J., "Moving object tracking system based on camshift and Kalman filter", International conference on CECNet, pp 1423–1426, 2011

[17] Nathan Funk, "A Study of the Kalman Filter applied to Visual Tracking", University of Alberta, Project for CMPUT 652,December 7, 2003