Multiple object tracking by employing shaped-based features and Kalman filter

Felix M. Philip*
Department of Electronics and Communication Engineering,
Hindustan University, Chennai, India
Email: felixmphilip8614@gmail.com
*Corresponding author

Rajeswari Mukesh
Department of Computer Science and Engineering,
Hindustan University, Chennai, India
Email: rajeswarim@hindustanuniv.ac.in

Abstract: There has been fast development happening in the multimedia and the related technologies, particularly associated with visual tracking and search operations. Moving target detection has been comprehensively engaged in various arenas but has the disadvantage that the scheme is frequently complex and also that tracking is affected numerous external factors. In this article, multiple objects recognition and tracking is projected so as to progress the method and make it more robust and general with assistance of shape-based features and Kalman filter. Primarily, the input video is rehabilitated to frames and then manually segmented for object segmentation. Consequently, the objects are tracked with the help of Kalman filtering. The method is assessed under standard evaluation metrics of error value and the score value. The technique achieved maximum score values of 95% and minimum error value of 25%. The results validate the effectiveness of the technique.

Keywords: moving object detection; tracking; shape-based features; Kalman filtering and segmentation.

Reference to this paper should be made as follows: Philip, F.M. and Mukesh, R. (2018) ‘Multiple object tracking by employing shaped-based features and Kalman filter’, Int. J. Business Intelligence and Data Mining, Vol. 13, Nos. 1/2/3, pp.331–346.

Biographical notes: Felix M. Philip obtained his Bachelor’s degree in Electronics and Communication Engineering from Mahatma Gandhi University, Kerala, India. Then he obtained his Master’s degree in VLSI Design from Anna University, Tamil Nadu, India and he is a Research Scholar in Electronics and Communication Engineering (‘image processing’), in Hindustan University, Chennai, India. He is currently an Assistant Professor at the Faculty of Electronics and Communication Engineering in, Jawaharlal College of Engineering and Technology.

Rajeswari Mukesh has received her PhD degree in Computer Science and Engineering from Jawaharlal Nehru University, Hyderabad. At present, she is a Professor and Head of Computer Science and Engineering Department at Hindustan University. She is guiding eight PhD candidates. She published more than 15 international journals and attended more than 15 international and
1 Introduction

Object tracking (Yilmaz et al., 2006) signifies a method that tracks an object into a number of frames of a video and trace its comparative motion in relation to other objects in the neighbourhood. It is, in fact, a very important performance in the domain of computer vision and has capitalised its zooming attention on the recognition of objects. The basic object of vision scheme is to keep an eye on the highways, buildings and over protracted period of time with the purpose of tracking the vehicles and humans so as to recognise anomalous conduct. The object detection is a method that identifies the moving objects in the vicinity. It is indispensable for the every tracking method to have knowledge of the objects in the neighbourhood (Hu et al., 2004). The video surveillance schemes have been doing their sophisticated rounds in security schemes and potential threats are recognised well-ahead. It has appeared as a fruitful and economical manner of supplying safety to people, buildings and highways. In addition to this the multiple object tracking has become a challenging research subject in computer vision. It is commended with the mission of addressing hassles faced in the single object tracking, like the changing appearances, non-rigid motion, dynamic illumination and occlusion, in addition to the contests related to the multiple objects tracking like the inter-object occlusion and multi-object confusion (Han et al., 2004).

The encounters in tracking the objects may instigate due to the fast object motion, (Anusha and Julie, 2014) varying appearance patterns of both the object and the scene, non-rigid object edifices, object-to-object and object-to-scene occlusions, and the camera motion. Usually, bottom-up and top-down methods are two types of methodologies for the visual tracking issue (Hu et al., 2009). Also, because of the complex nature of human body, tracking human in video systems is a very difficult mission (Wu and Huang, 2004) and comprises a number of hard problems like occlusion, self-occlusion, cluttered background, and high-dimensional motion parameters (WenJuan et al., 2010). Occlusion reasoning for object tracking is one among the most stimulating issues in visual surveillance (Ning et al., 2002). Furthermore, The SIFT that characterises an appropriate feature is an advanced method for recognising and extracting local feature descriptors that are logically invariant to adjustments in illumination, scaling, rotation, image noise and minor modification in the perspective (Anusha and Julie, 2014). In object tracking, assured features establish the colour, edge, position, moment and so on. The moment features provide assured strength to occlusion and background clutter. In the colour moments and wavelet moments, in object recognition mechanism, typically a group of numerical features are abstracted from an image. The Kalman filter tracking mechanism has a consistent forecast on the target movement, feature matching can be achieved in the moderately miniature arena (Lou et al., 2002). Thus the Kalman filter shows an extremely important portion in augmenting the processing tempo and efficacy of the tracking scheme.
In this regard, a feast of tracking algorithms was hurled including the method templates and local features (Raja et al., 1998) Kalman filters (Lou et al., 2002; Koller et al., 2011) and contours (Meier and Ade, 1999; Giachetti). The mean-shift algorithm was primarily engaged as an operative tracking method in Comaniciu et al. (2002). Henceforth, the identification and tracking of players in broadcast sports video sustained to be intricate. Further, a lot of endeavours have been made in the direction of the player tracking in sports video like (Needham and Boyle, 2001; Utsumi et al., 2002; Agbinya and Dees, 1999; Figueroa et al., 2004; Iwase and Saito, 2004). Firm methods trace the players by means of a single camera (Needham and Boyle, 2001; Utsumi et al., 2002; Agbinya and Dees, 1999) though some others (Figueroa et al., 2004; Iwase and Saito, 2004) have endeavoured to track players from the video furnished by multiple cameras. Furthermore, a number of approaches have intended for the motion segmentation (Zhang and Lu, 2001) like the image differencing (Rosin, 1998), temporal investigation of grey-levels on the basis of probabilistic models (Giaccone and Jones, 1998), and robust motion assessment (Sawhney and Ayer, 1996), or misalignment investigation on the basis of the normal flow (Irani et al., 1996). In Park and Aggarwal (2002), body parts are separated and tracked, by means of a body model to help in the tackling of uncertainties and tracking collapses. In addition, the crusade of the object is evaluated for automation of the pan-tilt mechanism (Murray and Basu, 1994). On the basis of the movement of the object the pan-tilt mechanism is achieved to position the object in the camera’s view. The frame differencing algorithm (Jain and Nagel, 1979; Haritaoglu et al., 2004) is engaged for the purpose that equips as output the position of the moving object in the image. The significant, recovering noises and unexpected variation is imperative for robust object tracking, the main control points of shape data tracking performance (Li et al., 2003).

The rest of the paper prearranged as follows: the recent research works is investigated in Section 2; the projected work is fleetingly elucidated in Section 3; in the 4th Section, the experimental results together with the comparison result are portrayed and the Section 5 signifies the summary of the paper.

2 Related works

Literature boons lot of works for visual tracking and in this segment, some of work associated with it is revised. He et al. (2015) valiantly hurled a universal optimisation method in the network flow model for multiple object tracking. At this time, a variable number of objects were routinely traced in the scene of a monocular and un-calibrated camera. They intended a global optimisation method in network flow model for multiple object tracking. This method protracted the recent work that expressed the tracking-by-detection into a maximum-a posteriori (MAP) data association issue. They reinvented the observation prospects and the kinship within observations to tackle permanent occlusions. Additionally, an improved greedy algorithm was industrialised to solve the min-cost flow, decreasing the amount of ID switches ostensibly. Furthermore, a linear hypothesis method was imagined to fill up the gaps in the trajectories. The test outcomes exemplified that the method was operative and effectual, and outpaces the high-tech approaches on numerous standard datasets.

Furthermore, Xi et al. (2015) was instrumental in suggesting the multiple object tracking using A* association algorithm. The multiple object tracking was intended as an integer programming dilemma of flow network. In this outline, the integer assumption
was abridged to a standard linear programming problem and henceforth the worldwide optimal solution was swiftly attained by the A* algorithm with dynamic weights. The inventive was fruitful in deterring the hassles of integer programming and more conspicuously, in relation to other modern parallel methods, knowledgeable lesser worst-case difficulty and was able to competently track the precision and strength in complex scenarios. The replication outcomes went a long way in inaugurating that the innovative method was cost-effective and was able to meet the simultaneous needs.

Correspondingly, Kim et al. (2012) expansively conceptualised the real-time interactive modelling and scalable multiple object tracking for AR. The epoch-making method encompassed both the scalable tracking and interactive modelling. Their engrossment was two-pronged: at the outset, they exemplified the technique to scale with the number of objects engaging the key frames. This was performed by integrating modern approaches for the image reclamation and online structure from motion that was implemented autonomously. Therefore, the tracking of 50 objects in 3D was performed within 6–35 ms per frame, in spite of adversative backdrops for tracking. Supplementary, they recommended a method to permit the user supplement novel objects very fast, buy just picking an image in a 2D region lying on the object. A 3D primitive was then built-in to the features within this area, and modified to produce the object 3D model. They demonstrated the modelling of polygonal and circular-based objects. The relative procedure was accomplished within the flick of a minute.

Tawab et al. (2013) assimilated amazing applause for encouraging the efficient multi-feature PSO for fast gray level object-tracking. The object model was reliant on the gray level strength. The eye-catching method amalgamated the effects of numerous object cases such as the zooming, scaling, rotating, and so on into a separate cost function. The method was not on the basis of the object type and shape and it was engaged for a host of object tracking implementations. More than 30 video series with 20,000 plus frames were affianced to evaluate the designed PSO-based object tracking method and compare it to traditional object tracking method and the earlier printed PSO-based tracking methods. The charismatic consequences vouchsafe the superlative efficacy and strength of our innovative method vis-a-vis the parallel examined methods.

Furthermore, Yin et al. (2011) methodically clarified the hierarchical Kalman-particle filter with adaptation to motion variation for the purpose of object tracking. At this time, they envisioned two methods in the tracking method. At the outset, they examined the object motion in a coarse-to-fine way and engaged the hierarchical method to assess it, in that the Kalman filter evaluated the global linear motion and the particle filter assessed the local nonlinear motion. Additionally, they intended a further physically important proposal distribution of the particle filter duly considered the character of motion. Fascinating test outcomes illustrated the efficacy of the tracking method in the real video sequences in that the target objects experience swift and unexpected motion. In addition, they were able to deliver quantitative contrasts within the modern and the new tracking algorithms.

Moreover, Mazinan (2014) astonishingly hurled the automated probabilistic assessment outline in concurrent objects tracking. The innovative method functioned in relation with a novel self-correcting particle filter to trace a multitude of moving objects. The idea was pertinent to trace a large majority of the real-time non-rigid objects, in view of the 3D image was endangered to scrutiny. As the captured frames were considered to be two dimensional information matrices, assured proper extracted features of the administered frames were organised to organise the third dimension. The general
Multiple object tracking by employing shaped-based features

appropriate features of moving objects that could not be implemented to the process of ensuing probability computation, had to be equipped a neural network to organise the third dimension. Later on, the probabilistic assessment of the modern self-correcting particle filter in each frame was remedied by means of the neural network consequences to assess each and every recognised object, aptly, in its available structure.

3 Proposed multiple object tracking by employing shaped-based features and Kalman filter

There has been lot of investigation fashionable in moving object detection. The overall difficulty of object recognition in static images is a intricate one as the object detection scheme is necessary to discriminate a specific class of objects from all others. Additional, a robust object detection scheme should be able to discriminate objects in uneven illumination, rotations and occlusions. The incidence of noise also adds to the issues. There are a diversity of algorithms to deal with object tracking, each having its own strengths and weaknesses.

In our last paper (Philip and Mukesh, 2015), object detection and tracking scheme making usage of back ground-based object segmentation and feature extraction was projected. Disadvantage of the paper was that only a single object was tracked which may not work in high object videos. Henceforth, in order to progress the method and make it more tough and over-all, in this paper multiple objects recognition and tracking is projected. The method employ shaped-based features and Kalman filter. The block diagram of the projected method is given in Figure 1.

3.1 Converting video to frames

The first stage convoluted in the procedure is to convert input video to frames. The conversion is performed with the help of the shot segmentation method. In most video implementations such as video retrieval, video watermarking and video summarisation video retrieval, shot segmentation forms the primary stage. In shot segmentation, video in consideration is firstly split into non-overlapping units where each unit is termed a shot. We can perceive important modifications in every shot that is aggressively subsidised by
the background noise and the objects in the videos. There are numerous video shot segmentation methods in the literature.

We have engaged wavelet transform-based shot segmentation. The number of frames measured in each shot is reliant on the distance measure and the transformation. The preliminary two frames are parted into set of blocks and DWT is engaged for each block of the frame. Wavelets can be well-defined as mathematical functions that have an average value of zero in a well-defined constrained interval. Understand, the input is the series of real numbers signified as \(y_1, \ldots, y_m\) then the DWT \(n\) of series of \(x\) real numbers can be signified as:

\[
\psi_{\alpha \beta}(t) = \frac{1}{\sqrt{a_{jk}}} \psi \left( \frac{t - \beta_{\alpha \beta}}{\alpha_{\alpha \beta}} \right)
\]

where \(\psi\) provides the mother wavelet, \(t\) is the time, \(\alpha\) provides the scaling parameter and \(\beta\) provides the shifting parameter correspondingly. Consequently, the distance measure \((d_{ij})\) is intended within the frames in consideration signified by \(f_i\) and \(f_j\):

\[
d_{ij} = \sqrt{\sum_{j=1}^{m} (f_i - f_j)^2}
\]

**Figure 2**  The schematic diagram of conversion from video to frames

Once the distance is calculated for the first two frames, analogous methodology is performed for the rest of successive frames. Lastly, the frames within a shot can be documented by minimum distance. Minimum distance would apply maximum similarity within the frames. Let the input video sequence be signified by \(V[j, k]\) and the shot segmented non-overlapping shots be signified by \(Z[j, k]\). The schematic diagram of conversion is provided in Figure 2.
3.2 Object detection phase

In this stage, the object is perceived and its dimensions are found out. The object detection is performed with help of manual marking and object dimensions include object centre, width and height. The flow diagram of the phase is specified in Figure 3.

**Figure 3** Flow diagram of the object detection phase (see online version for colours)

Primarily, first frame is measured and the objects in the frame are marked manually. In order to cross check that marked arena is of significance, the arena constraint is introduced. Area constraint necessitates the object to be of minimum permitted arena so that noise is circumvented. Let the minimum permitted area be characterised as $Th_{area}$, the arena of the object pixel $ob_i$ be signified as $Area_i$, then the criteria can be stated as:

$$\text{if} \ (Area_i < Th_{area}) \ , \ \text{then discard} \ Ob_i$$  \hspace{1cm} (4)

That is, clusters without a minimum arena are not measured as objects. After finding the object, the four corners are renowned and it is utilised to calculate the object centre. Let the objects perceived be characterised as $Ob = \{Ob_1, Ob_2, \ldots, Ob_m\}$. For an object let the extreme corner points be epitomised by $a_{x_i}, a_{y_i}, b_{x_i}, b_{y_i}, c_{x_i}, c_{y_i}, d_{x_i}$, and $d_{y_i}$. Then the centre point $(cen_{x}, cen_{y})$ is presumed to be as:

$$cen_{x} = \frac{a_{x} + b_{x} + c_{x} + d_{x}}{4} ; \ cen_{y} = \frac{a_{y} + b_{y} + c_{y} + d_{y}}{4}$$  \hspace{1cm} (5)
Consequently, the object height and width is also found out. Height is considered as the on the basis of difference within the top and bottom edge. Likewise, width is considered as the on the basis of the difference between the right and left edge. Tilt is also found out.

Suppose, the left and right edges are represented by \(x_L\) and \(x_R\), top and bottom edges are signified by \(x_T\) and \(x_B\). Width and height can be calculated as:

\[
\text{Width} = |x_R - x_L|, \quad \text{Height} = |x_T - x_B|
\]  

That is from the objects are detected signified by \(O_{bi}\); for \(0 < i \leq M\) and each object is defined by \(O_{bi} = \{cen_{i,j}, width_i, height_i\}\).

### 3.3 Object tracking PHA

In this stage, the objects are tracked by finding out successive positions of the objects in successive frames. That is, once the object in deliberation is perceived in first frame, then the same object is found out in the second frame. The procedure is continued in the succeeding frames to keep in track of the object. The procedure is performed for all objects in the frame. The flow diagram of the phase is specified in Figure 4.

**Figure 4** The flow diagram of the object tracking phase
Multiple object tracking by employing shaped-based features

The procedure is performed for single object at a time. Primarily, implement Kalman filtering to predict the new centroid. The condition is such that if both the old and new centroid positions are similar, then keep the object location as same.

Kalman filtering is an algorithm that utilises sequences of capacities detected over time, containing noise and other inaccuracies, and yields assessments of unknown variables that tend to be more precise than those on the basis of a single measurement alone. The Kalman filter is a recursive estimator. This revenues that only the assessed state from the previous time step and the current measurement are desirable to calculate the estimate for the current state. The Kalman filtering algorithm works in a two-step procedure. In the prediction stage, the Kalman filter yields assessments of the current state variables, together with their uncertainties. Once the consequence of the next measurement is perceived, these assessments are updated with the help of a weighted average, with more weight being provided to assessments with higher certainty. In our case, Kalman filter is engaged so as to envisage the subsequent centroid value.

If the two are not same, the dimensions of the cropped image are prolonged from all sides. That is the cropped image is stretched through right side $X_d(i)$, left side $X_l(i)$, bottom side $X_b(i)$ and top side $X_t(i)$ to form novel dimensions of $U_r(i)$, $U_l(i)$, $U_d(i)$ and $U_u(i)$. For this updated cropped image, extract the features that comprise of density, correlogram and histogram and wavelet moment.

Density is frequently well-defined in pixels per inch (PPI) or pixels per centimetre (PPCM) that as the name proposes provides the number of pixels in the cropped image per unit measure. Understand the number of pixels limited in the cropped image is $g$ and the centimetre covered is $h$, then PPCM density $PD$ is specified by:

$$PD = \frac{g}{h}$$

(7)

Correlogram provides the plot of auto-correlation of the cropped image values. It can be also being measured as the cross-correlation of image with itself. Autocorrelation can be well-defined by the formula (where the input image is epitomised by $f(x, y)$):

$$f(x, y) * f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') f(x + x', y + y') dx' dy'$$

(8)

where $f(x, y)$ the two-dimensional brightness is function that defines the image, and $f(x', y')$ are the dummy variables of integration.

Third feature extracted is histogram. Histogram is a graphical illustration of the distribution of pixel values. A histogram is a representation of tabulated frequencies, erected over discrete intervals, with an arena proportional to the frequency of the pixel in the cropped image.

Image moments establish a significant feature extraction technique that engenders high discriminative features. Wavelet moment hails from orthogonal moment family that usage an orthogonal wavelet function as kernel. These moments syndicate the compensations of the wavelet and moment investigates in order to concept amended moment descriptors. Wavelet moment for an image characterised by $f(x, y)$ having dimension $m \times m$ can be specified by:
\[ M_{ab}(r) = \sum_{x=1}^{w} \sum_{y=1}^{h} \psi_{ab}(r)e^{-j\varpi f(r, \varpi)} \]  

where \( r = \sqrt{x^2 + y^2} \), \( \varpi = \tan^{-1}(y/x) \) and \( \psi_{ab}(r) \) is the mother wavelet basis.

Henceforth from the expanded cropped image, we have found out the features. Let the extracted features of density, Correlogram, histogram and wavelet moments be characterised as \( PD_i, AC_i, H_i \) and \( M_i \) for many locations. Amongst all the features, take the higher value-based location of the density, histogram, wavelet moment and lower value-based location for the correlogram. If all the taken locations are same, then take that location and crop the image of the second frame consequently. That is, it can characterise as:

\[
\text{if } \left( \lambda(\text{max } PD_i) = \lambda(\text{min } AC_i) = \lambda(\text{max } H_i) = \lambda(\text{max } M_i) \right),
\]

Then take the location

where, \( \lambda \) provides the location, max provides the maximum amongst all values and min gives the minimum amongst all values.

The procedure is performed for consequent frames. In the ensuing step, third frame image is cropped with the help of the object location of second frame and similarly. Generalising the procedure for all frames, the image in the \( i + 1 \)st frame is cropped with the help of the object location of \( i \)th frame. Save all the cropped region and object locations to detect the tracked object. The procedure is performed for all objects detected.

4 Result and discussion

The proposed multiple object tracking procedure deliberated in the preceding section is applied. The performance of the multiple object tracking is tested under numerous images and here we with the help of the two kinds of videos. In Section 4.1, database engaged experimental set up and assessment metrics utilised are designated. Section 4.2 provides the performance investigation.

<table>
<thead>
<tr>
<th>Sample input video images (see online version for colours)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image_url" alt="Input video images" /></td>
</tr>
</tbody>
</table>
4.1 Database employed, experimental setup and evaluation metrics

The database utilised includes numerous videos and for assessment purpose, we have taken two car videos from website. These videos-based generated frames. The application is performed with the help of MATLAB which was simulated in a scheme having system specifications of 6 GB RAM and 3.2 GHz Intel i-7 processor. The assessment metrics engaged are error value and the score value. Error value has to be the minimum and score value has to be maximum for an operative method. Error value is attained as the difference of the original value and the attained value. The Table 1 first and second row displays the sample video images.

4.2 Performance analysis of proposed approach

The function of multi objective tracking is frequently measured in terms of error value and score value that are the most important performance parameters. The rudimentary notion of our work is to multiple objects tracking by engaging FCM-based segmentation and shaped-based features. The assessment is made with the assessment metrics of error value and score value. Comparison is also made with respect to the available method so as to have an extensive investigation. The Table 2 first and second row displays the input video image with tracking result.

Table 2  Original image with tracking result (see online version for colours)

<table>
<thead>
<tr>
<th>Input video images</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
</tbody>
</table>

4.3 Discussion

Figures 5 to 8 provide the performance investigation of the projected methods with available methods. The assessment is made up of error value and score value. The scheme is good means we attain the minimum error value and maximum score value. At this time, Figures 5 and 6 provides the car 1 analysis; Figures 7 and 8 gives the car 2 video investigation. When investigating the Figure 1, in our projected work we attain the minimum error value of 33% that is 50% for available work. From the Figure 6, we attain the maximum score value of 95% but available work attains only 15% of score value. In Figures 7 and 8 displays the error value of car video 2 and score value of car video 2. From the Figure 7, we acquire the minimum error value of 25% but in existing work attain the 48% error value. Correspondingly, we investigating Figure 8, we attain the maximum score value of 95% that is 80% for available methods. A progression of examinations was directed to ruling the subsequent precision of our method. We
evaluated the execution of our subsequent method by using constant video that is taken straightforwardly from the camera.

**Figure 5** Error value for car 1 video (see online version for colours)

**Figure 6** Score value for car 1 video (see online version for colours)
There are frequent approaches for characterising ideal, subordinate upon the criteria evaluated execution. It will be established that, under the suppositions to be made in the subsequent arena, the Kalman channel is ideal as for all intents and drives any paradigm that bodes well. One portion of this optimality is that the Kalman channel fuses all
information that can be specified to it. A Kalman molecule channel (KPF) and an unscented. In above figures available work as the object tracing scheme existing method (Philip and Mukesh, 2015) wavelet and histogram are utilised.

This projected inspecting scheme sufficiently controlled the procedure of particles towards locales with high probability, and consequently it meaningfully lessened the quantity of particles obligatory for following. Molecule channels can speak to multimodal appropriations in which as the Kalman channels, are limited to Gaussian probability circulations. From the inspection Kalman channel is best looked at to other channels in visual protest subsequent in numerous messed and dissimilar situations. From the figures, we can evidently comprehend that the projected method realised better score values accomplishing maximum values of 95% and minimum error value of 25%. We can also perceive smaller error value with respect to the available method. Low error value and high score value displays the efficacy of the projected method.

5 Conclusions

Multiple objects recognition and tracking is projected in this paper engaging shape-based features and Kalman filter. The method comprises of object detection phase and object tracking phase. In object detection phase, shot segmentation and manual segmentation is performed. In object tracking phase, Kalman filtering is engaged. The method is assessed under standard assessment metrics of error value and the score value. The method attained maximum score values of 95% and minimum error value of 25%. The results establish the efficiency of the method.

References


Multiple object tracking by employing shaped-based features


