A fast region-based active contour for non-rigid object tracking and its shape retrieval

Hiren Mewada¹, Jawad F. Al-Asad¹, Amit Patel², Jitendra Chaudhari², Keyur Mahant² and Alpesh Vala²

¹ Electrical Engineering, Prince Mohammad Bin Fahd University, Al Khobar, Kingdom of Saudi Arabia
² CHARUSAT Space Research & Technology Center, Charotar University of Science and Technology, Changa, Gujarat, India

ABSTRACT

Conventional tracking approaches track objects using a rectangle bounding box. Gait, gesture and many medical analyses require non-rigid shape extraction. A non-rigid object tracking is more difficult because it needs more accurate object shape and background separation in contrast to rigid bounding boxes. Active contour plays a vital role in the retrieval of image shape. However, the large computation time involved in contour tracing makes its use challenging in video processing. This paper proposes a new formation of the region-based active contour model (ACM) using a mean-shift tracker for video object tracking and its shape retrieval. The removal of re-initialization and fast deformation of the contour is proposed to retrieve the shape of the desired object. A contour model is further modified using a mean-shift tracker to track and retrieve shape simultaneously. The experimental results and their comparative analysis concludes that the proposed contour-based tracking succeed to track and retrieve the shape of the object with 71.86% accuracy. The contour-based mean-shift tracker resolves the scale-orientation selection problem in non-rigid object tracking, and resolves the weakness of the erroneous localization of the object in the frame by the tracker.

INTRODUCTION

The captured video contains moving objects and sounds representing a real-life scene. It gathers the sequence of images and plays them continuously. The human eye is sensitive to watch these sequences as a video if at least 24 sequences are played per unit second (Larson, 2020). In the tracking process, a trajectory of the desired object is estimated in the space domain as it moves from one frame to another frame. Thus it is the process of assigning a particular label in each frame. According to the application requirements, object-centric information like its shape, orientation, area, and recognition can be obtained after tracking. Visual object tracking (VOT) is a challenging task due to many issues like poor quality of the camera (i.e. low resolution, color discrimination, low frame rate), other environmental challenges like the small size of the object, clutter background, non-rigid shape, occlusion, appearance variation and lastly real-time tracking.
requirement. To target these challenges, various object appearance forms have been used in VOT. These tracking forms are illustrated in Li et al. (2013).

Major VOT literature implements a robust tracking algorithm for the aforementioned challenges using either rectangular bounding box or ellipse. A contour-based tracking referred to as non-rigid object tracking is more challenging because it needs more accurate object shape and background separation in contrast to rigid bounding boxes. Non-rigid object tracking and its analysis have broad applications in real-life, i.e. gait analysis and human recognition, automated object classification, detection of suspicious activities in an automated surveillance application, annotation of retrieval of a particular video from the dataset in video indexing application, traffic monitoring, tracking in medical imaging for clinical diagnosis, gesture detection and tracking for robotic applications and so on.

The non-rigid tracking algorithms generally rely on local pixel information (Duffner & Garcia, 2013), superpixel (Wang, Lu & Yang, 2017), or patch-based segmentation approach (Godec, Roth & Bischof, 2013). In local pixel information, object’s pixels are localized and discriminated from the background using a Hough voting scheme and probabilistic segmentation algorithm. In superpixel, a confidence map is obtained from using a discriminative algorithm to segment and track the object. Whereas, in patch-based segmentation, a Hough forest voting is used to classify the patches of the desired object from the frame. A radon transform features from the images were extracted to identify non-rigid patterns from the images in Santosh, Lamiro & Wendling (2011, 2013). They used histogram at various projection angle in the Radon domain. To avoid the problem of compressed representation due to a single vector formation of the histogram, the authors proposed a dynamic time wrapping method. They found that this method is more robust to is deformed shape due to degradation, distortion and occlusion. These algorithms use handcrafted features, thus limiting the accuracy for the complex object. Besides, there is no correlation between the tracking and detection process.

Active contour has played a significant role in non-rigid image segmentation and it has been widely used in medical image segmentation (Beaupré, Bilodeau & Saunier, 2018; Mewada et al., 2020) e.g. tissue shape extraction, tumor detection, aerial, and natural image segmentation (Patel, Mewada & Patnaik, 2012), and texture image segmentation (Mewada, Patel & Patnaik, 2015). The energy calculation based deforming nature of contour in an image plane can well extract the shape of the object automatically. Therefore, this paper proposes a non-rigid object tracking using the active contour model. The major challenge in contour-based tracking is the computation time. Hence, a concurrent approach for non-rigid object tracking and its shape extraction using active contour model and mean shift tracking algorithm is proposed. The overall structure of the paper is as follows: “Related Work” classifies the tracking algorithms and then presents the study of contour-based tracking algorithms. “Proposed Non-Rigid Object Tracking Algorithm” introduces fast active contour model initially and later the proposed concurrent approach is explained in detail. “Results and Discussion” discusses the obtained results and their analysis. In the end, a conclusion and future scope are presented in “Conclusion”.

Mewada et al. (2021), PeerJ Comput. Sci., DOI 10.7717/peerj-cs.373
RELATED WORK
The three crucial steps involved in the tracking process of moving objects in a video are: (a) detection of the desired object (b) motion tracking of that object and (c) behavioural analysis or monitoring of the object to fulfil a particular task. The detection of the object requires a mechanism either in all sequences or in one particular frame of the sequences referred to as a reference frame. Based on the detection approach, the tracking process can be classified as detection inclusive or detection free tracking. The use of a particular object detection algorithm and its linking to trajectories in the tracking process (Bose, Wang & Grimson, 2007; Yu & Medioni, 2008) is referred to as detection inclusive tracking whereas, manual initialization of object in the first frame and localizing it in subsequent frames is referred to as detection free tracking (Hu et al., 2012; Yang, Lu & Yang, 2014; Zhang & Van der Maaten, 2013). Detection free tracking is the more efficient however it requires user interaction and thus it is a semi-automatic method.

The most commonly used detection mechanisms are background subtraction, static approach, time difference, and optical flow (Yang, Shi & Yi, 2012). In optical flow (Mahraz, Riffi & Tairi, 2014; Beaupré, Bilodeau & Saunier, 2018), the motion vectors obtained from the features set of the object to be tracked are passed through the optical flow function for each frame. The motion vector calculation requires patches of the images and hence it is a time-consuming process causing difficulties in real-time implementation. Time difference mechanism finds the motion regions and calculates the difference between two consecutive frames’ pixels. It is mostly adaptive but erroneous in the detection. In the static approach, a reference frame not containing moving objects i.e. static background is used. All pixels in the current frame are modeled by Gaussian distribution function (GDF) by finding the maximum intensity difference between these two frames. The major constraint of this method is the requirement of the background frame. In the background subtraction process, pixels are modeled by Gaussian mixture model (GMM) using few initial frames to find the foreground object. Later, pixels are processed using morphological operation and thresholding to find actual objects (Israni & Mewada, 2018b).

Object tracking methods can be categorized based on utilization of the features and approaches used for the detection. In features based VOT, various features of the object like color (Aoki, Oki & Miyamoto, 2020), texture (Khilar et al., 2020), edges, and statistical features (Israni & Mewada, 2018a) are used to track the object. Detection based tracking algorithms can be classified further as point tracking, kernel tracking, Silhouette/contour tracking, and supervised learning approach. The tracking using supervised learning requires pre-training of the algorithm and hence it is largely used for the online tracking process. Point tracking is robust to the illumination but it requires an additional mechanism to find the object in each frame. The kernel-based tracking is poor to track a non-rigid object in the video. Silhouette/contour-based tracking can handle a large variety of shapes of the desired object.

Major tracking algorithms have used edge-based active contour (i.e. snake) (Leymarie & Levine, 1993), active meshes (Valette, Magnin & Prost, 2001), or region-based active contour (Pateux, 2000). A kernel and contour-based hybrid model was presented in...
(Yilmaz, 2007), where authors generated the kernel using implicit contour and estimated the object’s new location with its scale and orientation. Yi, Ahn & Choi (2008) used a mean shift algorithm in the tracking process and presented a probabilistic approach to estimate the orientation and scale of the object. The use of the fixed shape kernel failed to detect the shape when the object rotates out of the plane. Nakhrani & Tannenbaum (2012) eliminated the false loop generation in parametric contour and used the parametric contour to track the human under noisy environment.

Silhouette based shape representation and tracking of moving object were presented by Toshev, Makadia & Daniilidis (2009). They used the Kanade-Lucas-Tomasi (KLT) tracker to extract the features and track the object. Using the obtained track labels, they segmented the object and obtained its silhouettes. They evaluated their model on a dataset containing a car, airplane, and helicopter. In Cremers (2006), tracking of multiple objects using posterior probability framework was proposed and the shapes of objects were retrieved using the level-set. Non-rigid object tracking in video captured by moving camera was proposed in Lefèvre & Vincent (2004). They used edge-based ACM to track and segment the object. To achieve real-time performance on outdoor scene video sequences, the gradient information was computed in the region of the object only once per frame. To quantify animal behaviour, a body and leg tracking process of Drosophila was presented by Uhlmann et al. (2017). A semi-automatic active contour-based tracking method reduced the user intervention and opened a new direction to annotate the video of species. An optical flow based background modelling and car’s shape extraction and its classification using Hidden Markov Model (HMM) were presented in Thakoor & Gao (2005). The feature set including color and texture was integrated to level set and proposed in Xian, Hu & Zheng (2013) to extract non-rigid shape for surveillance dataset.

Baum (2013) utilized depth images along with color information of the object to track it. By integrating the depth of the object, the shape of the object was generated. Lucas-Kanade algorithm was used in optical flow measurement and the Generalized Probabilistic Descent model was used with HMM. A scale and orientation adaptive kernel, driven through an active contour model was proposed by Sun et al. (2020). They evaluated the algorithm on VOT2014 and VOT2016 dataset achieving an average accuracy of 65.93% and 54.53% respectively. A fragment-based object tracking in the contour framework was proposed in Chockalingam, Pradeep & Birchfield (2009). A saliency map generation using Deep Neural network for object tracking and shape extraction was proposed in Zhang et al. (2020). This supervised algorithm trained for one dataset is not applicable to another dataset unless the network is re-trained. In summary, a very limited tracking algorithm was proposed which provides the shape of the tracking object. The drawback of this algorithm is represented in its high computational cost and limited accuracy and that is in comparison with bounding box based tracing algorithm.

A sparse representation can reduce the data vector size and increase the speed of execution (Patel & Mewada, 2018). A sparse dictionary formation and visual object tracking was presented in Joshi & Mewada (2020). Multiple patches are extracted from each frame to form the sparse dictionary, and likelihood measurement is carried out to track singular object handling long term occlusion. They conclude that patch wise joint
sparse tracking produced better results under various constraints of occlusion, illumination variation, deformation and blurring. Automated tracking using Gaussian Mixture Model (GMM) in real-time for the robotic application was presented in Patoliya & Mewada (2019). They achieved 85% tracking accuracy in the ROS environment. Supervised non-rigid object tracking algorithms using RGB-depth sensors are also proposed in many literatures. A convolutional neural network based learnable approach for non-rigid tracking was proposed in Bozic et al. (2020). They used RGB-depth images and predicted dense correlates for tracking the object. Sengupta (2020) reduced a combination of photometric and depth error and presented tracking of the deformable object using RGB-depth camera on both synthetic and real data. They used visual information from the RGB-depth camera and evolved the contour by calculating the contour forces. Li et al. (2020) generated pre-conditioner from a large number of training samples and proposed conditionNet in their process and increased the accuracy and speed of non-rigid RGB-depth object tracking process employing a learnable optimization process.

To track the vehicles and get information about the vehicles in a complex environment, a mean-shift and active contour-based tracking algorithm was proposed in Cai et al. (2017). Similarly, a contour of the face region is tracked by Zhou et al. (2013) using the Gradient Vector Flow (GVF) contour model and mean-shift algorithm. They proposed the tracking even under partial occlusion and full occlusion scenarios. The proposed algorithm is similar to the non-rigid tracking algorithm proposed in Sun et al. (2020), but it also uses a mean-shift algorithm for tracking and region-based ACM for shape extraction. The main objective of the proposed model is to reduce the computational complexity of ACM. Therefore, a region-based contour model with fast contour topology adoption was integrated with the mean-shift algorithm in contrast to modifying the kernel of the mean-shift algorithm used in Sun et al. (2020).

PROPOSED NON-RIGID OBJECT TRACKING ALGORITHM
As discussed in previous section, many algorithms were proposed using a bounding box with high tracking accuracy. The classification of algorithms was also presented in “Related Work”. Among them, and as a kind of Kernel-based tracking algorithm, the mean shift tracking algorithm is the most popular due to its simplicity and tracking efficiency. The block diagram of the proposed hybrid contours based mean-shift algorithm is shown in Fig. 1.

As shown in Fig. 1A fast region-based active contour model is used to obtain the contour of the desired object. A color probability density function (PDF) is used as a feature of the object to be tracked in the proposed algorithm. The object in the first frame is selected manually. The shift of the target in the next frame is calculated using the mean-shift procedure. The last contour shape is translated to this new position in the next frame. The contour evolves in the next frame to obtain the shape of the object. The process is repeated for all frames.
Fast active contour

The level set method is largely utilized in image segmentation applications due to its capability to handle topological changes automatically. It is a deformable contour that evolves based on applied certain energies. A Region-based ACM using level set method was initially proposed by Chan & Vese (2001). This piecewise approximated region-based ACM on image \( I(x,y) \) over the image plane \( \omega \) is defined by the energy function i.e. \( E_{CV} \) in (1). The first two terms represent the regularization term and referred to as length and area of the closed contour (C). These two terms control the smoothness of the contour and propagation speed of the evolving contour respectively.

\[
E_{CV}(m_1, m_2, \phi) = \int_{\Omega} \int \|I(x,y) - m_1\|^2 H(\phi(x,y)) dxdy \\
+ \int_{\Omega} \int \|I(x,y) - m_2\|^2 H(1 - \phi(x,y)) dxdy \\
+ \mu \int_{\Omega} \int \delta_\varepsilon(\phi(x,y)) |\nabla(\phi(x,y))| dxdy + \nu \int_{\Omega} \int H(\phi(x,y)) dxdy
\]

where, \( m_1 \) and \( m_2 \) are the mean intensity values inside and outside the contour and \( \phi \) is signed distance function. \( H_\varepsilon(z) \) and \( \delta_\varepsilon \) are slightly regularized version of Heaviside function (H) and dirac \( \delta \) function as \( \varepsilon \rightarrow 0 \), defined as follow.

\[
H(z) = \begin{cases} 
1, & \text{if } z \geq 0 \\
0, & \text{if } z \leq 0 
\end{cases} \\
\delta(z) = \frac{d}{dz} H(z)
\]

The minimization of these energies pushes the contour to a desired boundary of the object. These energies can be minimized by calculating the first derivative of the level-set function as follows:

\[
\frac{\partial \phi}{\partial t} = \delta_\varepsilon \phi \left[ \mu K - \nu - \lambda_1 (I - m_1)^2 + \lambda_2 (I - m_2)^2 \right]
\]

Where, \( K = \text{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \) is the Kappa function representing Euclidean curvature of the contour \( C \), and \( \lambda \) is a positive constant. Figure 2 shows the implied representation of the
contour $C$ using the level-set function. The contour deforms by moving the contour point either inside or outside the shape based on the minimization of the energy function.

To prevent the contour being too at or steep near the interface, a level is generally initialized as signed distance function (SDF) in the kappa equation, i.e. on the contour; all values of the point are zero, outside the contour; all values are positive constant and inside the contour; all values are negative constant as shown in the Fig. 2. However, the re-initialization of the contour shifts the zero level-set function away from its present boundary. The suggested solutions were to incorporate partial differential equations ([Sussman & Fatemi, 1999]) and penalty terms ([Li et al., 2005]) in the length calculation of the contour. However, re-initialization is the most time-consuming process in the level-set function, i.e. 90% of the time in contour evaluation is spared to re-initialize and to regularize the contour ([Patel, Mewada & Patnaik, 2012]) and hence re-initialization cannot be used in video tracking applications. [Patel, Mewada & Patnaik (2012)] used the Gaussian smoothing function and binarization process by eliminating the re-initialization used in the Kappa equation. Thus they modified the evaluation of the contour using the following minimization function.

$$\frac{\partial \phi}{\partial t} = \delta_s [\lambda_1 (I - m_1)^2 + \lambda_2 (I - m_2)^2]$$  (4)

Another weakness of the standard region-based ACM is that resultant contour is highly sensitive to the placement of the initial contour in an image. In contrast to regular region-based ACM, the fast region-based ACM extracts the object boundary accurately irrespective of the placement of contour as shown in Fig. 3. This helps to track the object accurately even when the mean-shift tracker fails to locate the object. Therefore, in this paper, we used this solution to avoid the regularization and we modified the active contour to using the mean-shift tracking method to track the object and extract the non-rigid object’s shape.
A fast region-based active contour using mean-shift tracking

In the mean-shift tracking method, a mean of the cluster is calculated and this mean is shifted to a new location in the next frame by finding a similar mean in the next frame. An object is selected using a rectangle bounding box in the reference frame. Let $x_i$, where $i = 1, 2, \ldots, n$, represents the pixel location of the object in the reference frame and the object is modelled using the histogram of the color of the object, then the probability (i.e. $q$) of the color (i.e. $c$) can be represented as

$$q_c = \frac{\sum_{i=1}^{n} \exp\left(-\frac{1}{2}||x_i||^2\right) \delta(b(x_i) - c)}{\sum_{i=1}^{n} \exp\left(-\frac{1}{2}||x_i||^2\right)}$$  

(5)

Where $b$ is histogram bin and $\delta$ is Kronecker delta function. If target object in the next frame is centred at $y_i$, then the probability ($p$) of the color ($c$) in target object can be represented as

$$p_c(y) = \frac{\sum_{i=1}^{n} \exp\left(-\frac{1}{2}||\frac{y - y_i}{h}||^2\right) \delta(b(x_i) - c)}{\sum_{i=1}^{n} \exp\left(-\frac{1}{2}||\frac{y - y_i}{h}||^2\right)}$$

(6)

where $h$ is the scale where points $x_i$ are considered for probability estimation. Thus, object is localized in the next frame by finding the similarity between the models in Eqs. (5) and (6). The similarity is calculated using Bhattacharya coefficient. Using these color density models, the weight matrix at pixel ($x, y$) can be represented as

$$W(x, y) = \sqrt{\frac{q_c}{p_c(y)}}$$

(7)

The large value of the weight $W$ represents that the obtained model belongs to the target object model and vice versa. Therefore, the weight $W$ is incorporated into region-based
active contour model to obtain the non-rigid shape of the object. Therefore, Eq. (1) can be modified to look for object in the frame as follows:

\[
E_{CV}(m_1, m_2, \phi) = \int_\Omega \int ||W(x, y) - m_1|^2 H(\phi(x, y))|dx\,dy \\
+ \int_\Omega \int ||W(x, y) - m_2|^2 H(1 - \phi(x, y))|dx\,dy \\
+ \mu \int_\Omega \int \delta_c(\phi(x, y)) \nabla(\phi(x, y)) |dx\,dy + \nu \int_\Omega \int H_c(\phi(x, y)) |dx\,dy
\]

(8)

Where, \( m_1 = \frac{\int_\Omega \int W(x,y) H(\phi(x,y)) |dx\,dy}{\int_\Omega \int H(\phi(x,y)) |dx\,dy} \) and \( m_2 = \frac{\int_\Omega \int W(x,y)(1-H(\phi(x,y))) |dx\,dy}{\int_\Omega \int (1-H(\phi(x,y))) |dx\,dy} \)

The minimization of the energy function in Eq. (8) gives the final position of the contour over the object boundary and is given by

\[
\frac{\partial \phi}{\partial t} = \delta_c \phi [\lambda_1 (W(x,y) - m_1)^2 + \lambda_2 (W(x,y) - m_2)^2]
\]

(9)

Therefore, the proposed model which integrates the mean-shift algorithm with ACM is having a fast evaluation in comparison with the model proposed in Sun et al. (2020).

RESULTS AND DISCUSSION

This section analyses quantitatively the results obtained through the proposed model using various video datasets. The proposed algorithm is designed in MATLAB V2020, installed on Pentium I7 processor with 8GB RAM, 1.90 GHz PC. The proposed algorithm is also tested for real-time videos. In all cases, the color model with 16 bins is used as a feature space in the mean-shift model. Initially, a rectangle bounding box is set manually. The first experimental result of table tennis video is shown in the Fig. 4. This video consists of 52 frames sequence with a resolution of 352 × 240. The frame rate is 15 frames/second. These are the basic steps of object tracking in real-time video using a mean shift algorithm.

In Fig. 4A, top left frame shows an initial frame of the video. Taking this as a reference frame, a square mask is created on it. In all frames, the object is tracked in rectangle throughout the video as shown in following images. As shown in Fig. 4, the mean-shift cannot track objects when the frame goes relatively fast and the object is moving fast in that frame. In Fig. 4B, proposed ACM model using mean-shift tracker is depicted. As explained in “Related Work”, mean shift tracker may fail to localize the object in the next frame and hence it fails to track object. However, the contour can shift its position by minimizing the energy function and locate the object in the deformation process irrespective of its placement carried out by the mean-shift tracker. Therefore, it tracks object correctly with retrieval of the shape as shown in Fig. 4B.

Another experiment was done on real-time video with complex background to validate processing capability of fast ACM. Figure 5 shows the hand tracking in real-time video captured using the standard camera where background contains a variety of shapes. The visual results of the mean-shift and proposed model show that the proposed model gives robust and accurate tracking of the hand object. In some frames of video, the rectangle window obtained using mean-shift is not covering the hand region from the centre of the bounding box. The proposed model gives correct tracking of the actual shape of the object.
in most of the frames throughout the video. Other experimental analysis was obtained
for the video containing slow and rapid movement of the object.

Figure 6 shows a video clip of foreman with warped background and moving his
head dynamically. At the end of video, the movement is very rapid. The warped
background also creates the blocking artefacts and hence contour may fail to retrieve the
desired boundary from of the tracking object. The object’s boundary is degraded at some
point but mean-shift tracker assist contour to stick with defined boundary and hence it
extracts the shape precisely.

Here another experimental result is shown in Fig. 7, for a man tracking video containing
91 frames sequence with a resolution of 380 × 216. The frame rate is 30 frames/second.
A tracking target is a man with a briefcase. Here in this video sequence, the proposed
model succeeds to follow the boundary of the man in all sequences. In some of the frames,
the contour did not evolve completely (i.e. near the face region in the frame 69, 95 and 130) because the number of iterations used in the contour evaluation is not competitive enough. Large number of iterations is required due to large object’s size.

The last video sequence is of news reader, Akiyo. It contains 300 frames with 192 × 144 resolution and a frame rate of 30 frames/second. The tracking target is the face of Akiyo whereas the moving objects are lips and eyes only. Here, the face is an object to detect and track. The result of the proposed model is shown in the Fig. 8. It is visible that at some instant the result is not accurate. This is due to background matches with the hair of Akiyo.
and also a television in the background is a non-targeted object for contour detection. Else overall, tracking and detection are good and accurate.

Few frames from the different videos, where the algorithm can track the object correctly but not able to segment accurately are shown in Fig. 9. In the first two figures, the initialized contour has more aspect ratio than the desired object’s aspect ratio. Therefore, the assigned number of the iteration is less to evolve the contour over the desired boundary. In 155th frame of foreman sequences, the face is occluded by the hand, and hence contour fails to grow over the face region.

![Tracking result of foreman sequences](image-url)

**Figure 6** Tracking result of foreman sequences (A) using mean-shift tracking (B) using proposed algorithm (Frames 1, 6, 51, 84, 114, 126 are displayed). Image credit: Li et al. (2020). DOI: 10.7717/peerj-cs.373/fig-6
Table 1 shows the quantitative analysis of all videos. Table 1 proposes that the reduction of the computation complexity in a fast-region based ACM model allows the mean-shift tracker to extract the shape of the object even in 30 frames/second videos. Further the failure of tracking due to mean-shift algorithm is also handled by the contour model. The foreman video contains rapid movement in few of the frames and hence mean-shift tracker fails to localize the head of foreman and due to clutter background, the result has been degraded in comparison with other videos.

Table 1 shows the quantitative analysis of all videos. Table 1 proposes that the reduction of the computation complexity in a fast-region based ACM model allows the mean-shift tracker to extract the shape of the object even in 30 frames/second videos. Further the failure of tracking due to mean-shift algorithm is also handled by the contour model. The foreman video contains rapid movement in few of the frames and hence mean-shift tracker fails to localize the head of foreman and due to clutter background, the result has been degraded in comparison with other videos.

Traditional level-set based active contour model requires a computational expensive re-initialization step to keep the level set function closed to \( \phi \). The use of smoothing and
The binarization process in the active contour model removed the re-initialization process making the evolution of active contour fast. Also, the proposed active contour is placement insensitive. When mean-shift tracker initialize the contour (i.e. rectangle bounding box) far from the desired object, the evaluation of contour correctly segment the desired object.

Table 2 presents and compares accuracy using correctly tracked frames. Irrespective of dataset, the proposed approach is compared with state-of-art methods of non-rigid object tracking. Most of the state-of-art methods used the VOT2014 dataset (Kristan et al., 2014) consisting of 25 sequences with a total size of more than 10,000 frames. The first comparison is established with Hough transform based pixel voting (PixVot) method which is integrated with probabilistic segmentation algorithm (Duffner & Garcia, 2013) and a patch tracking (PT) method. Region of tracking was divided into patches and correlation filter was used to track in Li, Zhu & Hoi (2015). Sevilla-Lara & Learned-Miller (2012) estimated the distribution field (DF) using features of target objects. Another
Hough tracker (HT) based graph cut algorithm was presented in Godec, Roth & Bischof (2013) for non-rigid object tracking. In (Sun et al., 2015), a supervised level-set model (SLM) was proposed where target was tracked using online boosting approach. The mean-shift kernel is sensitive to the orientation and scale of the object. Hence, adaptive tracker driven by data referred to as data-driven kernel (DDK) using level-set was proposed in (Sun et al., 2020). Due to placement insensitive nature of active contour used in the proposed model boosted the performance in comparison with (Sun et al., 2020).

**CONCLUSION**

In contrast to rectangle boundary representation in visual tracking, this paper proposes the appearance based representation in the tracking process. Simultaneous tracking of an object and its non-rigid shape extraction is a crucial task. This paper modified and proposed an active contour based tracking algorithm. A mean-shift tracker is incorporated in active contour model to track object using feature density function. To reduce the computational complexity, a re-initialization and regularization less active contour model was used in the proposed model. The independent contour deforms according to the shape of the object and hence overcome the weakness of false tracking of mean-shift tracker. The result analysis propagates that the proposed model outperforms other models by increasing accuracy to 71.86%. The proposed model is applicable to single object tracking. The fixed number of iterations have been used in evaluating the contour. This iteration highly depends on the size of the object to be tracked. Therefore, if the size of the object scales down dramatically, then it may not be able to segment the tracked object accurately. Therefore, a stopping criterion to stop contour evaluation is required. A further modification is required to track multiple objects using the active contour model as well as to tackle the occlusion scenario.

| Method               | PixTrac (Duffner & Garcia, 2013) | DF (Sevilla-Lara & Learned-Miller, 2012) | HT (Godec, Roth & Bischof, 2013) | SLM (Sun et al., 2015) | PT (Li, Zhu & Hoi, 2015) | DDK (Sun et al., 2020) | Proposed |
|----------------------|-----------------------------------|-----------------------------------------|-------------------------------|-----------------------|-----------------------|------------------------|----------|
| Average accuracy     | 41.33                             | 44.99                                   | 51.72                         | 56.41                 | 65.48                 | 65.92                  | 71.86    |

**ADDITIONAL INFORMATION AND DECLARATIONS**

**Funding**

The authors received no funding for this work.

**Competing Interests**

The authors declare that they have no competing interests.

**Author Contributions**

- Hiren Mewada conceived and designed the experiments, performed the experiments, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.
• Jawad F. Al-Asad conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
• Amit Patel performed the experiments, performed the computation work, authored or reviewed drafts of the paper, and approved the final draft.
• Jitendra Chaudhari performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.
• Keyur Mahant conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
• Alpesh Vala conceived and designed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

Data Availability
The following information was supplied regarding data availability:
Sample videos for non-rigid object tracking are available in the Supplemental Files.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.373#supplemental-information.

REFERENCES
Aoki R, Oki T, Miyamoto R. 2020. Accuracy improvement of human tracking in aerial images using error correction based on color information. In: Proceedings of the 2020 9th International Conference on Software and Computer Applications. 124–128.
Baum M. 2013. Simultaneous tracking and shape estimation of extended objects. Vol. 13. Karlsruhe: KIT Scientific Publishing.
Beaupré D-A, Bilodeau G-A, Saunier N. 2018. Improving multiple object tracking with optical flow and edge preprocessing. arXiv. 1–6. Available at http://arxiv.org/abs/1801.09646.
Bose B, Wang X, Grimson E. 2007. Multi-class object tracking algorithm that handles fragmentation and grouping. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 1–8.
Bozic A, Palafox P, Zollhöfer M, Dai A, Thies J, Niessner M. 2020. Neural non-rigid tracking. Advances in Neural Information Processing Systems 33:1–11.
Cai I, Wang L, Li B, Zhang L, Lv W. 2017. A novel vehicle tracking algorithm based on mean shift and active contour model in complex environment. In: Automated Visual Inspection and Machine Vision II. Vol. 10334. International Society for Optics and Photonics, 103340K.
Chan TF, Vese LA. 2001. Active contours without edges. IEEE Transactions on Image Processing 10(2):266–277 DOI 10.1109/83.902291.
Chockalingam P, Pradeep N, Birchfield S. 2009. Adaptive fragments-based tracking of non-rigid objects using level sets. In: 2009 IEEE 12th international conference on computer vision. Piscataway: IEEE, 1530–1537.
Cremers D. 2006. Dynamical statistical shape priors for level set-based tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(8):1262–1273 DOI 10.1109/TPAMI.2006.161.
Duffner S, Garcia C. 2013. Pixeltrack: a fast adaptive algorithm for tracking non-rigid objects. In: Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2480–2487.
Godec M, Roth PM, Bischof H. 2013. Hough-based tracking of non-rigid objects. *Computer Vision and Image Understanding* **117**(10):1245–1256 DOI 10.1016/j.cviu.2012.11.005.

Hu W, Li X, Luo W, Zhang X, Maybank S, Zhang Z. 2012. Single and multiple object tracking using log-euclidean riemannian subspace and block-division appearance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(12):2420–2440 DOI 10.1109/TPAMI.2011.283.

Israni D, Mewada H. 2018a. Feature descriptor based identity retention and tracking of players under intense occlusion in soccer videos. *International Journal of Intelligent Engineering and Systems* **11**(4):31–41 DOI 10.22266/ijies2018.0831.04.

Israni D, Mewada H. 2018b. Identity retention of multiple objects under extreme occlusion scenarios using feature descriptors. *Journal of Communications Software and Systems* **4**(4):291–300.

Joshi Y, Mewada H. 2020. Object tracking in occlusion and contrast conditions using patch-wise sparse method. *International Journal of Intelligent Engineering and Systems* **13**(5):295–305 DOI 10.22266/ijies2020.1031.27.

Khilar R, Sahoo SK, Rani C, Shanmugam PK. 2020. An efficient dynamic background subtraction algorithm for vehicle detection tracking system. In: Das K, Bansal J, Deep K, Nagar A, Pathipooranam P, Naidu R, eds. *Soft Computing for Problem Solving: Advances in Intelligent Systems and Computing*. Vol. 1048. Singapore: Springer, 551–562.

Kristan M, Pfugfelder R, Matas J, Cehovin L, Nebehay G, Vojir T, Fernandez G, L. A. 2014. The visual object tracking vot2014 challenges results. In: *European Conference on Computer Vision 2014 Workshops*. Cham: Springer, 191–217.

Larson J. 2020. How many frames per second can the human eye see? Available at *https://www.healthline.com/health/human-eye-fps* (accessed 10 November 2020).

Lefèvre S, Vincent N. 2004. Real time multiple object tracking based on active contours. In: *International Conference on Image Analysis and Recognition*. Berlin: Springer, 606–613.

Leymarie F, Levine MD. 1993. Tracking deformable objects in the plane using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(6):617–634 DOI 10.1109/34.216733.

Li Y, Bozic A, Zhang T, Ji Y, Harada T, Nießner M. 2020. Learning to optimize non-rigid tracking. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 4910–4918.

Li X, Hu W, Shen C, Zhang Z, Dick A, Hengel A. 2013. A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology (TIST)* **4**(4):1–48.

Li C, Xu C, Gui C, Fox MD. 2005. Level set evolution without re-initialization: a new variational formulation. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Vol. 1. Piscataway: IEEE, 430–436.

Li Y, Zhu J, Hoi SC. 2015. Reliable patch trackers: robust visual tracking by exploiting reliable patches. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 353–361.

Mahraz MA, Riffi J, Taiiri H. 2014. Motion estimation using the fast and adaptive bidimensional empirical mode decomposition. *Journal of Real-Time Image Processing* **9**(3):491–501 DOI 10.1007/s11554-012-0259-4.

Mewada H, Patel AV, Chaudhari J, Mahant K, Vala A. 2020. Composite fuzzy-wavelet-based active contour for medical image segmentation. *Engineering Computations* **37**(9):3525–3541 DOI 10.1108/EC-11-2019-0529.
Mewada H, Patel R, Patnaik S. 2015. A novel structure tensor modulated Chan–Vese model for texture image segmentation. *The Computer Journal* 58(9):2044–2060 DOI 10.1093/comjnl/bxu143.

Nakhmani A, Tannenbaum A. 2012. Self-crossing detection and location for parametric active contours. *IEEE Transactions on Image Processing* 21(7):3150–3156 DOI 10.1109/TIP.2012.2188808.

Patel H, Mewada H. 2018. Dictionary properties for sparse representation: implementation and analysis. *Journal of Artificial Intelligence* 11(1):1–8 DOI 10.3923/jai.2018.1.8.

Patel R, Mewada H, Patnaik S. 2012. Fast and regularization less active contour. *International Journal of Computer Applications* 47(6):26–31.

Pateux S. 2000. Tracking of video objects using a backward projection technique. In: *Visual Communications and Image Processing*. Vol. 4067. Bellingham: International Society for Optics and Photonics, 1107–1115.

Patoliya JJ, Mewada HK. 2019. Automated tracking and real time following of moving person for robotics applications. *Journal of Automation, Mobile Robotics and Intelligent Systems* 13(4):31–37 DOI 10.14313/JAMRIS/4-2019.35.

Santosh K, Lamiroy B, Wendling L. 2011. Dtw for matching radon features: a pattern recognition and retrieval method. In: *International Conference on Advanced Concepts for Intelligent Vision Systems*. Berlin: Springer, 249–260.

Santosh K, Lamiroy B, Wendling L. 2013. Dtw-radon-based shape descriptor for pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 27(3):1350008 DOI 10.1142/S0218001413500080.

Sengupta A. 2020. Visual tracking of deformable objects with RGB-D camera. PhD thesis, INRIA Rennes-Bretagne Atlantique and University of Rennes, France.

Sevilla-Lara L, Learned-Miller E. 2012. Distribution fields for tracking. In: *2012 IEEE Conference on computer vision and pattern recognition*. Piscataway: IEEE, 1910–1917.

Shi Y, Karl WC. 2008. A real-time algorithm for the approximation of level-set-based curve evolution. *IEEE Transactions on Image Processing* 17(5):645–656 DOI 10.1109/TIP.2008.920737.

Sun X, Wang W, Li D, Zou B, Yao H. 2020. Object contour tracking via adaptive data-driven kernel. *EURASIP Journal on Advances in Signal Processing* 2020(1):1–13 DOI 10.1186/s13634-020-0662-0.

Sun X, Yao H, Zhang S, Li D. 2015. Non-rigid object contour tracking via a novel supervised level set model. *IEEE Transactions on Image Processing* 24(11):3386–3399 DOI 10.1109/TIP.2015.2447213.

Sussman M, Fatemi E. 1999. An efficient, interface-preserving level set redistancing algorithm and its application to interfacial incompressible fluid flow. *SIAM Journal on Scientific Computing* 20(4):1165–1191 DOI 10.1137/S1064827596298245.

Thakoor N, Gao J. 2005. Automatic video object shape extraction and its classification with camera in motion. In: *IEEE International Conference on Image Processing*. Vol. 3. Piscataway: IEEE, 1–4.

Toshev A, Makadia A, Daniilidis K. 2009. Shape-based object recognition in videos using 3d synthetic object models. In: *2009 IEEE conference on computer vision and pattern recognition*. Piscataway: IEEE, 288–295.

Uhlmann V, Ramdy P, Delgado-Gonzalo R, Benton R, Unser M. 2017. Flylimbtracker: an active contour based approach for leg segment tracking in unmarked, freely behaving drosophila. *PLOS ONE* 12(4):e0173433 DOI 10.1371/journal.pone.0173433.
Valette S, Magnin I, Prost R. 2001. Active mesh for video segmentation and objects tracking. In: *International Conference on Image Processing* (Cat. No. 01CH37205). Vol. 2. Piscataway: IEEE, 77–80.

Wang L, Lu H, Yang M-H. 2017. Constrained superpixel tracking. *IEEE Transactions on Cybernetics* 48(3):1030–1041 DOI 10.1109/TCYB.2017.2675910.

Xian S, Hu H-M, Zheng J. 2013. Moving object segmentation using level set method with shape prior, color and texture. *Journal of Multimedia* 8(5):589–595 DOI 10.4304/jmm.8.5.589-595.

Yang F, Lu H, Yang M-H. 2014. Robust superpixel tracking. *IEEE Transactions on Image Processing* 23(4):1639–1651 DOI 10.1109/TIP.2014.2300823.

Yang J-B, Shi M, Yi Q-M. 2012. A new method for motion target detection by background subtraction and update. *Physics Procedia* 33:1768–1775 DOI 10.1016/j.phpro.2012.05.283.

Yi KM, Ahn HS, Choi JY. 2008. Orientation and scale invariant mean shift using object mask-based kernel. In: *19th International Conference on Pattern Recognition*. Piscataway: IEEE, 1–4.

Zhang L, Van der Maaten L. 2013. Structure preserving object tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 1838–1845.

Zhou Z, Wu J, Yang W, Zhu Z. 2013. Face contour tracking based on mean shift and GVF snake. *Information Technology Journal* 12(10):1884–1894 DOI 10.3923/itj.2013.1884.1894.