Adaptive Non-Rigid Object Tracking by Fusing Visual and Motional Descriptors

Hadi Firouzi
Okanagan School of Engineering
University of British Columbia
Kelowna, Canada
Hadi.Firouzi@ubc.ca

Homayoun Najjaran
Okanagan School of Engineering
University of British Columbia
Kelowna, Canada
H.Najjaran@ubc.ca

Abstract—This paper presents a framework to track non-rigid objects adaptively by fusion of visual and motional feature descriptors. The proposed technique can automatically detect an object from different points of view as soon as the object starts moving. Moreover, an object model is created and gradually updated using both new and previous features. As a result, the proposed technique is able to track a non-rigid object even if the object is rotating or distorting. The efficacy of the proposed method is verified using the experimental results obtained from a grayscale camera.

Keywords-component: non-rigid object tracking; feature fusion; SIFT;

I. INTRODUCTION

Early object Visual object tracking is essential for many real-world applications involving machine vision. Some important examples include autonomous navigation and motion control, automated surveillance systems, video tracking, scene monitoring, and medical imaging. In its simplest form, visual object tracking can be defined as the problem of locating 3D objects in a 2D image plane as they move around a scene. Estimating the trajectory of objects in the image plane has a long tradition in computer vision research. A robust method often used to deal with this problem is the mean-shift method. The mean-shift method is an algorithm to find a region in a new image that is most similar to a specific region in the previous image based on an empirical density function. Zivkovic proposed a new 5-DOF color-histogram-based non-rigid object tracking based on a natural extension of the mean-shift algorithm. The algorithm simultaneously estimates the position of the local mode and the covariance matrix that describes the approximate shape of the local mode. Color based algorithm has recently been used to track non-rigid objects. One of the efficient color-histogram-based tracking methods is based on the mean-shift procedure. Zhou et al. proposed a new algorithm for optimally adapting the ellipse outlining the objects of interest. This algorithm is an improvement of the Zivkovic method in complex scenes. This method reduces the residuals between the estimated probability distribution and the expected one to adapt the object shape during the tracking. Brox et al. propose a method for tracking of rigid and articulated objects by combining surface-region matching, optical flow, and SIFT tracking. The method is designed to prevent the accumulation of errors using the object region and handle larger transformations using optical flow and SIFT.

Generally, there are three different categories of object tracking algorithms: 1) point tracking, 2) kernel tracking and 3) silhouette tracking. Among them, the latter is the most practical technique for real-time applications for its computational efficiency, but it can be limited in dynamic environments and presence of non-rigid objects. In this work, an adaptive approach to track non-rigid objects is proposed. The proposed method fuses visual and motional descriptors of a set of features extracted from a gray-scale camera images. In the proposed approach, tracking is carried out similar to point tracking but a silhouette is created to detect new objects and update visual descriptors so that visual descriptors can be specified for a general non-rigid object i.e., without training for a specific object or template. The visual descriptors are utilized to estimate local motion vectors of individual features by matching two sequential images. Since a non-rigid object can consist of a set of semi-connected components defined in an image, each feature can move independently while the whole object moves toward a certain direction. Consequently, the local motion vectors are used to update the motional descriptors of the features only after they pass an outlier rejection test involving comparisons with both local and global motion vectors. The updated motional descriptors are then used to estimate a global motion vector that defines the object motion using a weighted averaging scheme based on the quality of the features i.e., how the visual and motional descriptors of a feature are in agreement with the majority of the features and the previous state of the object and the feature.

II. NON-RIGID OBJECT

In Figure 1(A) some examples of non-rigid objects are shown. In this research, a non-rigid object \( O \) is defined as a semi-connected area; see Figure 1(B). This area is described by a set of features \( F_0 \) extracted by SIFT algorithm.

\[
F_0 = \{f_1, f_2, \ldots, f_M\}
\]
\( f_j = \{ \varphi_j, \theta_j \} \)

where \( M \) is the number of object features. Each feature \( f_j \), i.e. an interest point, is defined not only by a visual descriptor \( \varphi_j \) provided by SIFT, but also by a motional descriptor \( \theta_j \). Thus this motional descriptor is defined by feature position \( X_j \) and velocity \( V_j \).

\[
\theta_j = \{ X_j, V_j \}
\]  

Based on the definition of a non rigid object, each feature point can belong to a different part of the object and may move in different direction with different velocity while the object as a whole is going to a certain location with a specific velocity. As a result, feature position and velocity are defined based on their own history.

\[
X_j^t = (1 - \alpha)X_j^{t-1} + \alpha X_j^t
\]

\[
V_j^t = (1 - \alpha)V_j^{t-1} + \alpha (X_j^t - X_j^{t-1})
\]  

where \( V_j^{t-1} \) and \( V_j^t \) are the feature velocities in time \( t-1 \) and \( t \) respectively. Also \( X_j^{t-1} \) and \( X_j^t \) are the feature position in time \( t-1 \) and \( t \) respectively. \( \alpha \), a number between 0 and 1, is the feature position/velocity updating rate. By increasing the \( \alpha \), the current feature history is more ignored and vice versa.

\[
w_j = \frac{h_j}{(h_o + 1)} \| V_j^t - \bar{V}^t \| \| V_j^t - V_j^{t-1} \|
\]  

where \( h_j \) and \( h_o \) are the history length of feature and object respectively. Also \( \bar{V}^t \) is the average value of all feature motion vectors.

### III. DETECTING A NEW NON-RIGID OBJECT

To detect a non-rigid object usually a multiple appearance model or object template is used. However training a complete and accurate object model is not feasible in many practical cases such as video tracking. Object motion detection method is a practical way to detect a new object in a complex scene. In this research a combination of a background subtraction and a feature-based image motion detection method is used to detect a new non-rigid object. First a preliminary change image is generated by a typical background subtraction method (see Figure 2), then all interest points inside the change image are detected based on SIFT algorithm. These points are matched to the feature points extracted from the previous image. After matching SIFT features between the current and the previous images a motion image is generated based on matched features. Then according to the magnitude and direction of the motion vectors, the motion image is segmented into similarity regions. A connected region which satisfies a certain criteria (e.g. size and motion magnitude), is considered as a new non-rigid object, see Figure 4. The algorithm to detect new non-rigid object is described in Figure 3.
Figure 3: A block diagram of detecting a new non-rigid object

Figure 4: Detecting new non-rigid object, red point: basis features, white rectangle: new non-rigid object

A. Creating a new object model

Once a non-rigid object is detected, a new object model based on extracted SIFT features is created. Initially the object model is consisted of those features called basis features which are used to detect the non-rigid object, red points in Figure 4. However to increase the possibility of detection and improve the accuracy of tracking, all extracted features inside the object area i.e. unmatched features, blue points inside white rectangle in Figure 4, are added to object features. Indeed, adding unmatched features are useful when the object appearance is changing over the time.

\[ F_0 = \{ f_1, f_2, \ldots, f_N, f_{N+1}, \ldots, f_M \} \]  

where \( f_j \) for \( j \leq N \) are matched features and \( f_j \) for \( j > N \) are unmatched features. And \( N \) is the number of matched features.

In addition to visual descriptors, each object feature has a motional descriptor. For a matched feature, the motion vector is calculated by simply subtracting the previous position and new position of the matched features.

\[ V_j^t = X_j^t - X_j^{t-1}; j \leq N \]  

However, for an unmatched feature no motion vector can be directly calculated. Therefore based on the motion vector of the neighbor matched features, a new motion vector is estimated.

\[ V_j^t = \sum_{i=1}^{k} \partial_{j,i} V_i^t; j > N, i \leq N \]  

\[ \partial_{j,i} = 1 - \frac{\| V_j^t - V_i^t \|}{\sum_{k=1}^{N} \| V_j^t - V_k^t \|}; j > N, i \leq N \]

B. Updating object model

Updating the object model is an online process. In each image frame, the object model is adaptively updated by adding new features and removing unused ones. Assigned features are kept as object features as long as they are useful for detection and tracking. Generally features are removed from object feature vector if they are not matched for a certain number of consecutive images or if the total number of features exceeds a predefined number. Updating object feature vector based on object appearance changes, improve the tracking reliability and also increase the processing efficiency by ignoring unnecessary features. Moreover, different parts of a non-rigid object can move independently or hide behind other parts; therefore the object model must be adaptively updated to handle different situations by adding and removing the object features.

IV. TRACKING A NON-RIGID OBJECT

A non-rigid object is tracked through its features. First object features are roughly matched with the new extracted features from the current image based on their visual descriptors, and then matched features are filtered using their motional descriptors. In the other word, the estimated motion vector of a preliminary matched feature must be in agreement with its motional descriptor. In the following sections, two sequential steps to track a non-rigid object are described.

A. Tracking features based on visual descriptors

In this step all object features are roughly matched with the new extracted features based on their visual descriptors. To compare an object feature (\( f_j^{t-1} \)) with a new extracted feature (\( f_j^t \)), a Euclidean distance (\( D_j^t \)) based on their visual descriptors is computed. The more similar features have the smaller distance. If the shortest distance is smaller than a
predefined threshold \((D_{\text{thr}})\) the relevant features are considered as a matched pair. Thus this threshold is set to a proportionally large value to accept all possible matched pairs.

\[
D_f^j = \|f_j^{t-1} - \tilde{f}_j^t\| \\
( f_j^{t-1}, \tilde{f}_j^t) \text{ are a possible match if } D_f^j < D_{\text{thr}} \tag{9}
\]

B. Filtering matched features based on motional descriptors

After two features are matched based on their visual descriptor vectors, another similarity function is used to verify the correctness and reliability of the matched pair. In this function the actual motion vector is compared with a predicted motion vector. If the direction and magnitude of both motion vectors satisfy the similar criteria, the pair matched feature is considered as a correct and reliable one.

The similarity function is defined as following equation.

\[
V_f^j = \sigma \left( \frac{1}{2} + \frac{V_j^t \cdot \tilde{V}_j^t}{\|V_j^t\|\|\tilde{V}_j^t\|} \right) + (1 - \sigma) \left( 1 - \frac{\|V_j^t - \tilde{V}_j^t\|}{\sqrt{S_x + S_y}} \right) \\
( f_j^{t-1}, \tilde{f}_j^t) \text{ are a match if } V_f^j > V_{\text{thr}} \tag{10}
\]

where \(S_x\), \(S_y\) are image width and height respectively. \(\sigma \in [0,1]\) is a factor specifying the weight of velocity direction with respect to its magnitude. \(V_{\text{thr}}\) is a velocity similarity threshold. \(V_j^t\) and \(\tilde{V}_j^t\) are the real and predicted feature velocities. To predict the feature velocity both local and global velocities are used. Local feature velocity \((V_j^{t-1})\) is a velocity which is estimated based on only feature motion and the global feature velocity \((V_0^{t-1})\) is the object velocity. Using the following equations, the actual \((V_j^t)\) and predicted \((\tilde{V}_j^t)\) motion vectors are calculated.

\[
V_j^t = (1 - \alpha)V_j^{t-1} + \alpha(\tilde{X}_j^t - X_j^{t-1}) \\
\tilde{V}_j^t = (1 - \delta)V_0^{t-1} + \delta V_j^{t-1} \tag{11}
\]

where \(\tilde{X}_j^t\) is the position of the new matched feature and \(\delta \in [0,1]\) is a factor to make a balance between local and global feature velocities. Also \(\alpha\) is the feature position/velocity updating rate.

After tracking all object features, the object motion vector is obtained based on (4). The matched features are considered as the basis features. And similar to the object detection section A, all new features inside the object boundary (defined by basis features) are added to the object feature vector.

V. EXPERIMENTAL RESULTS

To show the accuracy and reliability of the proposed non-rigid object tracking method. A gray-scale camera is placed in an environment in which a person is picking up a cubic object with his hand and then he is passing the cube from one hand to the other hand. The combination of the cube and person’s hands makes a non-rigid object. Indeed, this non-rigid object includes different parts which can move independently and the hands can be add or removed from the whole object during the time. It is shown that the proposed technique is able to adaptively track the object whereas some parts are either adding or removing such as person hands.

VI. CONCLUSION

In this paper a new feature-based object tracking technique is presented. In essence, the new technique fuses visual and motional descriptors to track an object more effectively than the traditional techniques. A combination of typical background subtraction and image motion detection.
methods is used to automatically detect new non-rigid objects. The proposed method is designed to update the object model adaptively even if the object is rotating or distorting. Thus, newly extracted features and previous object features are integrated to increase the robustness and accuracy of the detection and tracking, especially where the object appearance changes. As a result, the proposed technique is less sensitive to visual descriptors than the traditional point tracking approaches; and hence, it can be executed with simpler and higher efficiency image analysis techniques. Finally, the experimental results show the advantage of fusing visual and motional information to improve the robustness and accuracy of the object tracking in dynamic environments.

VII. REFERENCES


