An Enhanced Video Tracking Technique Based on Nature Inspired Algorithm

1. Dr. Israa Hadi, 2 Mustafa Sabah
1: first author, Assistant Professor College of Information Technology University of Babylon, Israa_hadi1968@yahoo.com
2: Corresponding Author Ph.D. Student, AL-Manour University College, muustafa_bayat@yahoo.com

Abstract

One of the main goals of computer vision is to enable computer to replicate the basic functions of human vision such as motion perception and scene understanding. To achieve the goal of intelligent motion perception, much effort has been spent on object tracking, which is one of the most important and challenges research topic in computer vision.

In this paper, a new framework of video tracking based on cat swarm optimization is proposed to enhance the tracking techniques.

The proposed video tracking algorithm has multi stage: first stage is preprocessing step that convert video into multi frame, the second stage is object detection that deal with identifying an object of interest, in this stage we propose new idea that split the target region into four non overlapping area, for each area are we compute some important feature such as (color histogram, accumulated histogram, shape ratio, spatiogram), the third stage is the object tracking, in this stage we use block matching for track moving object but we proposed new method based on cat search optimization algorithm in order to reduce complexity, finally we propose new method to deal with the trajectory by converting the trajectory points into approximation function using curve fitting function and extract important features such as slope, intersection point. As the proposed algorithm does not consider any fixed search pattern or any other different assumption, a high probability for finding the true minimum (accurate motion vector) is expected.

Keywords: object tracking, hybrid feature, block matching, swarm algorithm, curve fitting.

1. Introduction & Background

The moving object tracking in video pictures has attracted a great deal of interest in computer vision. The aim of object tracking and detection is to establish a correspondence between objects or object parts in consecutive frames and to extract temporal information about objects such as trajectory, posture, speed and direction. Tracking detected objects frame by frame in video is a significant and difficult task.

Every tracking method requires an object detection mechanism either in every frame or when object first appears in the video. The existing object detection methods roughly classified into four categories: Point Detectors, Background Subtraction, Segmentation, Supervised Learning.

The Point detectors are used to find interest point in the images which have an expressive texture in their respective localities, Background Subtraction which object detection can be achieved by building a background model then finding the deviation from the model for each incoming frames where any significant changes in an image region from the background model signify a moving object the third type is Segmentation, the aim of image segmentation algorithms is to partition every image into perceptually similar regions, finally Supervised Learning which the object detection can be performed by learning different object views automatically from a set of examples by means of supervised learning mechanism. Typically, tracking over time consists of matching moving objects in successive frames. Tracking categories are Point Tracking, Kernel Tracking, Silhouette Tracking, and Shapes Matching.
Point Tracking where object detected in consecutive frames is represented by points and the association of the points based on the previous object state which includes object position and motion. This approach requires an external mechanism to detect the object in every frame. Kernel Tracking refers to object shape and appearance. Objects are tracked by computing the motion of the kernel in the consecutive frames. Silhouette Tracking is performed by estimating the object region in every frame. Silhouette tracking make the use of information encoded inside the object region. In shape matching object silhouette and its associated model is searched in the current frame. The search is performed by computing the similarity of the object with the model generated from the hypothesized object silhouette based on previous frame [5][13].

In video tracking, the high correlation between successive frames can be exploited to improve coding efficiency, which is usually achieved by using motion estimation (ME). Many ME methods have been studied in an effort to reduce the complexity of video coding, such as block matching (BM) algorithms. BM seems to be the most popular method due to its effectiveness and simplicity for both software and hardware implementations [4][6].

Because of computational drawbacks of conventional numerical methods in solving complex optimization problems, researchers may have to rely on meta-heuristic algorithms. Over the last decades, many meta-heuristic algorithms have been successfully applied to various engineering optimization problems. For most complicated real-world optimization problems, they have provided better solutions in comparison with conventional numerical methods. To imitate natural phenomena, most meta-heuristic algorithms combine rules and randomness. These phenomena include the biological evolutionary processes, such as genetic algorithm (GA), animal behavior, such as particle swarm optimization (PSO) and a colony algorithm, cuckoo search algorithm (CS) and cat swarm optimization (CSO).

Cat Swarm Optimization (CSO), a recent met heuristic technique firstly reported by Chu and Tsai, models the behavior of cats to solve the optimization problem. The optimal solutions obtained by the CSO are far better than the best solutions obtained by efficient particle swarm optimizers and genetic algorithms. However, only a few evolutionary approaches have specifically addressed the problem of BM, such as genetic block matching and the PSO-BM. Although these methods support an accurate identification of the motion vector, their spending times are very long, when compared with other BM techniques [8][11].

In order to relate with a trajectory point as a function, Curve fitting function will be an essential component of any mathematical interface. The main advantage of such mathematical interfaces over other methods is that they can more naturally embody the usual two-dimensional mathematical notations. This method of evaluating empirical formulas was developed over a century ago, the procedure involves approximation a function such that the sum of the squares of differences between the approximation function and the actual function [12].

2. Related work

Tracking is an important topic in computer vision and it has been studied for several decades. In this section we summarize studies that are related to our proposed tracking work:

S. J. Rajput, S.D.Oza [3] propose a new method for detection and tracking of multiple moving objects under static background, based on motion detection, background subtraction using canny edge detection with edge difference between consecutives frames.

Chittaliya, Trivel[6] propose A novel, simple and fast block matching algorithm has been proposed using predictive motion vector for object tracking. Color histogram is used for matching criteria for the motion tracking.

Pinar Civicioglu, Erkan Besdok [8]. In this paper, the algorithmic concepts of the Cuckoo-search (CK) and Particle swarm optimization (PSO) have been analyzed. The numerical optimization problem solving successes of the mentioned algorithms have also been compared statistically by testing over 50 different functions. The run-time complexity and the required function-evaluation number for acquiring global minimize by the cuckoo algorithm is generally smaller than the comparison algorithms. The CK algorithm supply more robust and precise results than the PSO.

Steven[12] explain different method for interpolation and curve fitting function, In this book, look at various ways of approximating functions from a given set of discrete data points. Interpolation is a
method for constructing a function $F(x)$ that fits a known set of data points $(x_k, y_k)$, i.e. a method for constructing new data points within the range of a discrete set of known data points. There are various ways in which this can be done.

Yongguo Liu and Yidong Shen[9] In this article, a recent meta heuristic method, cat swarm optimization, is introduced to find the proper clustering of data sets. The clustering approach based on cat swarm optimization called Cat Swarm Optimization Clustering (CSOC) is proposed. In the proposed methods, seeking mode and tracing mode are adopted to exploit and explore the solution space. Experimental results on six real life data sets are given to illustrate the effectiveness of the proposed algorithms.

Yuan, Shen[6] propose a fast block matching algorithm for motion estimation (ME) and compare the algorithm with other popular fast block-matching algorithms for ME. A real-world example shows that the block matching algorithm based on PSO for ME is more feasible than others. Moreover, the initial values of parameters in PSO are empirically discussed, since they directly affect the computational complexity. Thus, an improved PSO algorithm for ME is empirically given to reduce computational complexity.

3. Scope and the limitation

The main challenges that have to be taken into account when designing and operating a tracker are related to the similarity of appearance between the target and other objects in the scene, and to appearance variations of the target itself. The appearance of other objects and of the background may be similar to the appearance of the target and therefore may interfere with its observation. In such a case, image features extracted from non-target image areas may be difficult to discriminate from the features that we expect the target to generate. This phenomenon is known as clutter. In addition to this problem, video tracking is made difficult by changes of the target appearance in the image plane that are due to one or more of the following factors:

- Ambient illumination. The direction, intensity and color of the ambient light influence the appearance of the target. For example, ambient light changes when clouds obscure the sun.
- Changes in pose. A moving target varies its appearance such as rotation, translation operation.
- Noise. The image acquisition process introduces into the image signal a certain degree of noise, which depends on the quality of the sensor.
- Occlusions. A target may fail to be observed when partially or totally occluded by other objects in the scene. Occlusions are usually due to: A target moving behind a static object, such as a column, a wall, or other moving objects obscuring the view of a target.

To address this challenge, different approaches can be applied that depends on the expected level of occlusion: Partial occlusions that affect only a small portion of the target area can be dealt with by the target appearance model or by the target detection algorithm itself. The invariance properties of some global feature representation methods (e.g. the histogram) are appropriate to deal with occlusions. Also, the replacement of a global representation with multiple localized features that encode information for a small region of the target may increase the robustness of a video tracker.

- Change of the orientation of tracked object with passing of time will increase the complexity of the process.
- How do the motion estimation when the moving of object is too complex and fast?
- Color histogram neglects the similarity of colors thus when color of background is similar to moving object.
- There is no trajectory if the object is static [1] [3] [5] [7] [13].

4. Block matching algorithm

In video tracking, the high correlation between successive frames can be exploited to improve coding efficiency, which is usually achieved by using motion estimation(ME). Many ME methods have been studied in an effort to reduce the complexity of video coding, such as block matching(BM) algorithms.

BM seems to be the most popular method due to its effectiveness and simplicity for both software and hardware implementations. In a BM algorithm, the current frame is divided into non-overlapping
macro blocks (MB) of NXN pixel dimension. For each block, in the current frame, the best matched block within a search window of size \((2W+1)\times(2W+1)\) in the previous frame is determined, where \(W\) is the maximum allowed displacement. The position difference between a template block in the current frame and the best-matched block in the previous frame is called the motion vector. A commonly used matching measure is the sum of absolute differences (SAD), which is computationally expensive and represents the most consuming operation in the BM process.

The full search algorithm (FSA) is the simplest BM algorithm that can deliver the optimal estimation solution regarding the minimal matching error, because it checks all candidates one at a time. However, such exhaustive search and full-matching error calculation at each checking point yields an extremely computational-expensive FSA method that seriously constraints real-time video applications. To decrease the computational complexity of the BM process, several BM algorithms have been proposed, which are based on the following three techniques:

1. Using a fixed pattern, which means that the search operation is conducted on a fixed subset of the total search window, and some famous example include, the three step search (TSS). Such approach has been algorithmically considered as the fastest.

2. Reducing the search points signifies that the algorithm chooses only such locations as search points, which iteratively minimize the error-function (SAD values). In this category, some example include the adaptive road pattern search (ARPS). This approach assume that the error-function behaves monotonically, which holds well for slow-moving sequences; however, such properties do not hold true for other kind of movements in video sequences, yielding that the algorithms may get trapped into local minima.

3. Decreasing the computational overhead for each search point means that the matching cost (SAD operation) is replaced by a partial or simplified version with less complexity. These techniques are based on the assumption that all pixels within each block move by the same amount, while a good estimate of the motion could be obtained by using only a fraction of the pixels. However, as only a fraction of the pixels enters into the matching computation, the use of these regular sub-sampling techniques can seriously affect the accuracy of the detection of motion vectors due to noise or illumination changes [2][6][7].

5. Proposed algorithm

The novel framework of video tracking consists of the following logical component steps:

5.1 Preprocessing Step

Before performing any video processing operation the quality of the frame is very essential. In this phase improving the quality of frame is taken into the consideration and the video is tested with one type of noises, the each noise types are de noised using various filtering techniques and the best suited filters are taken into the consideration for different noises.

- Create a frame sequence from a video input.
- Take current image and previous image.
- Perform the filtering operation to remove the noise from the image [10].

5.2 Object Detection Step

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections.

Segmentation involves partitioning an image into a set of homogeneous and meaningful regions, such that the pixels in each partitioned region possess an identical set of properties or attributes. These sets of properties of the image may include gray levels, contrast, spectral values, or textural properties. The result of segmentation is a number of homogeneous regions, each having a unique label.

An image is thus defined by a set of regions that are connected and non overlapping, so that each pixel in the image acquires a unique region label that indicates the region it belongs to. Segmentation
algorithms are based on one of the two basic properties of gray-level values: discontinuity and similarity among the pixels [1][10].

5.3 Feature Selection & Target representation Step

Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness so that the objects can be easily distinguished in the feature space. Feature selection is closely related to the object representation. Feature extraction is the first step in the tracking pipeline and allows us to highlight information of interest from the images to represent a target. The features to be extracted can be grouped into three main classes, namely:

- Low-level (e.g. color, gradient, motion)
- Mid-level (e.g. edges, corners, regions)
- High-level (objects).

In the proposed algorithm use hybrid features in order to keep track the target, in each portion of object (one can split the target area into four regions with a regular grid use the following features:

- Color histograms have been used as target models for their invariance to scaling and rotation, robustness to partial occlusions, data reduction and efficient computation. Color histograms encode the statistical distribution of the pixel values, as each histogram bin contains global information, this representation, although less descriptive, is more invariant in case of partial occlusions and pose changes.
- Spatial information can be introduced by associating with each bin of the histogram the first two spatial moments of the pixel coordinates of the corresponding color (spatiograms). A spatiogram can alternatively; one can split the target area into regions with a regular grid and compute multiple histograms (one per region). In this case the target is represented by a vector concatenating the values of the histograms of each region. For example, a multi-region orientation histogram can be used, where the regions are the four sectors of the bounding ellipse.
- Accumulation histogram: general color histogram is one of the most basic color features, which indicates the occurring probability of various colors in the image. In general color histogram, color value is the abscissa, and the color occurring frequency in the image is the ordinate. Let \( \text{Sum}(P, x_i) \) denotes the number of pixels of feature data \( x_i \) in the image \( P \), and \( N \) denotes the total number of pixels in the image \( P \). Then, the general color histogram of the image \( P \) is shown as:

\[
H(P) = (h_{x_1}, h_{x_2}, \ldots, h_{x_i}, \ldots, h_{x_n})
\]

Where:

\[
h_{x_i} = \frac{\text{Sum}(P, x_i)}{N}
\]

In accumulation histogram, color value is the abscissa, while the color accumulation occurring frequency in the image is the ordinate. Supposing the general color histogram of some feature data: \( H(P) = (h_{x_1}, h_{x_2}, \ldots, h_{x_i}, \ldots, h_{x_n}) \) of image \( P \) is known, its accumulation histogram is shown as:

\[
\lambda(P) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_i, \ldots, \lambda x_n)
\]

Where:

\[
\lambda x_i = \text{the summation of all values of }\lambda x_i \text{ from 1 to I } [1][5][7].
\]

In this paper the target is represented by a vector concatenating the values of the histograms of each region. For example, a multi-region orientation histogram can be used, where the regions are the four sectors of the bounding region and compute multiple histograms (one per region) to describe object moving characteristics instead of general histogram where this histogram reflect the similarity of color distribution more accurately.

A target representation is a model of the object of interest that is used by a tracking algorithm. This model includes information about the shape and the appearance of the target. The model for a specific target can be computed in different ways: Point approximation, area approximation and histograms [7][13].
5.4 Localization Step

In this stage we will describe how to localize a target over time, given its initial position. After initialization, the localization step of a video tracker recursively estimates the state $x_k$, given the features extracted from the video frames and the previous state estimates $x_{k-1}$. Localization methods classify into two major classes:

- Single-hypothesis localization (SHL) methods, where only one track candidate estimate is evaluated at any time.
- Multiple-hypothesis localization (MHL) methods, where multiple track candidates are evaluated simultaneously. The ability to propagate multiple hypotheses can improve the performance of a tracker.

The second type considered in proposed algorithm is for motion estimation through a BM algorithm, BM can be approached as an optimization problem aiming for finding the best MV within a search space as we shown in figure 1 below.

![Figure 1. Calculation of Motion vector](image)

Fast motion estimation algorithms yield the motion vector by minimizing a cost function which is generally based on the distance criterion. Some of the commonly used distance criteria such as: MSE (mean square error) – It can be interpreted as the Euclidean distance between two MBs, which is close to the human visual perception. MSE also suffers from high computational complexity.

$$\text{MSE} (dx, dy) = \frac{1}{N^2} \sum_{m=x}^{x+N-1} \sum_{n=y}^{y+N-1} [I_k (m, n) - I_{k-1} (m + dx, n + dy)]^2 \quad (4)$$

$$\text{MV}_x, \text{MV}_y \text{ (min (dx, dy)) CR2 MSE (dx, dy)} \quad \text{………………. (5)}$$

Also there are many cost function such as MAE (mean absolute error), SAD (sum of absolute differences and Bhattacharyya) that measure and compare two probability distributions such as histogram [2] [6] [7].

5.5 Swarm intelligent Step

Swarm Intelligence is a property of a system whereby the collective behaviors of (unsophisticated) agents interacting locally with their environment cause coherent functional global patterns to emerge. SI is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates [8].

In this paper we implement the CA to optimization the computation time of search; the cat algorithm has two modes: [9]
5.5.1 Seeking mode

This mode is used to model the cat during a period of resting but being alert, looking around its environment for next move. In the seeking mode, four factors are defined as follows:

- **Seeking Memory (SM):** SM is used to set the size of seeking memory of the cat \(N_{SM}\). That is, \(N_{SM}\) denotes the number of neighboring positions of the cat.
- **Self Position Flag (SPF):** SPF indicates whether the current position of the cat will be one of neighboring positions. If \(F_{SPF} = 1\) the current position of the cat is viewed as a new neighboring position, otherwise neighboring positions of the cat should be different from its current position.
- **Changed Dimension (CD):** CD is used to set the number of dimensions to be mutated \(N_{CD}\).
- **Seeking Range (SR):** SR describes the mutative ration of the dimension to be mutated \(P_{SR}\).

5.5.2 Tracing mode

This mode is used to model the case of the cat in tracing targets. Once a cat goes into the tracing mode, it moves according to its velocities for each dimension.

\[
\begin{align*}
V_{ij} &= V_{ij} + r_1 \times c_1 \times (x_{ij} - x_{ji}) \quad \text{......... (6)} \\
x_{ij} &= x_{ij} + v_{ij} \quad \text{......................(7)}
\end{align*}
\]

5.6 Trajectory Step

The output points of moving object in video (summation of motion vector) the result of proposed algorithm.

5.7 Curve fitting approximation of function Step

Consider two commonly used methods for curve fitting, namely interpolation and least squares. For interpolation, we use first polynomials, Then introduce least squares curve fitting using simple polynomials and later generalize this approach sufficiently to permit other choices of least squares fitting functions for the output of tracking operation (trajectory).

Interpolation is a method for constructing a function \(F(x)\) that fits a known set of data points \((x_k, y_k)\), i.e. a method for constructing new data points within the range of a discrete set of known data points. There are various ways in which this can be done. One of the most important aspects of regression analysis is related to the question of what functional relationship \(F(x)\) should be assumed, we must consider the following rules (plot the data and look for obvious trends such as linear, quadric or high order behavior, then see if the data is symmetrical, then consider the periodically, finally break the data set into groups and consider possibly of assuming different functions for data subsets) [12].

Using two types of function as we shown in figure 2:

- **Linear regression (linear function)**
  
  \(F(x) = a_1 + a_2 x \quad \text{......................... (8)}\)

- **Non linear approximation function** can be transformed into linear, we use two types:
  
  \(F(x) = a_1 x + a_2 \quad \text{......................... (9)}\)
  
  \(F(x) = a_1 e^{a_2 x} \quad \text{......................... (10)}\)

*Figure 2. Types of curve fitting function*
5.8 Extract feature Step

Extracting the property for each function such as (domain, scope, slope and intersection points…) to generate feature set for each trajectory.

5.9 Build Database file

Build a database file that contain a profile for each trajectory included the unique index for each one, no of the frame taken into account, the trajectory function, features for each function in order to understand the behavior for each one as we shown in table 1.

<table>
<thead>
<tr>
<th>Path #</th>
<th>No of frames</th>
<th>Trajectory function</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>F(x)=0.368+0.878x</td>
<td>Slope=S Intersection point =I</td>
</tr>
</tbody>
</table>

Finally, the overall proposed work summary as we shown in figure 4 below:

![Figure 4. Block diagram of overall framework](image-url)
6. Proposed tracking algorithm:

Figure 5 below show the steps of proposed tracking algorithm:

- **Frame after apply region growing method**
- **Split the target area into 4 non overlapping area**
- **Feature pool for each region (color histogram, spatiogram, accumulated histogram, 2D shape ratio...)**
- **Apply cat swarm optimization to new search window then find the search region and find appropriate similarity region**
  - Seeking mode
  - Tracing mode
- **Update the search window location by drawing a rectangle using the selected feature points**
- **Find the most outer feature points around the target object**
- **Using one of the similarity measures such as Bhattacharyya measure, chi-square for measure the similarity and one of the cost functions.**
- **Estimation motion vector, then create a trajectory**
- **Build approximation function for each trajectory using curve fitting, using different types of function such as linear & non linear function.**
- **Create some feature for each trajectory function such as slope, intersection points.**

**Figure 5. Block diagram of proposed tracking algorithm**

7. Proposed idea for each object

Using 8 cats for whole object, two cats for each sub region one of them in tracing mode all others in seeking mode in order to select the optimum block, where each region of object assumed it’s direction into two direction using the concept of holistic methods that calculated only global motion as we show in the figure 6 below
8. Conclusion

The objective of this paper is to construct a novel block based framework for video tracking, this method extract the features from vectors (trajectory for each moving object tracking after implement the interpolation operation) we extract two famous features such as slop, intersection point. Also Build a formulation of mathematical models for several points (each set correspond to object motion) in order to evaluate their behavior by fitting data with an approximation function.

In this paper develop new algorithm tracking based on mixed of more than on level of feature extraction classes, The proposed tracking algorithm based on one of swarm intelligent (such as cat algorithm) is proposed to reduce the number of search locations in the BM process. In our proposed algorithm, the computation of search location is reduced by considering a fitness calculation strategy when it estimate new search location.

Since proposed algorithm does not consider any fixed search pattern or any other movement assumption as most as other BM approaches, a high probability for finding accurate motion vector is expected. So that the proposed algorithm reduces the computation cost and enhances the search speed. The Proposed algorithm based on split the target area into four region then compute for each part pool of feature such as (accumulated histogram, color histogram, spatiogram, 2D shape ratio) also develop new method to detect the interest block depending on some of similarity measurement according to some feature for each part of object based on cut swarm optimization to detect the correct direction finally The proposed algorithm accelerated detection the location of moving object.

9. References


