CS 6/79995 ST: Advances in Internet based Systems and Applications
Spring 2001

Object Detection and Tracking in Video

Author:Zhong Guo

Email: zguo@mcs.kent.edu, Homepage: http://www.mcs.kent.edu/~zguo

Prepared for Prof. Javed I. Khan
Department of Computer Science, Kent State University
Date: November 2001

Abstract: This survey paper reviews briefly research works on object detection and tracking in videos. The definition and tasks of object detection and tracking are first described, and the potential applications are mentioned. Followed is the summation of major research highlights and widely used approaches. Reference is included at the end of the paper.

Other Survey's on Internetwork-based Applications
Back to Javed I. Khan's Home Page

Table of Contents

Introduction

Object detection and tracking tasks

Videos are actually sequences of images, each of which called a frame, displayed in fast enough frequency so that human eyes can percept the continuity of its content. It is obvious that all image processing techniques can be applied to individual frames. Besides, the contents of two consecutive frames are usually closely related.
Visual content can be modeled as a hierarchy of abstractions. At the first level are the raw pixels with color or brightness information. Further processing yields features such as edges, corners, lines, curves, and color regions. A higher abstraction layer may combine and interpret these features as objects and their attributes. At the highest level are the human level concepts involving one or more objects and relationships among them.

Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Object tracking is to monitor an object�s spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem, the problem of matching the target region in successive frames of a sequence of images taken at closely-spaced time intervals. These two processes are closely related because tracking usually starts with detecting objects, while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

Applications

Object detecting and tracking has a wide variety of applications in computer vision such as video compression, video surveillance, vision-based control, human-computer interfaces, medical imaging, augmented reality, and robotics. Additionally, it provides input to higher level vision tasks, such as 3D reconstruction and 3D representation. It also plays an important role in video database such as content-based indexing and retrieval.

Challenges

Although has been studied for dozens of years, object detection and tracking remains an open research problem. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how you define the object to be detected and tracked. If only a few visual features, such as a specific color, are used as representation of an object, it is fairly easy to identify all pixels with same color as the object. On the other extremity, the face of a specific person, which full of perceptual details and interfering information such as different poses and illumination, is very hard to be accurately detected, recognized and tracked. Most challenges arise from the image variability of video because video objects generally are moving objects. As an object moves through the field of view of a camera, the images of the object may change dramatically.This variability comes from three principle sources: variation in target pose or target deformations, variation in illumination, and partial or full occlusion of the target [HB98].

There are two sources of information in video that can be used to detect and track objects: visual features (such as color, texture and shape) and motion information. Combination of statistical analysis of visual features and temporal motion information usually lead to more robust approaches. A typical strategy may segment a frame into regions based on color and texture information first, and then merge regions with similar motion vectors subject to certain constraints such as adjacency. A large number of approaches have been proposed in literature. All these efforts focus on several different research areas each deals with one aspect of the object detection and tracking problems or a specific scenario. Most of them use multiple techniques and there are combinations and intersections among different methods. All these make it very difficult to have a uniform classification of existing approaches. So in the following sections, we would review most of the approaches separately in association with different research highlights.

Object Detection and Tracking Approaches

Feature-based object detection

In feature-based object detection, standardization of image features and registration (alignment) of reference points are important. The images may need to be transformed to another space for handling changes in illumination, size and orientation. One or more features are extracted and the objects of interest are modeled in terms of these features. Object detection and recognition then can be transformed into a graph matching problem.

1.Shape-based approaches

Shape-based object detection is one of the hardest problems due to the difficulty of segmenting objects of interest in the images. In order to detect and determine the border of an object, an image may need to be preprocessed. The preprocessing algorithm or filter depends on the application. Different object types such as persons, flowers, and airplanes may require different algorithms. For more complex scenes, noise removal and transformations invariant to scale and rotation may be needed. Once the object is detected and located, its boundary can be found by edge detection and boundary-following algorithms. The detection and shape characterization of the objects becomes more difficult for complex scenes where there are many objects with occlusions and shading. [FBFH94]

[PPS94]

2.Color-based approaches

Unlike many other image features (e.g. shape) color is relatively constant under viewpoint changes and it is easy to be acquired. Although color is not always appropriate as the sole means of detecting and tracking objects, but the low computational cost of the algorithms proposed makes color a desirable feature to exploit when appropriate. [VM97]

[GBT98] developed an algorithm to detect and track vehicles or pedestrians in real-time using color histogram based technique. They created a Gaussian Mixture Model to describe the color distribution within the sequence of images and to segment the image into background and objects. Object occlusion was handled using an occlusion buffer. [FT97] achieved tracking multiple faces in real time at full frame size and rate using color cues.This simple tracking method is based on tracking regions of similar normalized color from frame to frame. These regions are defined within the extent of the object to be tracked with fixed size and relative positions. Each regionis characterized by a color vector computed by sub-sampling the pixels within the region, which represents the averaged color of pixels within this region. They even achieved some degree of robustness to occlusion by explicitly modeling the occlusion process.

Template-based object detection

If a template describing a specific object is available, object detection becomes a process of matching features between the template and the image sequence under analysis. Object detection with an exact match is generally computationally expensive and the quality of matching depends on the details and the degree of precision provided by the object template. There are two types of object template matching, fixed and deformable template matching.[ADGB99]

1. Fixed template matching

Fixed templates are useful when object shapes do not change with respect to the viewing angle of the camera. Two major techniques have been used in fix template matching.

(1)Image subtraction

In this technique, the template position is determined from minimizing the distance function between the template and various positions in the image. Although image subtraction techniques require less computation time than the following correlation techniques, they perform well in restricted environments where imaging conditions, such as image intensity and viewing angles between the template and images containing this template are the same.

(2)Correlation

Matching by correlation utilizes the position of the normalized cross-correlation peak between a template and an image to locate the best match. This technique is generally immune to noise and illumination effects in the images, but suffers from high computational complexity caused by summations over the entire template. Point correlation can reduce the computational complexity to a small set of carefully chosen points for the summations. [KMZ94]

2. Deformable template matching

Deformable template matching approaches are more suitable for cases where objects vary due to rigid and non-rigid deformations. These variations can be caused by either the deformation of the object per se or just by different object pose relative to the camera. Because of the deformable nature of objects in most video, deformable models are more appealing in tracking tasks.

In this approach, a template is represented as a bitmap describing the characteristic contour/edges of an object shape. A probabilistic transformation on the prototype contour is applied to deform the template to fit salient edges in the input image. An objective function with transformation parameters which alter the shape of the template is formulated reflecting the cost of such transformations. The objective function is minimized by iteratively updating the transformation parameters to best match the object [JZL96]. The most important application of deformable template matching techniques is motion detection of objects in video frames which we will review in the following section. [SL01] [ZJD00]

Motion detection

Detecting moving objects, or motion detection, obviously has very important significance in video object detection and tracking. A large proportion of research efforts of object detection and tracking focused on this problem in last decade. Compared with object detection without motion, on one hand, motion detection complicates the object detection problem by adding object�s temporal change requirements, on the other hand, it also provides another information source for detection and tracking.

A large variety of motion detection algorithms have been proposed. They can be classified into the following groups approximately.

1.Thresholding technique over the interframe difference

These approaches [DN90] rely on the detection of temporal changes either at pixel or block level. The difference map is usually binarized using a predefined threshold value to obtain the motion/no-motion classification.

2.Statistical tests constrained to pixelwise independent decisions

These tests assume intrinsically that the detection of temporal changes is equivalent to the motion detection [NSKO94]. However, this assumption is valid when either large displacement appear or the object projections are sufficiently textured, but fails in the case of moving objects that preserve uniform regions. To avoid this limitation, temporal change detection masks and filters have also been considered. The use of these masks improves the efficiency of the change detection algorithms, especially in the case where some a priori knowledge about the size of the moving objects is available, since it can be used to determine the type and the size of the masks. On the other hand, these masks have limited applicability since they cannot provide an invariant change detection model (with respect to size, illumination) and cannot be used without an a priori context-based knowledge.

3.Global energy frameworks

The motion detection problem is formulated to minimize a global objective function and is usually performed using stochastic (Mean-field, Simulated Annealing) or deterministic relaxation algorithms (Iterated Conditional Modes, Highest Confidence First). In that direction, the spatial Markov Random Fields [PT99] have been widely used and motion detection has been considered as a statistical estimation problem. Although this estimation is a very powerful, usually it is very time consuming.

Object tracking using motion information

Motion detection provides useful information for object tracking. Tracking requires extra segmentation of the corresponding motion parameters. There are numerous research efforts dealing with the tracking problem. Existing approaches can be mainly classified into two categories: motion-based and model-based approaches [PD00]. Motion-based approaches rely on robust methods for grouping visual motion consistencies over time. These methods are relatively fast but have considerable difficulties in dealing with non-rigid movements and objects. Model-based approaches also explore the usage of high-level semantics and knowledge of the objects. These methods are more reliable compared to the motion-based ones, but they suffer from high computational costs for complex models due to the need for coping with scaling, translation, rotation, and deformation of the objects.

Tracking is performed through analyze geometrical or region-based properties of the tracked object. Depending on the information source, existing approaches can be classified into boundary-based and region-based approaches.

1.Boundary-based approaches

Also referred to as edge-based, this type of approaches rely on the information provided by the object boundaries. It has been widely adopted in object tracking because the boundary-based features (edges) provide reliable information which does not depend on the motion type, or object shape. Usually, the boundary-based tracking algorithms employ active contour models, like snakes [NP99] and geodesic active contours. These models are energy-based or geometric-based minimization approaches that evolve an initial curve under the influence of external potentials, while it is being constrained by internal energies.

(1) Snakes

Snakes is a deformable active contours used for boundary tracking which was originally introduced by Terzopoulos et al[KWT88]. Snakes moves under the influence of image-intensity �forces,� subject to certain internal deformation constraints. In segmentation and boundary tracking problems, these forces relate to the gradient of image intensity and the positions of image features. One advantage of the force-driven snake model is that it can easily incorporate the dynamics derived from time-varying images. The snakes are usually parameterized and the solution space is constrained to have a predefined shape. So these methods require an accurate initialization step since the initial contour converges iteratively toward the solution of a partial differential equation. [AMTY98] [LC97] [QL98]

Considerable work has been done to overcome the numerical problems associated with the solution of the equations of motion and to improve robustness to image clutter and occlusions. [CB92] proposed a B-spline representation of active contours, [DLJ96] employed polygonal representation in vehicle tracking problems, and [MT93] proposed a deformable superquadric model for modeling of shape and motion of 3D non-rigid objects.

(2) Geodesic active contour models

These models are not parameterized and can be used to track objects that undergo non-rigid motion. In [CC96], a three step approach is proposed which start by detecting the contours of the objects to be tracked. An estimation of the velocity vector field along the detected contours is then performed. At this step, very unstable measurements can be obtained. Following this, a partial differential equation is designed to move the contours to the boundary of the moving objects. These contours are then used as initial estimates of the contours in the next image and the process iterates. More recently, in [BSR99], a front propagation approach that couples two partial differential equations to deal with the problems of object tracking and sequential segmentation was proposed. Additionally, in [GKRR99], a new, efficient numerical implementation of the geodesic active contour model has been proposed which was applied to track objects in movies.

2.Region-based approaches

These approaches rely on information provided by the entire region such as texture and motion-based properties using a motion estimation/segmentation technique. In this case, the estimation of the target's velocity is based on the correspondence between the associated target regions at different time instants. This operation is usually time consuming (a point-to-point correspondence is required within the whole region) and is accelerated by the use of parametric motion models that describe the target motion with a small set of parameters. The use of these models introduces the difficulty of tracking the real object boundaries in cases with non-rigid movements/objects, but increases robustness due to the fact that information provided by the whole region is exploited.

Optical flow [LW00] [PBA00] is one of the widely used methods in this category. In this method, the apparent velocity and direction of every pixel in the frame have to be computed. It is an effective method but time consuming. Background motion model can be calculated using optic flow, which serves to stabilize the image of the background plane. Then, independent motion is detected as either residual flow, the flow in the direction of the image gradient that is not predicted by the background plane motion. Although slightly more costly to compute, this measure has a more direct geometric significance than using background subtraction on a stabilized image. This method is very attractive in detecting and tracking objects in video with moving background or shot by a moving camera.

Utilization of available motion information in video streams

All of the approaches mentioned above have to extract motion information from pixel values of video frames. This is done on raw images or decoded/reconstructed images though certain complicated analysis at pixel level which are extremely time-consuming. Actually, some kind of motion information has already been included in video streams encoded by widely accepted video format standards. For example, the MPEG standard serials adopted a technique called motion estimation/compensation, which calculates the displacements of consecutive matching blocks as motion vectors and encode into standard MPEG streams. In many cases, especially well textured objects, the motion vector values reflect the movement of objects in the scene very well. So instead of having a separate module to extract motion information, some approaches [KGO01][JDD99] utilized these motion vector values directly. Of course this will reduce the accuracy of object boundary because motion vectors are associated only with block of pixels. But this does not result in a major problem in many applications because the goal of tracking is not to determine the exact correspondence for every image location in a pair of images, but rather to determine, in a global sense, the movement of an entire target region over a long sequence of images. The ease of motion information acquisition and highly efficient tracking algorithms at block level are very attractive.

Summary

Along with the increasing popularity of video on internet and versatility of video applications, availability, efficiency of usage and application automation of videos will heavily rely on object detection and tracking in videos. Although so much work has been done, it still seems impossible so far to have a generalized, robust, accurate and real-time approach that will apply to all scenarios. This will require, I believe, combination of multiple complicated methods to cover all of the difficulties, such as noisy background, moving camera or observer, bad shooting conditions, object occlusions, etc. Of course, this will make it even more time consuming. But that does not mean nothing has been achieved. In my opinion, research may go more directions, each targeting on some specific applications. Some reliable assumption can always be made in a specific case, and that will make the object detection and tracking problem much more simplified. More and more specific cases will be conquered, and more and more good application products will appear. As the computing power keeps increasing and network keeps developing, more complex problem may become solvable.

Reference

Research papers

[AMTY98]Shoichi Arakil, Takashi Matsuoka, Haruo Takemura, and Naokazu Yokoya,Real-time Tracking of Multiple Moving Objects in Moving Camera Image Sequences Using Robust Statistics.1051-4651/98, 1998 IEEE.

[BSR99]M. Bertalmio, G. Sapiro, and G. Randall,Morphing Active Contours.Proc. Int'l Conf. Scale-Space Theories in Computer Vision, pp. 46-57, 1999.

[CB92]R. Curwen and A. Blake,Dynamic Contours: Real-Time Active Splines.*Active Visio*n, A. Blake and A. Yuille, eds., pp. 39-58. MIT Press, 1992.

[CC96]V. Caselles and B. Coll,Snakes in Movement.SIAM J. Numerical Analysis, vol. 33, pp. 2,445-2,456,

1996.
[DLJ96]M.P. Dubuisson, S. Lakshmanan, and A.K. Jain,Vehicle Segmentation and Classification using Deformable Templates.IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 3, pp. 293-308, 1996.
[DN90]N. Diehl,Object-Oriented Motion Estimation and Segmentation in Image Sequences.IEEE Trans. Image Processing, vol. 3, pp. 1,901-1,904, Feb. 1990.
[FBFH94]C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, Efficient and Effective Querying by Image Con-tent. J. Intelligent Information Systems, vol. 3, no. 1, pp. 231-262, 1994.
[FT97]Paul Fieguth,Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates.Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97).
[GBT98]T.D.Grove,K.D.Baker,and T. N. TAN, Color based object tracking.14th International Conference on Pattern Recognition (CV41).
[GKRR99]R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky,Fast Geodesic Active Contours.Proc. Int'l Conf. Scale-Space Theories in Computer Vision, pp. 34-45, 1999.
[HB98]Gregory D. Hager and Peter N. Belhumeur,Efficient Region Tracking With Parametric Models of Geometry and Illumination.IEEE transactions on pattern analysis and machine intelligence, vol. 20, no. 10, pp. 1025-39 October 1998.
[JDD99]Ryan C. Jones, Daniel DeMenthon, David S. Doermann,Building mosaics from video using MPEG motion vectors.1999 ACM 1.58113.239~5/99/0010...$5.60.
[JZL96]A.K. Jain, Y. Zhong, and S. Lakshmanan,Object Matching Using Deformable Templates.IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 3, pp. 267�278, Mar. 1996.
[KGO01] Javed I. Khan, Zhong Guo and Wansik Oh, Motion based object tracking in MPEG-2 video stream for perceptual region discrimination rate transcoding.Proceedings of the 9th ACM international conference on multimedia. Pp. 572-576.
[KMZ94]W. Krattenthaler, K.J. Mayer, and M. Zeiller,Point Correlation: A Reduced-Cost Template Matching Technique.*Proc. ICI*P, pp. 208�212, 1994.
[KWT88]M. Kass, A. Witkin, and D. Terzopoulos,Snakes: Active Contour Models. Int'l J. Computer Vision, vol. 1, pp. 321-332, 1988.
[LC97]Yun-Ting Lin and Yuh-Lin Chang,Tracking Deformable Objects with the Active Contour Model.O-8186-7819-4/97 $10.00 0 1997 IEEE.
[LW00]L. Wixson,Detecting Salient Motion by Accumulating Directionally-Consistent Flow.IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 8, August 2000.
[MT93]D. Metaxas and D. Terzopoulos,Shape and Nonrigid Motion Estimation Through Physics-Based Synthesis.IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 15, no. 6, pp. 580-591, 1993.
[NSKO94]H.H. Nagel, G. Socher, H. Kollnig, and M. Otte,Motion Boundary Detection in Image Sequences by Local Stochastic Tests.Proc. European Conf. Computer Vision, vol. II, pp. 305-315, 1994.
[NP99]Natan Peterfreund,Robust Tracking of Position and Velocity With Kalman Snakes.IEEE transactions on pattern analysis and machine intelligence, vol. 21, no. 6, June 1999.
[PBA00]Robert Pless, Tomas Brodsky, and Yiannis Aloimonos,Detecting Independent Motion: The Statistics of Temporal Continuity.IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 8, August 2000
[PD00]Nikos Paragios and Rachid Deriche,Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects.IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 3, pp. 266-280, march 2000.
[PPS94]A. Pentland, R. Picard, and S. Sclaroff,Photobook: Tools for Content-Based Manipulation of Image Databases.Storage and Retrieval of Image and Video Databases II, Paper No. 2185-05, San Jose, Calif., pp. 34-47, SPIE, Feb. 1994.
[PT99]N. Paragios and G. Tziritas,Adaptive Detection and Localization of Moving Objects in Image Sequences.Signal Processing: Image Comm., vol. 14, pp. 277-296, 1999.
[QL98]Lili Qiu, Li Li,Contour Extraction of Moving Objects.
[SL01]Stan Sclaroff and Lifeng Liu,Deformable Shape Detection and Description
via Model-Based Region Grouping.IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 5, pp. 475-489, May 2001.
[VM97]V. V. Vinod and Hiroshi Murase,Video Shot Analysis using Efficient Multiple Object Tracking.O-8186-7819-4/97 $10.00 0 1997 IEEE.
[ZJD00]Yu Zhong, Anil K. Jain, M.-P. Dubuisson-Jolly,Object Tracking Using Deformable Templates.IEEE transactions on pattern analysis and machine intelligence, vol. 22, no. 5, pp544-549, May 2000.

Research groups and links

http://www-iplab.ece.ucsb.edu/Tracking/head.html
http://http.cs.berkeley.edu/projects/vision/
http://atwww.hhi.de/~blick/
http://visual.ipan.sztaki.hu/
http://www.cv.iit.nrc.ca/research.html
http://www-cv.mech.eng.osaka-u.ac.jp/research/tracking_group/tracking.html
http://www.irisa.fr/temics/Demos/Analysis4/

Scope

This survey is based on electronic search in IEEE digital library and their citations using key words �object detection� and �object tracking�. The search was done on November 8, 2001 and around 150 relative references were found from which the above representative papers were chosen.