Universidad del Norte

Master Thesis

Visual Object Tracking applying ensemble of multiple trackers

Author:
Jorge Martinez Gomez

Supervisor:
Juan Carlos Niebles

A thesis submitted in fulfilment of the requirements
for the degree of Master of Science

in the

Computer Vision Research Group
Electrical and Electronics engineering department

February 2015

Declaration of Authorship

I, John Smith, declare that this thesis titled, 'Thesis Title' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- lelele
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism."

Dave Barry

UNIVERSITY NAME (IN BLOCK CAPITALS)

Abstract

Faculty Name
Department or School Name

Doctor of Philosophy

Thesis Title

by John Smith

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

> Not enough people do things that leave others to wonder. RT @BrianMendicino: Wondering why @neiltyson is watching Glee.
>
> Neil deGrasse Tyson

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

Contents

Declaration of Authorship i

Abstract iii

Acknowledgements iv

Contents v

List of Figures vii

List of Tables viii

Abbreviations ix

Physical Constants x

Symbols xi

1 Introduction 1
1.1 Goals . . . 1

2 Moving Object Detection Approaches, Challenges and Object Tracking 3
2.1 Object Representation . . . 3
2.2 Moving Object Detection . . . 3
2.2.1 Background Substraction . . . 4
2.2.2 Temporal differencing . . . 4
2.2.3 Statistical Approaches . . . 5
2.2.4 Point detectors . . . 5
2.2.5 Challenges . . . 5
2.3 Object Tracking . . . 6
2.3.1 Point Tracking . . . 6
2.3.2 Kernel Tracking . . . 7
2.3.3 Silhouette Tracking . . . 8
2.3.4 Tracking applying fusion of trackers or features . . . 9

Bibliography **10**

List of Figures

List of Tables

Abbreviations

LAH **L**ist **A**bbreviations **H**ere

Physical Constants

Speed of Light $c = 2.997\ 924\ 58 \times 10^8\ \mathrm{ms}^{-\mathrm{S}}$ (exact)

Symbols

a	distance	m
P	power	W (Js^{-1})
ω	angular frequency	rads^{-1}

For/Dedicated to/To my...

Chapter 1

Introduction

Visual object tracking is an important problem in computer vision. This field has a wide range of applications such as surveillance,en success building trackers for specific object classes, a generic object tracker represents a challenging task. In general case, tracking an arbitrary object in an unknown scenario is still considered unsolved. Common challenges are for example, object deformations, illumination human-computer interaction and motion analysis. Althought there has be changes, partial and complete occlusions, drifting, background clutter and similar objects in the scene. This particular situations make some trackers better than others. However, there is no algorithm that has mastered all possible problems that a scenario might generate.

Recently, the evaluation performed in [1] shows, each tracking algorithm performs well on different sequences. This explains that different tracking algorithms avoid challenges that can occur in general object tracking. We consider an approach that combines the virtues of different algorithms while evading their weaknesses could outperform each single algorithm. Just as "two heads are better than one", making trackers perform this task together in an unknown scenario, may result in a higher level of performance and achievement, than could be obtained individually. This is what in psychology states as "positive interdependence", the ability of group members to encourage and facilitate each other's efforts [2].

1.1 Goals

In this paper, we focus on the problem of tracking an arbitrary object in videos, with no prior knowledge other than its location in the first frame, also known as "model free tracking". We are motivated to link trackers together so one cannot succeed, unless all

group members succeed. A common tracking system consists of an appearance model, which can evaluate the likelihood of the object of interest at a given location. A motion model, that stores and contains the locations of the object over time. Finally, a search strategy to find the best location in the current frame [3].All methods share the same goal, meaning that each tracker's individual "effort" is required and is indispensable for group success. Using these models, we make trackers correct each other, increasing performance and ensuring the group is united to a common goal, a concrete reason of being, a purpose for existence.

Chapter 2

Moving Object Detection Approaches, Challenges and Object Tracking

2.1 Object Representation

An object can be considered simply as nothing but an entity of interest used for further analysis. These elements can be represented by their shape and appearance. In this section, we describe different object shape and appearance representations employed in tracking.

2.2 Moving Object Detection

In a video, there are two sources of information that can be used for object detection and tracking: Visual features (color, texture and shape) and motion information. Robust approaches suggest that combining the statistical analysis of visual features and temporal analysis of motion information. Moving object detection targets the extraction of moving objects that are of interest in sequences (e.g. people and vehicles).

A large number of methodologies have been proposed for object tracking, focusing on the task of object detection first. Most of them apply combinations and intersections among different methodologies, making it very difficult to create a uniform classification of existing approaches. This section classifies different approaches available for object detection from videos.

2.2.1 Background Substraction

Background subtraction is a commonly used technique for object segmentation in static scenarios [4]. This task consist in detecting moving regions by subtracting the current image pixel-by-pixel from a reference background image. The pixels above some threshold are classified as foreground (belongs to an object). The background image is created averaging images over time in an intiialization period, and is updated with new images to adapt to dynamic scene changes. Also, the foreground map is followed by morphological operations such as closing and erosion (elimination of small-sized blobs).

Although background subtraction techniques extracts well most of the relevant pixels, this method is sensitive to changes when some background and foreground pixels have similar value.

2.2.2 Temporal differencing

In temporal differencing, objects are detected by taking pixel-by-pixel difference of consecutive frames (generally two or three) in a video sequence. This method is most common for moving object detection in scenarios where camera is moving. Unlike static camera scenarios, the background is changing in time for moving camera (not appropiate to create a background model). Alternatively, the moving object is detected by taking the difference between frames $t-1$ and t.

This method is highly adaptive to dynamic changes in the scene as most recent frames are involved in the process. However, it fails detecting small regions as moving objects (ghost regions). Detection will not be correct also, for objects that preserve uniform regions (static objects).

A two-frame differencing method is presented in [5], where the pixels that satisfy the following equation are marked as foreground.

$$|I_t(x,y), I_{t-1}(x,y)| > Th$$

Other methods were developed in order to overcome drastic changes of two frame differencing in some cases. For instance, a three-frame differencing method [6] and a hybrid method that combines three-frame differencing with an adaptive background subtraction model [7].

2.2.3 Statistical Approaches

Statistical characteristics of pixels have been used, in order to overcome shortcomings between frames of basic background subtraction methods. The approaches consist in keeping and updating pixels statistics that belong to the background model. Foreground pixels are identified by comparing each pixel's statistics with that of the background model. These methods are becoming more popular due to its reliability in scenes that contain noise, illumination changes and shadows **Cite here!**.

The statistical method proposed in **Cite here** describes and adaptive background model for real-time tracking. Every pixel is modeles by a mixture of Gaussians which are udpated online using incoming image data. Then, the Gaussians distributions of the mixture model for each pixel is evaluated in order to detect whether a pixel belongs to foreground and background.

2.2.4 Point detectors

Point detectors are used to find interesting points in objects which have an expressive texture in their respective localities. An interest point should have invariance to changes in illumination and camera viewpoint. One important detector uses optical flow approach. These methods make use of the flow vectors of moving objects over time to detect moving blobs in an image. In this approach the apparent velocity and direction of every pixel in the frame must be computed.

2.2.5 Challenges

Object detection and tracking is still an open research problem in computer vision. A robust, accurate and high performance approach is still a great challenge. The level of difficulty depends on how the object of interest is defined in terms of features. For instance, Using color as object representation method, it is not difficult to identify all pixels with same color as the object. However, there is always a probability of existence a background region with same color information (background clutter). In addition, illumination changes in the scene does not guarantee that the pixel values of an object will be the same in all frames. These variabilities or challenges which are random in object tracking causes wrong object tracking, and are listed below.

- **Illumination Changes:** It is desirable that background model adapts to gradual changes of the appearance of the environment.

- **Dynamic background:** Some scenery regions contain movement, but should be still remain as background, acoording to their relevance. Such movement can be periodical or irregular (e.g. traffic lights, waving trees).
- **Occlusion:** Partially or full, occlusion affects the process of computing the background frame. In real life situations, occlusion can occur anytime the object of interest passes behind another object with respect to a camera.
- **Background clutter:** As stated before, this challenge makes the segmentation task difficult. It is hard to create ans separate background model from moving foreground objects.
- **Shadowing:** Shadows cast by foreground objects complicates processes such as background subtraction. Overlapping shadows hinder their separation and classification. Researchers have proposed different methods for detection of shadows.
- **Camera motion:** Sometimes, video may be captured by unstable (e.g. vibrating) cameras.
- **motion:** The speed of a moving object plays an important role in its detection and track. If an object is moving too slow, the temporal differencing methods fails to detect object, because it preserves uniform region between frames. In the other case, fast moving object leaves ghost regions in a detected foreground model.
- **Object rotation and deformation:** Since natural objects move freely, they can appear slightly or completely transformed. Such rotations and transformations in or out of plane on the images affect object tracking considerably.

2.3 Object Tracking

The goal of an object tracker is to generate an object path over time. This trajectory consists of the object position over time in every frame of the video. The tracker may provide complete region in the image that is occupied by the object at every time instant. There are a wide variety of applications of object detection and tracking in computer vision. Certainly, this list is not meticulous and covers popular approaches on each category.

2.3.1 Point Tracking

Tracking can be formulated as the correspondence of objects represented by points across frames. This category can be divided into two subcategories:

- **Deterministic Methods:** These approaches for point correspondence define a cost of associating each object in frame $t-1$ to a single object in frame t using motion constraints, such as proximity, velocity, rigidity and motion. Minimization of the correspondence cost is formulated as a combinatorial optimization problem. A solution, which consists in one-to-one correspondence among all possible associations, can be obtained by optimal assignment methods. For instance Hungarian Algorithm **cite Here!** or greedy search methods.

- **Statistical methods for Point Tracking:** Statistical correspondence methods solve tracking problems whose measurements obtained from video sensors contain nose, or object motion can undergo random perturbations. These approaches take measurements and model uncertainties into account during object state estimation. Applying state space approach to model the object properties such as position, velocity and acceleration. In single object state estimation, the optimal state of an object is giben by the Kalman Filter, assuming measurement noise have a Gaussian distribution. In the general case, that is, object state is not assumed as Gaussian, estimation can be performed using particle filters.

  In the case of multiobject data association, state estimation using Kalman or particle filters, it is necessary to solve first correspondence problem before these filters can be applied. However, in cases when two objects are close each other, the correspondence could be incorrect. Then, an incorrectly associated measurement can cause the filter to fail to converge. In order to tackle this problem, Joint Probability Data Association Filtering (JPDAF) and Multiple Hypothesis Tracking (MHT) are two used techniques for data association.

2.3.2 Kernel Tracking

In this type of tracking, object motion is computed using representations of a primitive object region, from one frame to the next. These algorithms differ in terms of appearance representation (features extraction) used, the number of objects tracked, and the method used for object motion estimation.

- **Density-based tracking:** According to **Cite master thesis**, the object is modelled with one or more probability density functions, such as Gaussian, mixture of Gaussian, Parzen windows or histogras, that describe the probability of object appearance. Mean-shift is an approach to feature space analysis. This method shifts a data point to the average of data points in its neighborhood. Mean shift

uses fixed color distribution. A similar approach is called CAMSHIFT that handles dynamically changing color distribution by adapting the search window size and computing color distribution in the search window.

- **Template-based tracking:** These approaches apply templates of the object to calculate appearance probability on every frame of the video sequence. The most common is *Template matching* that searchs accross the image, a region similar to the object template, defined in previous frames. The similarity measure is calculated using normalized cross correlation. A limitation of this method is its high computational cost due to brute force search. To reduce this cost, some methods limit the object search to a neighborhood near previous position.

  Instead of templates, other object representations can be used for tracking. For example, color histograms or mixture models can be computed using the appearance of pixels inside the rectangular or ellipsoidal regions. To reduce computational complexitu, the similarity between object model and the hypothesized position, is computed evaluating the ratio betweem color means between model and position. The position with highest ratio is selected as current object location.

  "Tracking by detection" or "Tracking by repeated recognition" [8] systems generally perform target object appearance learning. These methods are closely related to object detection (an area with great progress in computer vision) and has encouraged some successful real-time tracking algorithms [9, 10]. However, many tracking algorithms employ static appearance models that are defined manually or trained at the first frame only [11–15], these methods are often unable to deal with significant appearance changes. This situations are difficult when there is limited knowledge of the object of interest. In order to cope this problem, an adaptive appearance model that changes during the tracking process as the appearance of the object changes, gets better results [16–18].

  Boosting has been used in a wide field of machine learning tasks and applied to computer vision problems. Many tracking algorithms are based on the boosting framework [19] and is related to the work on Online Adaboost [20–22], multi-class boost [23] and MILBoost [24]. The goal of boosting is to combine many weak classifiers (usually decision stumps) into a linear strong classifier.

2.3.3 Silhouette Tracking

The object is tracked via estimation of the object region in each frame. Silhouette-based methods provide an accurate shape description for the objects that are tracked. These approaches can be divided into two main categories, shape matching and contour

tracking. Shape matching approaches search object silhouette in the current frame. Contour based, evolve initial contour to its new position in the current frame using state space models or direct minimization of some energy functional.

2.3.4 Tracking applying fusion of trackers or features

Tracking algorithms fusion can be performed actively, that means that each tracker receives feedback; or passively without feedback. The first algorithm that explicitly applies ensemble methods to tracking-by=detection is shown in [20]. The author extended the work of [25] using the Adaboost algorithm to combine a set of weak features and update object model with online update strategy. The authors in [26], combined a template-based tracker, optical flow tracker, and online-random forest tracking-by-detection method into a cascade of trackers. The best selection is summarized into a simple set of rules. Another active fusion approaches are VTD [27] and VTS [28]. In this articles, the trackers are sampled by proposing appearance models, motion models, state representation types, and observation types. Then, the sampled trackers run in parallel and interact with each other. The authors in [29] proposed a classifier ensemble framwork that uses Bayesian estimation theory to estimate the non-stationary distribution of sampled classifers. In [30] and [31], the authors present a passive approach based on the idea of attraction fields. The closer a fusion candidate is to a tracking result box, the stronger it is attracted by it.

Bibliography

[1] Yi Wu, Jongwoo Lim, and Ming Hsuan Yang. Online object tracking: A benchmark. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2411–2418, 2013.

[2] David W. Johnson, Roger T. Johnson, and Karl A. Smith. Cooperative Learning Returns To College What Evidence Is There That It Works?, 1998. ISSN 0009-1383.

[3] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *ACM Computing Surveys*, 38:13, 2006. ISSN 03600300.

[4] Am McIvor. Background subtraction techniques. *Proc. of Image and Vision Computing*, 2:13, 2000.

[5] A J Lipton, H Fujiyoshi, and R S Patil. Moving target classification and tracking from real-time video. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision WACV98 Cat No98EX201*, 98:8–14, 1998. ISSN 09031936.

[6] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36:585–601, 2003. ISSN 00313203.

[7] Robert T Collins, Alan J Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt, and Lambert Wixson. A System for Video Surveillance and Monitoring, 2000. ISSN 19406029.

[8] Greg Mori and Jitendra Malik. Recovering 3D human body configurations using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 28:1052–1062, 2006. ISSN 01628828.

[9] Xiaoming Liu and Ting Yu. Gradient feature selection for online boosting. In *Proceedings of the IEEE International Conference on Computer Vision*, 2007.

[10] Helmut Grabner, Michael Grabner, and Horst Bischof. Real-Time Tracking via On-line Boosting. *Technology*, 1:1–10, 2006. ISSN 0162-8828.

[11] M. Isard and J. MacCormick. BraMBLe: a Bayesian multiple-blob tracker. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, 2, 2001.

[12] Vincent Lepetit and Pascal Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1465–1479, 2006. ISSN 01628828.

[13] Michael J Black and Allan D Jepson. EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation. *International Journal of Computer Vision*, 26:63–84, 1996. ISSN 0920-5691.

[14] D Comaniciu, V Ramesh, and P Meer. Real-time tracking of non-rigid objects using mean shift. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:142–149, 2000. ISSN 01628828.

[15] Amit Adam, Ehud Rivlin, and Ilan Shimshoni. Robust fragments-based tracking using the integral histogram. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 798–805, 2006.

[16] David a. Ross, Jongwoo Lim, Ruei-Sung Lin, and Ming-Hsuan Yang. Incremental Learning for Robust Visual Tracking. *International Journal of Computer Vision*, 77:125–141, 2007. ISSN 0920-5691.

[17] Iain Matthews, Takahiro Ishikawa, and Simon Baker. The template update problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:810–815, 2004. ISSN 01628828.

[18] A D Jepson, D J Fleet, and T F El-Maraghi. Robust online appearance models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1296–1311, 2003. ISSN 0162-8828.

[19] Y Freund and R E Schapire. A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. *Journal of Computing Systems and Science*, 55:119–139, 1997. ISSN 00220000.

[20] Shai Avidan. Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:261–271, 2007. ISSN 01628828.

[21] Helmut Grabner, Christian Leistner, and Horst Bischof. Semi-supervised on-line boosting for robust tracking. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5302 LNCS, pages 234–247, 2008.

[22] NC Oza and S Russell. Online ensemble learning. *AAAI/IAAI*, 6837:1109–1109, 2000.

[23] Amir Saffari, Martin Godec, Thomas Pock, Christian Leistner, and Horst Bischof. Online multi-class LPBoost. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3570–3577, 2010.

[24] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. Visual Tracking with Online Multiple Instance Learning. *IEEE transactions on pattern analysis and machine intelligence*, pages 983–990, 2010. ISSN 1939-3539.

[25] Robert T. Collins, Yanxi Liu, and Marius Leordeanu. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1631–1643, 2005. ISSN 01628828.

[26] Jakob Santner, Christian Leistner, Amir Saffari, Thomas Pock, and Horst Bischof. PROST: Parallel robust online simple tracking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 723–730, 2010.

[27] Junseok Kwon and Kyoung M. Lee. Visual tracking decomposition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1269–1276, 2010.

[28] Junseok Kwon and Kyoung Mu Lee. Tracking by sampling trackers. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1195–1202, 2011.

[29] Qinxun Bai, Zheng Wu, Stan Sclaroff, Margrit Betke, and Camille Monnier. Randomized Ensemble Tracking. In *International Conference on Computer Vision*, pages 2040–2047, 2013.

[30] Christian Bailer, Alain Pagani, and Didier Stricker. A user supported tracking framework for interactive video production. *Proceedings of the 10th European Conference on Visual Media Production - CVMP '13*, pages 1–8, 2013.

[31] Christian Bailer, Alain Pagani, and Didier Stricker. A Superior Tracking Approach: Building a strong Tracker through Fusion. In *European Conference on Computer Vision*, pages 170–185, 2014.