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 **vi**

List of Tables **vii**

Abbreviations **viii**

Physical Constants **ix**

Symbols **x**

1 Introduction **1**
- 1.1 Goals 1

2 Moving Object Detection Approaches, Challenges and Object Tracking **3**
- 2.1 Moving Object Detection 3
  - 2.1.1 Background Substraction 3
  - 2.1.2 Temporal differencing 4
  - 2.1.3 Statistical Approaches 4
  - 2.1.4 Point detectors 5
  - 2.1.5 Challenges 5

Bibliography **7**

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 [**?**] 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 [**?**].

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 [**?**].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 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.

2.1.1 Background Substraction

Background subtraction is a commonly used technique for object segmentation in static scenarios [1]. 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.

2.1.2 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 [2], 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 [3] and a hybrid method that combines three-frame differencing with an adaptive background subtraction model [4].

2.1.3 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.

2.1.4 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.

2.1.5 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.

Bibliography

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

[2] 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.

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

[4] 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.