Omar Arif

Robust Target Localization and Segmentation

Application of Kernel-based statistical methods to computer vision

LAP LAMBERT Academic Publishing

Omar Arif

Robust Target Localization and Segmentation

Application of Kernel-based statistical methods to computer vision

LAP LAMBERT Academic Publishing

Impressum/Imprint (nur für Deutschland/ only for Germany)
Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Coverbild: www.ingimage.com

Verlag: LAP LAMBERT Academic Publishing AG & Co. KG
Dudweiler Landstr. 99, 66123 Saarbrücken, Deutschland
Telefon +49 681 3720-310, Telefax +49 681 3720-3109
Email: info@lap-publishing.com

Herstellung in Deutschland:
Schaltungsdienst Lange o.H.G., Berlin
Books on Demand GmbH, Norderstedt
Reha GmbH, Saarbrücken
Amazon Distribution GmbH, Leipzig
ISBN: 978-3-8433-5038-9

Imprint (only for USA, GB)
Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Cover image: www.ingimage.com

Publisher: LAP LAMBERT Academic Publishing AG & Co. KG
Dudweiler Landstr. 99, 66123 Saarbrücken, Germany
Phone +49 681 3720-310, Fax +49 681 3720-3109
Email: info@lap-publishing.com

Printed in the U.S.A.
Printed in the U.K. by (see last page)
ISBN: 978-3-8433-5038-9

Omar Arif

Robust Target Localization and Segmentation

SUMMARY

This thesis aims to contribute to the area of visual tracking, which is the process of identifying an object of interest through a sequence of successive images. Some of the challenges associated with these tasks are image noise, occlusions, background clutter, complex object shapes, etc.

The work contained in this thesis explores kernel-based statistical methods. These methods map the data to a higher dimensional space where the tasks of classification and clustering are easily carried out. There are two problems related to the mapping: The out-of-sample and the pre-image problem. A pre-image framework for some of the manifold learning and dimensional reduction methods is developed.

Two algorithms are developed for visual tracking that are robust to noise and occlusions. In the first algorithm (Chapter 3), a KPCA-based eigenspace representation is used. The de-noising and clustering capabilities of the KPCA procedure lead to a robust algorithm. This framework is further extended in Chapter 6 to incorporate the background information in an energy based formulation, which is minimized using graph cut. Chapter 7 extends this framework to track multiple objects using a single learned model.

In the second method, a robust density comparison framework is developed (Chapter 5) that is applied to visual tracking (Chapter 8), where an object is tracked by minimizing the distance between a model distribution and given candidate distributions.

The superior performance of kernel-based algorithms comes at a price of increased storage and computational requirements. A novel method is proposed in Chapter 4, that takes advantage of the universal approximation capabilities of generalized radial basis function neural networks to reduce the computational and storage requirements for kernel-based methods. The ideas developed are general and are applicable to other kernel-based methods, such as KCPA and support vector machines.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND BACKGROUND

Computer vision aims to artificially replicate the human visual perceptions. It deals with the science and technology of processes related to the acquisition of images, analysis of images and sequence of images to extract useful information, and to the development of artificial cognitive systems that "see."

Image segmentation and visual tracking are two important components of computer vision. The former deals with an image, while the latter is concerned with the sequence of images. Segmentation aims to partition an image into smaller more meaningful parts, which are related with respect to some common aspect. Visual tracking, on the other hand, is the process of locating an object of interest in a sequence of successive images. For a deformable object, the motion of the object, over a sequence of images, is described by an overall global motion (pose), and the local deformation of the object [94]. In this respect, at each frame of the sequence, segmentation is required to delineate the target from the background. Tracking a deformable object is therefore, the process of estimating the pose parameters (*target localization*) and the local deformation (*target segmentation*) of the object. In this sense, visual tracking encompasses segmentation. Algorithms differ in whether they perform localization only or both i.e. localization and segmentation. The former can be called *transformation based* and the latter *contour based*. Objects can be represented by their appearances such as color, texture, edges etc, which provide characteristic information about the target object. This characteristic information is encoded into a cost or similarity functional. Tracking algorithms then find the target object that is optimum with respect to a pre-determined similarity functional.

Visual tracking is a challenging task. In some cases, the image information (appearance) may not be sufficient enough to identify the target object. This may happen due to a number of reasons, such as noise, occlusions, complex object shapes and presence of target

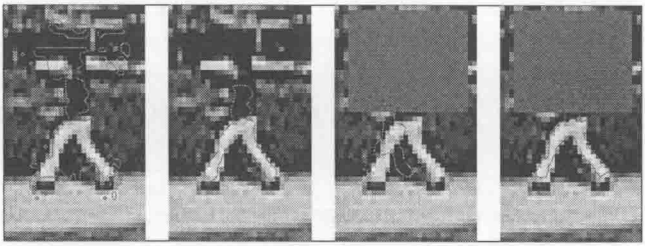

Figure 1: Segmentation: without shape prior, with shape prior, shape prior and occlusion, shape prior and occlusion with manual localization (from left to right).

features in the background etc. In such cases, visual tracking can be simplified by imposing a shape/spatial constraint on the tracking process. These constraints can come in the form of smoothness priors on the contour that segments the object from the background, or in the form of incorporating prior known information on the shape of the object to be segmented. For example, in the first image in Figure 1, the target object is segmented without any shape prior. The segmentation is noisy, due to the presence of the target features (color) in the background. The second image produces a correct segmentation by incorporating shape priors. However, the addition of shape information requires extra effort on the part of tracker. The prior shapes need to be aligned with the pose/location of the object, which is being tracked, for correct segmentation. This is not trivial since the pose is not known and has to be estimated along with the segmentation. If the pose can not be adequately estimated, then the tracking results are meaningless. The effect of incorrect target localization on target segmentation is shown in Figure 1-third image. In the last image, the target is localized manually, which results in correct segmentation. So correct target localization is critical to shape based trackers.

Apart from imposing constraints on the shape of the object, tracking can also be simplified by imposing constraints on object motion. The object motion can be constrained to be constant velocity or constant acceleration based on prior information [110]. Figure 2 shows an example of successful tracking using a constant velocity motion constraint. These methodologies form the basis of well known tracking algorithms like Kalman filter

(a) Without motion constraint

(b) With constant velocity constraint

Figure 2: Tracking with and without Kalman filtering

[50] and particle filter [44]. These techniques have been used in tracking algorithms such as [18, 28, 49, 101]. Filtering techniques will not be pursued in this research.

The computation complexity of visual tracking algorithms is an important issue. Tracking algorithms should be able to work in real time.

This thesis aims to contribute to the above mentioned challenges and requirements for visual tracking. It does this by developing novel algorithms for target localization and segmentation, which are robust to noise and massive occlusions. The algorithms are based on kernel-based statistical methods. Also, a novel method is provided to reduce the computation complexity of the algorithms.

Below, target segmentation and localization, as carried out by different methods, is explained briefly. We will be mostly concerned with how shape information is incorporated and used to perform localization and segmentation.

1.1 Target Segmentation

The aim of target segmentation is to delineate the object from its background. Several algorithms and techniques have been developed for segmentation. The more recent ones are discussed below.

1.1.1 Active Contour based

The basic idea in active contour based segmentation algorithms is to evolve a closed contour, starting from an initial estimate, such that it encircles the object region. Evolution of the contour is governed by an energy functional, which defines the fitness of the contour to the hypothesized object region [110]. The energy functional is minimized when the contour delineates the object from the background. The energy functional can be based on local information such as the image gradient [52, 54], global features such as color [41, 109, 114], or mean image intensity inside and outside of the evolving contour [25]. Level set methods [89] have been successfully used for implicit representation of the contour. The most important advantage of level set representation of the contour is its flexibility in allowing topology changes. There is a vast body of literature concerning level sets and active contour, see for example [68, 69, 70, 78].

More recently, shape information is incorporated into the active contour framework to make the segmentation robust to noise and occlusion [33, 34, 35, 40, 60, 79, 98]. Leventon et al. [60] define the shape term E_{shape} using a probabilistic approach based on principal component analysis (PCA). The set of training shapes are represented by their signed distance functions [89], and PCA is applied to obtain a reduced representation. A probability density function is defined over the parameters of the reduced representation to obtain the shape energy. The level set function is evolved using both the image and the shape terms, which draw the level set function towards the most probable shape according to the learned distribution. Tsai et al. [98] incorporate the shape model, derived also by performing PCA on a collection of signed distance maps of the training shapes, into region-based active contour ([25]). The problem is reformulated to directly optimize the parameters associated with the first few eigenvectors. Cremers et al. [33] use kernel density estimator to define the probability density on the space of signed distance function representing the prior shapes. They show that this approach captures nonlinear shape variability. Freedman et al. [40] track by combining density matching and shape priors. For density matching, Bhattacharyya measure is used to define the distance between a model intensity distribution and the intensity distribution of an estimated image region. The tracker is expressed as a PDE-based curve

evolution, which is implemented using level sets. Dambreville et al. [35] combine intensity based segmentation with prior shape knowledge learned using Kernel PCA (KPCA). A binary representation of shapes is used. KPCA is shown to outperform linear PCA, by allowing only shapes that are close enough to the training data. Dynamical shape priors have also been used to improve the tracking of deformable objects in the presence of noise and occlusions. [30, 77].

1.1.2 Graph cut

Image segmentation can also be formulated as a graph partitioning problem [23]. Let $\mathcal{R}$ be the set of all pixels in the image, and let $\mathcal{L} = \{0, 1\}$ be a label assignments on $\mathcal{R}$. The label 1 means the pixel belongs to the target, while the label 0 means it belongs to the background. The segmentation problem is cast as that of finding a labeling $l : \mathcal{R} \rightarrow \mathcal{L}$, minimizing an energy $E(l)$, modeled by:

$$E(l) = E_d(\mathcal{I}, l) + E_s(l). \tag{1}$$

E_d is the data term which measures how well the labeled pixels fit the image model. The standard data term is

$$E_d(\mathcal{I}, l) = \sum_{u \in \mathcal{R}} F_u(l_u), \tag{2}$$

where F_u measures how well label l_u fits pixel u. To realize the energy on the graph, each pixel is considered a node of the graph with two additional terminal nodes, the target and the background terminal nodes. To encode the data term on the graph, each pixel is connected to the target and background terminal nodes with edge weights $F_u(1)$ and $F_u(0)$, representing the cost of assigning a pixel to target and background respectively. E_s is the regularization or boundary term. Let $\mathcal{N}$ be a neighborhood system on $\mathcal{R}$, then E_s is realized by connecting each pair of pixels $(u, v) \in \mathcal{N}$ with a non-negative edge weight measuring the penalty for assigning two neighboring pixels to different regions. The mincut of the graph represents the segmentation that best separates the target from its background and minimizes the energy $E(l)$.

Prior shape information can be encoded in the graph cut by either imposing it on the edges between pixels and the terminals nodes [63, 105], or by defining the edge weights