Image and Vision Computing 27 (2009) 1157–1168

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A level set framework using a new incremental, robust Active Shape Model for object segmentation and tracking

Michael Fussenegger [a], Peter Roth [b], Horst Bischof [b], Rachid Deriche [c], Axel Pinz [a,*]

[a] Institute of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Kopernikusgasse 24, A-8010 Graz, Austria
[b] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
[c] Odyssée Project Team INRIA Sophia Antipolis – Méditerranée, France

ARTICLE INFO

Article history:
Received 20 November 2006
Received in revised form 30 September 2008
Accepted 28 October 2008

Keywords:
Level set
Segmentation
Tracking
Active Shape Model
Incremental robust PCA

ABSTRACT

Level set based approaches are widely used for image segmentation and object tracking. As these methods are usually driven by low level cues such as intensity, colour, texture, and motion they are not sufficient for many problems. To improve the segmentation and tracking results, shape priors were introduced into level set based approaches. Shape priors are generated by presenting many views a priori, but in many applications this a priori information is not available. In this paper, we present a level set based segmentation and tracking method that builds the shape model incrementally from new aspects obtained by segmentation or tracking. In addition, in order to tolerate errors during the segmentation process, we present a robust Active Shape Model, which provides a robust shape prior in each level set iteration step. For the tracking, we use a simple decision function to maintain the desired topology for multiple regions. We can even handle full occlusions and objects, which are temporarily hidden in containers by combining the decision function and our shape model. Our experiments demonstrate the improvement of the level set based segmentation and tracking using an Active Shape Model and the advantages of our incremental, robust method over standard approaches.

1. Introduction

In the past 20 years, various level set approaches for image segmentation and tracking have been proposed. The level set method for capturing moving fronts was introduced by Osher and Sethian [25] with some of the key ideas already proposed in Dervieux and Thomasset [12,13]. The two main advantages of this approach are that the implicit boundary representation does not depend on a specific parameterisation during the propagation and that the embedding function can model topological changes of the contour such as splitting and merging. Based on this Caselles et al. [4,35,31] introduced geometric deformable models. Paragios and Deriche [26] use coupled geodesic active regions and Chan and Vese [5] as well as Tsai et al. [38] proposed a level set implementation of the Mumford–Shah functional [24] for image segmentation.

Level set methods have been successfully applied to tracking tasks. An early work on region based level set tracking proposed by Bertalmio et al. [1] is based on morphing images. Paragios and Deriche [27] use a geodesic model, that combines motion and edge information. Using the difference between the current image and the reference background, a region based model was proposed by Besson et al. [2]. In Mansouri [23] and Yilmaz et al. [41] feature distributions of the object and the background are used for tracking. Freedman and Zhang [14] track a predefined distribution for the object region by minimising a Kullback–Leibler or Bhattacharyya distance. But all of these approaches are restricted to one level set function and can track only one region. Shi and Karl [36] propose a new fast level set implementation that can handle multiple regions, but in contrast to other approaches no shape prior information is used.

To increase accuracy several approaches were proposed, that successfully introduced the usage of prior shape knowledge for level set based segmentation and tracking. Leventon et al. [22] use a Gaussian model to describe their shape priors. They assume a uniform distribution over pose parameters that model translation and rotation. Rousson and Paragios [34] suggested to introduce shape information on the variational level. But as Leventon et al. [22] they can handle only one shape prior and unfamiliar image structures are ignored. Cremers et al. [11] presented an approach with dynamic labelling, that allows them to use more than one shape prior and does not suppress unfamiliar image structures, but all shape priors are assigned to one level set function. Riklin-Raviv et al. [32] present a novel approach that can handle a projective transformation of the shape prior, but their approach is also limited to one region. Furthermore, the projective transformation is

* Corresponding author. Tel.: +43 316 873 5021.
E-mail addresses: mfussenegger@iconmobile.com (M. Fussenegger), pmroth@icg.tugraz.at (P. Roth), bischof@icg.tugraz.at (H. Bischof), Rachid.Deriche@sophia.inria.fr (R. Deriche), axel.pinz@tugraz.at (A. Pinz).

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2008.10.014

computationally too expensive for tracking applications. Recently, Fussenegger et al. [15,16] proposed a framework for segmentation and tracking of an arbitrary number of objects with multiple shape priors and Cremers [9] introduced dynamical statistical shape priors for level set based tracking. For a more detailed review on the use of the level set segmentation, we refer to the recent review in Cremers et al. [10].

For all these approaches the shape models are learnt in advance and detached from the segmentation and tracking process. A reasonable training set has to be prepared in order to build the shape model before it can be used for segmentation or tracking of complex objects. Adding new images to an existing model usually results in a recomputation of the whole shape model. In addition, these shape models are not robust, but especially while tracking we might receive data that is corrupted by partial occlusions or imperfect segmentation.

To overcome these drawbacks we apply a robust, incremental model [17] to estimate the shape priors. The model is based on the idea of Active Shape Models (ASM) [7,8]. For an ASM, it is assumed that the shape model can be generated by a linear combination of implicit representations, which is realised by using Principal Component Analysis (PCA) [19]. We extended this approach by applying a robust and incremental PCA algorithm [37] to learn the ASM. Thus, in contrast to existing approaches, we need only a small data set representing the objects' shapes to initialise the ASM and we can cope with corrupted segmentations.

The outline of the paper is as follows: Section 2 gives an overview of the proposed framework consisting of the level set segmentation/tracking and the incremental, robust ASM. In Sections 3 and 4, we introduce our level set methods for multi region segmentation and tracking. Next, Section 5 describes the incremental, robust ASM. In Section 6, we show experimental results demonstrating the properties of our framework. We compare our incremental, robust ASM with the standard ASM and show its capabilities for segmentation and tracking tasks. In addition, we demonstrate the interaction and the capabilities of each individual part of our framework. Finally, conclusions are drawn in Section 7.

2. System overview and level set formulation

Fig. 1 depicts our proposed framework, which combines three components: (i) the segmentation module [15], (ii) the tracking module [16], and (iii) the ASM module [17].

In a first step, the ASM module is initialised with one or more training images of non corrupted, aligned shapes characterising different aspects of the objects of interest. For each object, an eigenspace is estimated from the corresponding aligned shapes represented as binary images. Since this representation can be updated later on, compared to existing methods, in this step only few training samples are necessary. These ASMs are then used to generate the shape priors in the subsequent segmentation process, which can be summarised as follows:

After each level set segmentation the level set representation of the current segmentation is passed from the segmentation module

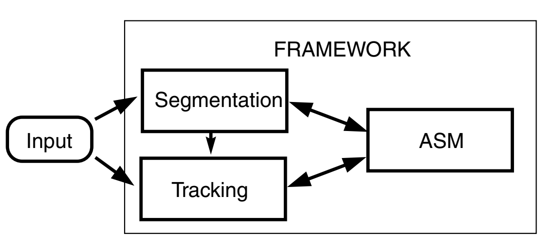

Fig. 1. Proposed level set segmentation and tracking framework using an ASM: the ASM model robustly provides shape priors for the segmentation/tracking process and the obtained results are directly used to retrain the ASM.

to the ASM module. In the ASM module, the level set function is first registered by mapping it onto the mean shape using a similarity transformation, where the mean shape is calculated over all already learnt shapes. Next, the thus aligned level set function is robustly reconstructed from the corresponding eigenspace (Section 5). Finally, this reconstruction is passed to the segmentation module and is used as a shape prior in the next level set iteration step. In this way, we can handle imperfect segmentations considerably better than other approaches. This is repeated until the segmentation process ends. The final result can be used to update the ASM. Thus, we obtain an incrementally better ASM, that improves the segmentation results, which in turn improves the ASM.

When the framework is used for tracking of objects, the processing is very similar to the one described for segmentation, but partial or full occlusions are handled in a different manner.[1] Furthermore, the thus obtained tracked shapes are used as training shapes for the incremental update of the ASM.

For such a framework, several level set formulations (e.g., [25,5,38,28]) may be applicable. But in particular, we use the level set formulation proposed by Paragios and Deriche [28,29] to minimise the energy for an object region:

$$E_D(\Phi, p_1, p_2) = -\underbrace{\int_\Omega (H(\Phi)\log p_1 + (1 - H(\Phi))\log p_2) d\boldsymbol{x}}_{\text{region homogeneity}} + \nu \underbrace{\int_\Omega |\nabla H(\Phi)| d\boldsymbol{x}}_{\text{contour length}}, \qquad (1)$$

where $\Phi : \Omega \to \mathbb{R}$ is the level set function with $\Phi(\boldsymbol{x}) > 0$ if $\boldsymbol{x} \in \Omega_1$ and $\Phi(\boldsymbol{x}) < 0$ if $\boldsymbol{x} \in \Omega_2$, $H(\Phi)$ is the regularised Heaviside function, and p_1 and p_2 are the probability density functions, that describe the probability that a pixel $\boldsymbol{x}$ belongs to region Ω_1 or to region Ω_2, which cover the whole image domain Ω without overlap. The first term describes the homogeneity in a region,[2] whereas the second term takes into account the weighted length of the contour. For noisy images a higher value of the weighting parameter ν provides a better segmentation result since the length of the contour becomes more important and very small region parts are suppressed. This original level set formulation equation (1) by Paragios and Deriche [28,29] models exactly one background and one foreground (object) region.

In general, segmentation results can be improved by using a shape prior. To demonstrate the importance of having a good shape prior, Fig. 2 shows the segmentation results when using different shape priors, that were estimated using an ASM. For Fig. 2(b), the segmentation is done without an ASM. In this case the segmentation fails completely. In Fig. 2(c), we show a shape prior that was obtained from 40 training shapes, and the corresponding segmentations. The segmentation has been improved significantly but there are still some errors present. Finally, Fig. 2(d) shows that the segmentation result can further be improved by increasing the number of shapes to train the ASM. In fact, 80 shapes were used to train the model for the shape prior. Thus, it is clear that a considerable number of shapes is necessary a priori to finally obtain an ASM of sufficient accuracy.

To add a shape prior to the level set formalism, Rousson and Paragios [34] extend the energy function in Eq. (1) straight forwardly:

$$E(\Phi, \Phi_S, p_1, p_2) = E_D(\Phi, p_1, p_2) + \gamma E_S(\Phi, \Phi_S) \qquad (2)$$

with

$$E_S(\Phi, \Phi_s) = \int_\Omega (\Phi - \Phi_S)^2 d\boldsymbol{x}, \qquad (3)$$

[1] A pixel in the input image may belong to more than one shape region, which is not true for the segmentation process – cf. the max criterion in Eq. (11).

[2] Our homogeneity measure depends on the obtained image data and is modelled as multivariate Gaussian density: $p_i = p(\boldsymbol{x}|\Omega_i) = \frac{1}{(2\pi)^{N/2}|\Sigma_i|}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right]$, with $N = 1$ for grey value and $N = 3$ for colour images.