Available online at www.sciencedirect.com
SCIENCE DIRECT®
Image and Vision Computing 24 (2006) 1218–1232
www.elsevier.com/locate/imavis

A comparison between feature-based and EM-based contour tracking

Arthur E.C. Pece [a,b,*], Anthony D. Worrall [c]

[a] Heimdall Vision, Bjørnsonsvej 29, 2500 Valby, Denmark
[b] Department of Computer Science, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark
[c] School of Systems Engineering, University of Reading, RG6 6AY Reading, UK

Received 7 April 2003; received in revised form 6 June 2005; accepted 7 June 2005

Abstract

Most active-contour methods are based either on maximizing the image contrast under the contour or on minimizing the sum of squared distances between contour and image ‘features’. The Marginalized Likelihood Ratio (MLR) contour model uses a contrast-based measure of goodness-of-fit for the contour and thus falls into the first class. The point of departure from previous models consists in marginalizing this contrast measure over unmodelled shape variations.

The MLR model naturally leads to the EM Contour algorithm, in which pose optimization is carried out by iterated least-squares, as in feature-based contour methods. The difference with respect to other feature-based algorithms is that the EM Contour algorithm minimizes squared distances from Bayes least-squares (marginalized) estimates of contour locations, rather than from ‘strongest features’ in the neighborhood of the contour. Within the framework of the MLR model, alternatives to the EM algorithm can also be derived: one of these alternatives is the empirical-information method.

Tracking experiments demonstrate the robustness of pose estimates given by the MLR model, and support the theoretical expectation that the EM Contour algorithm is more robust than either feature-based methods or the empirical-information method.

Keywords: Generative model; Active contour; EM algorithm; Kalman filter; Empirical information matrix

1. Introduction

Active contour methods find application to tracking when camera motion prevents the use of background subtraction methods, and/or when only specific kinds of objects need to be tracked, and the shape, but not the appearance, of these objects is known a priori.

Most active-contour methods can be classified as *feature-based* if the pose of the object is optimized by minimizing squared distances between contours and image features; and *contrast-based* if the pose of the object is optimized by maximizing the norm of the image gradient (or some related measure) underlying the contour. A recent overview of active-contour methods can be found in [13].

Feature-based methods have found wide application in tracking [2,3,17,35,37], but feature extraction is a process notoriously sensitive to noise, which leads to instabilities in tracking.

Contrast-based methods include the original snake model [20], the model-based tracker by Kollnig and Nagel [21], and several methods based on image statistics [12,38,34]. These methods have the disadvantage of not taking explicitly into account unmodelled variations of the contour shape. In addition, gradient-based optimization of the object pose/shape is not easily applicable to the underlying model (although it has been shown [21] that the Levenberg–Marquardt method can be applied).

The MLR (Marginalized Likelihood Ratio) contour model [29,31] is contrast-based, but allows for random, unmodelled shape variations. The basic assumptions of the model are as follows:

(1) grey-level values of nearby pixels are correlated if both pixels belong to the object being tracked, or both belong to the background, but there are no correlations if the two pixels are on opposite sides of an object boundary;
(2) the shape of the contour is subject to random local variability.

* Corresponding author. Address: Department of Computer Science, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen, Denmark.
E-mail address: aecp@heimdall-vision.com (A.E.C. Pece).

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2005.06.013

The first assumption leads to the use of a likelihood ratio as the objective function to be maximized in pose/shape refinement. The same principle is adopted in [12,38,34]. The second assumption implies that the likelihood ratio must be marginalized over possible deformations of the contour. A similar principle is adopted in [23]. Taking the two assumptions together means that no features need to be detected and matched to the model, leading to greater robustness against noise; while at the same time local shape variations are taken explicitly into account. As in many scenarios involving marginalization, the EM algorithm [24] is a natural choice for optimization of the marginalized likelihood ratio. Other modified Newton methods are also applicable.

The MLR model and the EM Contour algorithm have been developed in previous publications, as applied to vehicle tracking with 3D models [31,28,6,5] or eye tracking with 2D models [14–16]. The general MLR model is discussed more extensively in [29]. This paper contains a shorter introduction to the model (in the specific form described in the Appendix of [29]) and focuses on developing Newton-like methods for pose optimization based on the MLR model. The three methods considered in this paper are the EM Contour algorithm, an empirical-information method [25], and a feature-based contour algorithm. Both theoretical and experimental comparisons between these methods are presented.

No performance comparison can conclusively prove the superiority of an entire class of methods: the experiments with the empirical-information method are meant primarily to show that the EM algorithm is not the only feasible optimizer for the MLR model; the implementation of the feature-based algorithm is meant primarily to illustrate the similarities in practice between the EM Contour algorithm and the feature-based approach. Taken together with the theory developed in Section 3, however, the performance comparisons suggest that tracking with the EM Contour algorithm is more robust, while the computational cost is approximately equal for all algorithms.

Section 2 describes the MLR contour model. Section 3 derives modified Newton methods for pose refinement based on the MLR model. Section 4 extends the algorithms from pose refinement to tracking (Kalman filtering). Section 5 describes the results obtained by tracking motor vehicles. Finally, Section 6 contains a discussion of these comparisons.

2. Likelihood model

The object state is described by an m-vector $\mathbf{x}(t)$ which is a function of time t. Given the state and a geometrical model of the object (which can be a 2D model or a 3D model), the object contour is projected onto the image plane. The contour is then used to estimate the likelihood of the image, given the object state.

2.1. The observation

A finite set of n sample points on the contour are used to estimate the likelihood. The image coordinates and unit normals to these sample points are computed from the geometric model together with the estimated state parameters. The normal line to a sample point will be called *observation line*. Due to the aperture problem, only the normal component of the displacement of the object boundary can be locally detected. Therefore, only the intersection between the object boundary and the observation line is of interest in the pose refinement algorithms. A distinction must also be made between the predicted intersection (i.e. the contour intersection) and the actual intersection: these differ not only because of errors in the state estimate, but also because of errors in the geometric model.

In the following, the symbol ν_i will be used for the coordinate on the observation line indexed by i. The symbol μ_i will be used for the coordinate of the contour intersection. The distance between contour and actual intersection is denoted by ϵ_i. The actual intersection is of course unknown: in general it is only possible to estimate a pdf (probability density function) over values of ϵ_i. The subscript i will be dropped when not needed.

The grey-level profile on observation line i will be denoted by $I_i(\nu)$. Using a digital computer, only a finite set of grey levels can be measured on the observation line. Given regularly-spaced sampling of grey levels (with spacing $\Delta\nu$ and bilinear interpolation) we define the *observation* as $\boldsymbol{I}_i = \{I_i(j\,\Delta\nu)|j \in \mathbb{Z}\}$. In the following, the subscript j will always denote location on the observation line. Fig. 1 illustrates the meaning of the symbols μ, ν, ϵ, $\Delta\nu$.

2.2. Likelihoods of grey-level differences

In this paper, we consider a specific form of the MLR model: the general model is described in [29] and the specific model described here is analyzed in better detail in the Appendix of [29]. In the specific model, it is assumed that

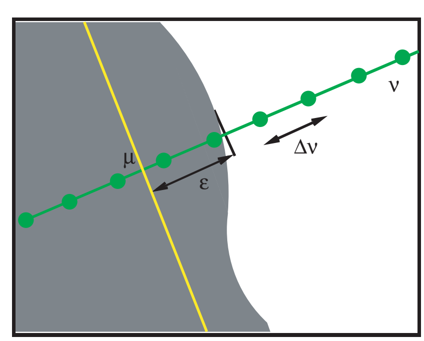

Fig. 1. The yellow line represents a model contour, mismatched with the object boundary, either because of misalignment or because of shape variability. The green line represents the normal to a sample point on the contour. ν is the coordinate on the normal line, μ is the coordinate of the intersection with the contour, and ϵ is the distance between the intersections of the normal line with the contour and with the object boundary. Grey levels are sampled on the normal with a regular spacing $\Delta\nu$. The useful range of sampling of grey levels is determined by the width σ of the Gaussian window (see text). (For interpretation of the reference to colour in this legend, the reader is referred to the web version of this article.)