7

MOTION BASED IMAGE SEGMENTATION

7.1 Introduction

In his book on human visual perception The perception of the visual world [1], 1950, J. J. Gibson studied the flow of projected surfaces which he called optical flow. He characterized optical flow by speed and direction at each point of the retinal image domain. Nakayama and Loomis [2] referred to optical flow later as the instantaneous velocity field over the image positional domain and provided an analytic working definition of optical velocity, the instantaneous velocity vector. In this book, we will retain the essence of Gibson’s and Nakayama’s definition and characterization which view optical flow as the field of optical velocities over the image domain. Taking image in its most common mathematical sense to mean the result of an application, a projection on the retinal domain in this case, we will also refer to optical flow as image motion or, when there is no ambiguity, simply motion as in the title of this chapter.

Image motion occurs whenever there is a relative motion of the viewing system and the viewed environmental surfaces. The processing of this ever-present motion by the human visual system plays several functions which result in a rich, effortless perceptual ability [2]. In machine vision, motion analysis has several essential roles as well [3, 4]. The fundamental roles are (motion) detection, to separate the image domain into two regions, one corresponding to the moving scene objects and the other to the background; (motion) segmentation to partition the image domain into regions of differing motion properties; three-dimensional interpretation to explain the image in terms of the structure and the motion of the moving real objects in the viewed environment; and tracking to locate and follow objects in their movement throughout an image sequence.

In this chapter, we will be concerned with motion based image segmentation, i.e., image domain partitioning using optical flow. Motion-based segmentation is of broad interest in vision because it serves various useful applications such as video coding [5], video indexing and retrieval [6, 7] video event understanding [4, 8, 9], tracking [10, 11, 12, 13], and three-dimensional interpretation [3, 14, 15, 16].
Optical flow is not a sensed variable; it must be estimated. One can estimate it, using a variational method [17], for instance, and then segment it, using, for instance, the Chan-Vese level set method [18] described in a preceding chapter. The issue, in this case, is to compute a boundary preserving estimate of optical flow via proper regularization [17, 19]. The Appendix describes two prevalent image motion estimation methods.

An advantageous alternative to this two-step processing is joint estimation and segmentation because estimation and segmentation are in reciprocal dependence. We will describe two such methods, both of which can be seen as applications to optical flow of the Mumford-Shah image segmentation functional [20] which, we recall, is

$$E(g, K) = \int_{\Omega} (g - g_0)^2 dx + \rho \int_{\Omega \setminus K} \|\nabla g\|^2 dx + \lambda \rho l(K) \quad (7.1)$$

We recall that $E$ is the ‘energy’, or ‘merit’, of the partition determined by the image approximation $g$ with boundaries $K$, $g_0$ is the sensed input image, $\rho$ (scale) and $\lambda$ (weight) are positive real constants, and $l$ is the length of $K$.

The transposition of (7.1) to optical flow will bring in a term of conformity of image motion to data, namely, the image spatial and temporal variations; a term to embed the desired properties of the flow, smoothness, for instance; and a term related to the length of the flow boundaries to produce regular boundaries according to a definition, albeit implicit, of a flow edge.

The two methods we are about to describe differ in their representation of the flow boundaries. One [21] uses the discrete Leclerc minimum description length (MDL) formulation [22]. It references the segmentation individual edge points and, as such, can be viewed as an implicit edge detection method. The other is a level set method, using closed curves to delineate (optical flow) segmentation regions, each described by the parameters of a general linear model of motion [23].

Before detailing these methods, we review briefly others which addressed joint optical flow estimation and segmentation via functional optimization [24, 25, 26, 27]. A set of label variables over the image domain describes the segmentation in [24]. The problem is cast in a Bayesian parametric motion estimation formulation which, under assumptions, is transformed into an energy function containing the classic quadratic terms of smoothness of flow and conformity of flow to data, in addition to two other terms related to the segmentation, one referencing the flow and the other the labels of the segmentation, both terms derived under a Markov/Gibbs field model. Three parameters control the relative contribution of the four terms. Optimization is carried out deterministically to ease the computational burden.

The objective functional in [27] contains five terms. Two terms are related to parametric estimation of the flow field under an affine approximation model. One is a measure of conformity of the flow to data and the other is a smoothness prior. Both terms use a statistically robust function rather than the quadratic function as in [28], and are given a semi-quadratic form via the introduction of two sets of auxiliary variables over the image domain, one set for each term. The other terms are related to the segmentation. One term is to favor low values along the segmentation boundaries of the smoothness related auxiliary variable. Another term is to enforce conformity