Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 29 (2008) 852–861
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Generalizing the Lucas–Kanade algorithm for histogram-based tracking

David Schreiber *

Smart System Division, Austrian Research Centers GmbH – ARC, Donau-City-Strasse 1, A-1220 Vienna, Austria

Received 27 July 2007
Available online 18 January 2008

Communicated by B. Kamgar-Parsi

Abstract

We present a novel histogram-based tracking algorithm, which is a generalization of the template matching Lucas–Kanade algorithm (and in particular of the inverse compositional version which is more efficient). The algorithm does not make use of any spatial kernel. Instead, the dependency of the histogram on the warping parameters is introduced via a feature kernel. This fact helps us to overcome several limitations of kernel-based methods. The target is represented by a collection of patch-based histograms, thus retaining spatial information. A robust statistics scheme assigns weights to the different patches, rendering the algorithm robust to partial occlusions and appearance changes. We present the algorithm for 1-D histograms (e.g. gray-scale), however extending the algorithm to handle higher dimensional histograms (e.g. color) is straightforward. Our method applies to any warping transformation that forms a group, and to any smooth feature. It has the same asymptotic complexity as the original inverse compositional template matching algorithm. We present experimental results which demonstrate the robustness of our algorithm, using only gray-scale histograms.

Keywords: Template tracking; The Lucas–Kanade algorithm; Histogram-based tracking; Kernel-based tracking; Robust least squares; Pedestrian tracking

1. Introduction

Tracking can be defined simply as follows: given a current frame of a video and the location of an object in the previous frame, find its location in the current frame. The three main categories into which most algorithms fall are feature-based tracking (e.g. Beymer et al., 1997; Smith and Brady, 1995), contour-based tracking (e.g. Isard and Blake, 1998; Yokoyama and Poggio, 2005), and region-based tracking. In the last category, the region's content is used either directly (template matching (e.g. Lucas and Kanade, 1981), or is represented by a non-parametric description such as a histogram (e.g. Perez et al., 2002; Adam et al. (2006)), and most notably, kernel-based tracking using the mean shift algorithm (e.g. Bradski, 1998; Comaniciu et al. (2003)).

Kernel-based methods track an object region represented by a spatially weighted intensity histogram. An object function that compares target and candidate kernel densities is formulated using the Bhattacharya measure, and tracking is achieved by optimizing this objective function using the iterative mean shift algorithm (e.g. Comaniciu et al., 2003). However, first introduced kernel-based approaches were restricted to visual tracking problems involving only location (Comaniciu et al., 2003) and location and scale (Collins, 2003). Additional limitations of these methods include: (1) slow convergence rate of the mean shift optimization; (2) the interaction between the spatial structure of the kernel and of the image. This causes, for example, Hager et al. (2004) and Fan et al. (2005) to carefully choose a small number of multiple

* Tel.: +43 (0) 50550 4282; fax: +43 (0) 50550 4150.
E-mail address: david.schreiber@arcs.ac.at

0167-8655/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.12.014

kernels for each specific sequence they process; (3) loss of spatial information, and (4) inability to handle occlusions due to the global nature of the template model.

These limitations were addressed by many consequent works. For example, Hager et al. (2004) and Georgescu and Meer (2004) replace the mean-shift optimization of the Bhattacharyya measure by Newton-style iterations of an SSD-like measure (the Matusita metric) and allow for more general transformations than translation and scale. In (Hager et al., 2004) it is shown that the latter optimization method is more efficient than the former, and makes fewer assumptions on the form of the underlying kernel structure. However, as the binning function of the histogram in (Hager et al., 2004) is a binary function, the analytical solution of the SSD minimization is cumbersome and requires the introduction of a large sifting matrix (number of pixels in the region times the number of bins in the histogram). In (Georgescu and Meer, 2004), a non-binary feature kernel is used, and minimization is done using an iterative weighted least-squares. Furthermore, to enhance the localization accuracy, optical flow-based registration is employed too, and both estimations are combined into a single estimation process using the sum of the two. However, both Georgescu and Meer (2004) and Hager et al. (2004) are still making use of a global spatial kernel.

The issue of the loss of spatial information has been treated in various forms. In (Elgammal et al., 2003), a joint feature-spatial distribution is used which takes into account both the intensities and their position in the window, where a local spatial domain kernel rather than a global one is used. The tracking is achieved by maximizing likelihood using mean shift. In (Yang et al., 2005), a new similarity measure for this joint distribution is introduced, which is the expectation of the density estimates over the model or target image. To alleviate the quadratic complexity, the improved fast Gauss transform is used. The new similarity measure allows the man shift algorithm to track more general motion models. In (Birchfield and Rangarajan, 2005), the notion of *spatiograms* has been introduced, by adding the spatial mean and covariance of the pixel positions to the histogram and then employing mean shift for spatiograms. In (Zhao and Tao, 2005), *correlogram* is used and the mean shift algorithm is extended to 3D (location and orientation of the correlogram).

A different approach to histogram-based tracking was introduced by Perez et al. (2002). Based on color histogram distances, color likelihood is build and then coupled with a dynamical state space model. The resulting posterior distribution is sequentially approximated with a particle filter. This probabilistic approach is further extended to patch-based color modeling to incorporate the spatial layout.

Recently, a new histogram-based tracking approach called *Frag-Track* (Adam et al., 2006), was introduced, which is not based on an optimization scheme but rather on an exhaustive search (over translations and scales) which is made efficient by using integral histograms. The template object is represented by multiple histograms of multiple rectangular patches of the template. Every patch votes on the possible positions and scales of the object in the current frame, by comparing its histogram with the corresponding image patch histogram. Next, a robust statistic measure is minimized in order to combine the vote maps of the multiple patches. The advantages of this method over optimization-based techniques, is that first it allows the use of any metric for comparing two histograms, and not just analytically tractable ones, and second, that it is less likely to be stuck on a local minima. However, as the method relies on the use of integral histograms, the number of bins used is limited and on colour images the method can become quite memory-consuming. Limited accuracy is another issue. The search over different positions and scales is discrete, not to mention sub-pixel accuracy associated with continuous optimisation schemes. In addition, the method is limited to a transformation consisting of translations and scale.

Template tracking, on the other hand, dates back to Lucas and Kanade (1981). The goal of the Lucas–Kanade algorithm is to minimize the sum of squared error between the template and a new image warped back onto the coordinate frame of the template (Baker and Matthews, 2004). The minimization is performed with respect to the warping parameters. Due to its non-linear nature, optimization is done iteratively solving for increments to the already known warping parameters. In particular, the *inverse compositional algorithm* (Baker and Matthews, 2004) is a more efficient version of the algorithm, where the roles of the template and the image are switched and as a result, the Hessian need not be updated each iteration.

To handle partial occlusions, appearance variations and presence of background pixels, robust versions of the template matching algorithm were proposed (e.g. Hager and Belhumeur, 1998; Ishikawa et al. (2002)). The goal of the robust algorithms is to use a weighted least-squares process, such that occluded regions, background pixels and regions where brightness have changed would be considered as outliers and would be suppressed. In practice, robust algorithms require a trade-off between efficiency and accuracy. Namely, in (Hager and Belhumeur, 1998), the Hessian matrix depends on outliers, while in (Ishikawa et al., 2002), the template is divided into patches, assuming a constant weight for each patch.

In this paper we introduce a novel optimization-based algorithm for histogram-based tracking which is a generalization of the Lucas–Kanade algorithm. In particular, we formulate it as an inverse compositional algorithm which is more efficient (Baker and Matthews, 2004). We remove altogether any spatial kernel, external or local, from our histogram definition, and introduce the warping parameters directly into a feature kernel. This fact helps us to overcome the limitations of kernel-based methods mentioned above. A fast convergence rate of the optimization is achieved by using a Gauss–Newton gradient descent; by avoiding the use of a spatial kernel altogether, the structure of the kernel becomes irrelevant; spatial information is kept