Int J Comput Vis (2012) 98:279–302
DOI 10.1007/s11263-011-0509-0

Globally Optimal Estimation of Nonrigid Image Distortion

Yuandong Tian · Srinivasa G. Narasimhan

Received: 30 July 2010 / Accepted: 17 November 2011 / Published online: 21 December 2011

Abstract Image alignment in the presence of non-rigid distortions is a challenging task. Typically, this involves estimating the parameters of a dense deformation field that warps a distorted image back to its undistorted template. Generative approaches based on parameter optimization such as Lucas-Kanade can get trapped within local minima. On the other hand, discriminative approaches like nearest-neighbor require a large number of training samples that grows exponentially with respect to the dimension of the parameter space, and polynomially with the desired accuracy $1/\epsilon$. In this work, we develop a novel data-driven iterative algorithm that combines the best of both generative and discriminative approaches. For this, we introduce the notion of a "pull-back" operation that enables us to predict the parameters of the test image using training samples that are not in its neighborhood (not ϵ-close) in the parameter space. We prove that our algorithm converges to the global optimum using a significantly lower number of training samples that grows only *logarithmically* with the desired accuracy. We analyze the behavior of our algorithm extensively using synthetic data and demonstrate successful results on experiments with complex deformations due to water and clothing.

Keywords Computer vision · Motion and tracking · Image alignment · Image registration · Nonrigid deformation · Feature correspondence · Distortion estimation · Global optimum · Iterative approach · Combine generative and discriminative approaches · Physics-based vision · Water distortion

Y. Tian (✉) · S.G. Narasimhan
The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: yuandong@cs.cmu.edu
url: http://www.cs.cmu.edu/~ILIM/projects/IM/globalopt/research_globalopt.html
url: http://www.cs.cmu.edu/~ILIM/

S.G. Narasimhan
e-mail: srinivas@cs.cmu.edu

1 Introduction

Images that capture non-rigid deformations of objects such as water, clothing and human bodies, exhibit complex distortions (Fig. 1). Aligning or registering such images despite the distortions is an important goal in computer vision that has implications for tracking and motion understanding, object recognition, OCR and medical image analysis. Typically, given a distorted image $I_{\mathbf{p}}$ (e.g., of a scene observed through an undulating water surface) and its template T (the scene observed when the water is still), the task is to estimate the parameters $\mathbf{p}$ of a distortion model that warps the image back to the template.[1]

Most techniques for non-rigid image alignment can be classified into three broad categories, i.e., feature-based, generative and discriminative approaches. *Firstly*, feature matching techniques aim to match a set of sparse local features in the distorted image with those in the template (Lowe 2004; Ling and Jacobs 2005; Pilet et al. 2008). Then, the parameters of a distortion model are estimated from the matchings. These methods work well when the dimension d of the parameter space is low (e.g., 6 for affine), but often fail in the presence of repetitive textures or high dimensional non-rigid distortions. *Secondly*, template matching techniques obtain dense correspondence between a distorted image and

[1] Other works (Tian and Narasimhan 2009; Learned-Miller 2006; Efros et al. 2004) use a set of distorted images or videos as the input and compute distortions and/or the template.