Neurocomputing 73 (2010) 3179–3190

Contents lists available at ScienceDirect

Neurocomputing

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

Robust object tracking using a spatial pyramid heat kernel structural information representation

Xi Li [a,*,1], Weiming Hu [a], Hanzi Wang [b], Zhongfei Zhang [c]

[a] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[b] University of Adelaide, Australia
[c] State University of New York, Binghamton, NY 13902, USA

ARTICLE INFO

Article history:
Received 6 February 2010
Received in revised form
18 May 2010
Accepted 6 June 2010
Communicated by Xiaofei He
Available online 3 August 2010

Keywords:
Object tracking
Visual surveillance
Appearance modeling

ABSTRACT

In this paper, we propose an object tracking framework based on a spatial pyramid heat kernel structural information representation. In the tracking framework, we take advantage of heat kernel structural information (HKSI) matrices to represent object appearance, because HKSI matrices perform well in characterizing the edge flow (or structural) information on the object appearance graph. To further capture the multi-level spatial layout information of the HKSI matrices, a spatial pyramid division strategy is adopted. Then, multi-scale HKSI subspace analysis is applied to each spatial pyramid grid at different levels. As a result, several grid-specific HKSI subspace models are obtained and updated by the incremental PCA algorithm. Based on the multi-scale grid-specific HKSI subspace models, we propose a tracking framework using a particle filter to propagate sample distributions over time. Theoretical analysis and experimental evaluations demonstrate the effectiveness of the proposed tracking framework.

1. Introduction

For visual tracking, handling appearance variations of an object is a fundamental and challenging task. Much work has been done in the domain of modeling object appearance variations.

Hager and Belhumeur [1] propose a tracking algorithm which uses an extended gradient-based optical flow method to handle object tracking under varying illumination conditions. However, the problem with this algorithm is the sensitivity to partial occlusion and noise. In [2], curves or splines are exploited to represent the appearance of an object to develop the Condensation algorithm for contour tracking. Due to the simplistic representation scheme, the algorithm is sensitive to the pose or illumination change, resulting in tracking failures under a varying lighting condition. Zhao et al. [3] present a fast differential EMD tracking method (DEMD) which is robust to illumination changes. But DEMD only considers the issue of color distribution matching without modeling the appearance changes and capturing the structural information of an object. Silveira and Malis [4] present a new algorithm for handling generic lighting changes. Yet, the algorithm performs poorly in capturing the object's shape information which is important to object tracking in complex scenarios.

Black et al. [5] employ a mixture model to represent and recover the appearance changes in consecutive frames. Jepson et al. [6] develop a more elaborate mixture model with an online EM algorithm to explicitly model appearance changes during tracking. Zhou et al. [7] embed appearance-adaptive models into a particle filter to achieve a robust visual tracking. Wang et al. [8] present an adaptive appearance model based on the Gaussian mixture model (GMM) in a joint spatial-color space (referred to as SMOG). SMOG captures rich spatial layout and color information. Yilmaz [9] proposes an object tracking algorithm based on the asymmetric kernel mean shift with adaptively varying the scale and orientation of the kernel. Nguyen et al. [10] propose a kernel-based tracking approach using maximum likelihood estimation. Yu et al. [11] propose a spatial-appearance model which captures non-rigid appearance variations and recovers all motion parameters efficiently. Li et al. [12] use a generalized geometric transform to handle the deformation, articulation, and occlusion of appearance. Ilic and Fua [13] present a non-linear beam model for tracking large deformations. Tran and Davis [14] propose robust regional affine invariant image features for visual tracking. Grabner et al. [15] develop a keypoint matching-based tracking method by online learning classifier-based keypoint descriptions. The common problem with the above tracking methods is that they perform poorly in characterizing both local and global interactions among pixels, which is crucial for robust visual tracking under bad conditions.

Lee and Kriegman [16] present an online learning algorithm to incrementally learn a generic appearance model for video-based

* Corresponding author.
E-mail addresses: lixichinanlpr@gmail.com (X. Li), wmhu@nlpr.ia.ac.cn (W. Hu), Hanzi.Wang@ieee.org (H. Wang), zhongfei@cs.binghamton.edu (Z. Zhang).
[1] The author has moved to CNRS, TELECOM ParisTech, France.

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.017

recognition and tracking. The limitation of this algorithm is to rely heavily on a generic prior model, without which visual tracking cannot be implemented. Lim et al. [17] present a human tracking framework using robust system dynamics identification and nonlinear dimension reduction techniques. The limitation is that its computational cost is expensive. Black et al. [18] present a subspace learning based tracking algorithm with the subspace constancy assumption. A pre-trained, view-based eigenbasis representation is used for modeling appearance variations. However, the algorithm does not work well in the cluttered scene with a large lighting change due to the subspace constancy assumption. Ho et al. [19] present a visual tracking algorithm based on linear subspace learning. In order to make subspace learning more efficient, two incremental PCA algorithms are proposed in [20,21], respectively. Limy et al. [22] propose a generalized tracking framework based on the incremental image-as-vector subspace learning methods with a sample mean update. The common limitation in [19,20,22] is the ignorance of the spatial layout information of object appearance. To address this problem, Li et al. [23] present a visual tracking framework based on online tensor decomposition. The framework relies on image-as-matrix techniques for considering the spatial layout information. Besides, Porikli et al. [24] utilize several covariance matrices of image features to capture the spatial correlation information of object appearance. They utilize the affine-invariant Riemannian metric to make some basic statistics on the covariance matrices. Similarly, Li et al. [25,26] propose two appearance models under the Log-Euclidean Riemannian metric. In these two models, the Log-Euclidean covariance matrices of image features are used to represent object appearance. In this way, the local self-correlation information of object appearance is taken into account. Wu et al. [29] propose a tracking approach which is capable of incrementally learning a low-dimensional covariance tensor representation. However, the tracking methods [23–26,29] have a problem that their appearance models lack a competent object description criterion that captures the intrinsic structural properties of object appearance. Babenko et al. [30] present a tracking system based on online multiple instance learning. This system is able to update the appearance model with a set of image patches, which do not need to precisely capture the object of interest. But the limitation of this system is to use Haar-like image features for object representation, which is sensitive to complex appearance variations.

In this paper, we propose a tracking framework based on heat kernel structural information (HKSI) matrices. The HKSI matrices essentially reflect the edge flow (or structural) information of the object appearance graph as heat diffusion time progresses. Using the edge flow (or structural) information, the intrinsic properties of object appearance changes can be precisely captured. The main contribution of the tracking framework is summarized as follows. First, an object is represented as an object appearance graph, where a series of multi-scale heat kernel structural information (HKSI) matrices are extracted. These HKSI matrices are capable of capturing the edge flow (or structural) information of object appearance. Second, the spatial pyramid division mechanism is adopted for characterizing the multi-level spatial layout information of HKSI matrices. Third, a grid-specific HKSI subspace model for each pyramid grid is learned online by the incremental PCA algorithm [22]. Subsequently, the grid-specific HKSI subspace model, which serves as an observation model, is incorporated into a particle filter for tracking. Moreover, a novel criterion for the likelihood evaluation,

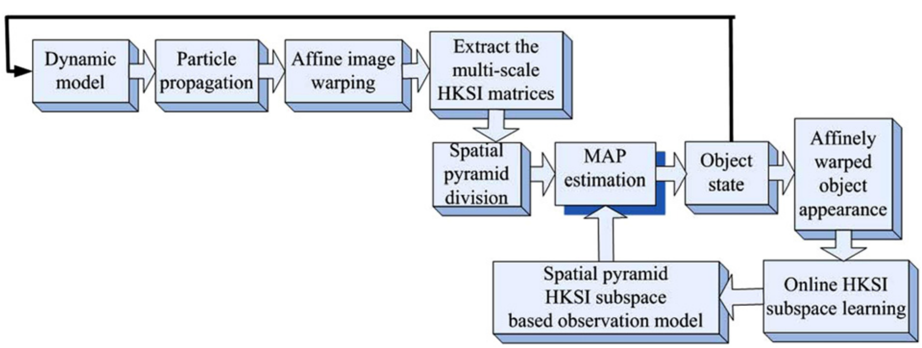

Fig. 1. The architecture of the tracking framework.

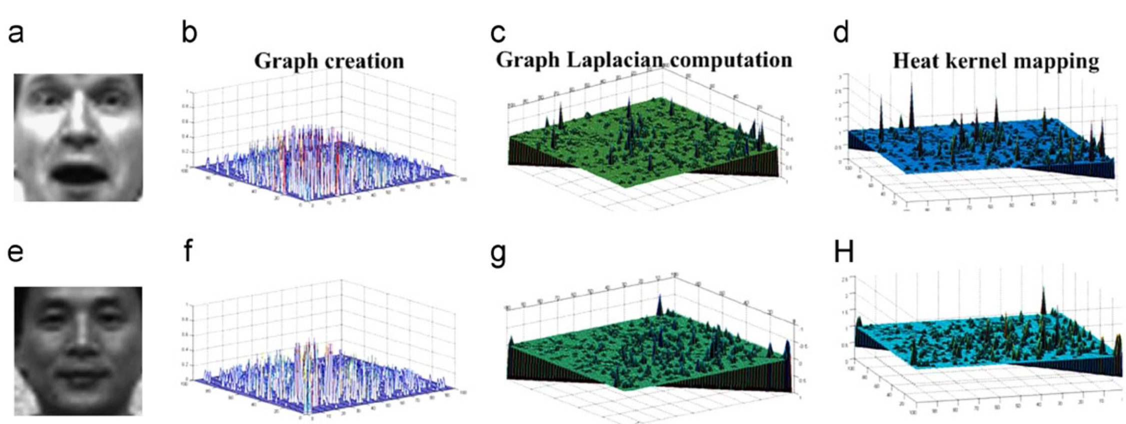

Fig. 2. Example of constructing the scale-0.1 heat kernel matrix. (a) and (e) show two different face images; (b) and (f) plot the corresponding edge-weight matrices in the 3D space; (c) and (g) display the corresponding normalized graph Laplacian matrices in the 3D space; (d) and (h) exhibit the corresponding scale-0.1 heat kernel matrices in the 3D space.