Journal of Network and Computer Applications 35 (2012) 1740–1748

Contents lists available at SciVerse ScienceDirect

Journal of Network and Computer Applications

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

Using incremental subspace and contour template for object tracking

Jihao Yin [a], Chongyang Fu [a], Jiankun Hu [b,*]

[a] School of Astronautics, Beihang University, Beijing 100191, China

[b] School of Engineering and Information Technology, University College, The University of New South Wales, Australian Defence Force Academy, Canberra ACT, NSW 2600, Australia

ARTICLE INFO

Article history:
Received 29 September 2011
Received in revised form
8 June 2012
Accepted 18 June 2012
Available online 26 June 2012

Keywords:
Contour model
Object tracking
Incremental learning
Affine transform

ABSTRACT

Object tracking in the presence of appearance variation and occlusion is a hot topic in research, many algorithms were proposed in recent years. Early contour tracking algorithms used particle filter in a high dimensional space. In practice, contour points can move independently, hence contour deformation forms a high dimensional deformation space. As a result, the application of particle filter is calculation expensive. In this paper, we address the problem of tracking contour in complex environments by involving subspace and a contour template. Specifically, our algorithm tracks the global motion and the local contour deformation separately. We track the global motion by weighted distance to subspace, which is adaptive to the complex environment variation by incremental learning, and then use contour model to track local deformation and evolve the contour to the edge points. The experimental results show that our method can track object contour undergoing partially occlusion and shape deforming, which verify the effectiveness of the proposed algorithm.

1. Introduction

Visual tracking is still very important in actual video-sequence processing (Akhriev and Kim, 2003), especially in network monitoring. The technology has been used for supervision and human–computer interaction (González-Ortega et al., 2010). Due to the deformable nature of objects in most of the tracking applications, contour tracking are appealing in tracking tasks. The contour tracking deals with non-rigid nature of real objects, which results in difficult trade off problem. Early algorithms on contour tracking used a fixed parametric contour model for representation such as B-spline. When multiple knots are used, the vector at each point consists of blending coefficients which is B-spline basis function appropriate to the order of the curve and its set of knots (Blake et al., 1994). Blake and Isard introduced particle filter to track object contour, used curves or splines to represent the boundary and developed the condensation algorithm (Blake and Isard, 1998). They used B-spline representation and the particle filter, and tracked objects in affine transform space, but could not track local deformations (Vaswani et al., 2006). Dynamic snake representation has emerged as a powerful algorithm for contour tracking (Kass et al., 1988). Kass proposed active contour model or snakes to evolve a curve subjecting to constraints from a given image (Chan and Vese, 2001). Osher represented the contour as the zero level set of a higher dimensional function, usually referred as level set function, and evolved initial contour until it minimized an image-based energy functional (Li et al., 2005). The simplicity and efficiency of this approach is affected by cluttered background. The contour model does not exactly follow the object of interest and may be trapped by a strong occasional edge, but too "soft" contour frequently follows noisy edge instead of actual one (Zhong et al., 2000). Rathi implemented particle filter in an infinite dimensional state space and involved geometric active curve to evolve for segmentation using Chan–Vese model. The algorithm can track contour deformation, but needs the average pixel values of target and occlusion object as prior knowledge (Rathi et al., 2007). Namrata proposed PF-MT (Particle Filter Mode Tracking) algorithm, which uses B-spline basis function to present the contour deformation velocity, presented the contour by level set and tracked the deformation in a subsampled effective basis space (Vaswani et al., 2010). The algorithm can track contour deformation in multi-model, but the calculation is time consuming.

Early algorithms on contour tracking track global transform and local deformation in a high dimensional space simultaneously by using particle filter, but the application of particle filter is calculation expensive (Arulampalam et al., 2002). Some researchers have reported contour tracking with global motion constraint (Yin et al., 2010). Most global motion tracking algorithms are able to work in stationary environments; however, these algorithms usually fail in the presence of appearance variation and occlusion (Cheng, 1995). Incremental learning is widely used to adapt to the additional samples variation recently. Ross used appearance model to track objects by incremental learning a low dimensional subspace representation (Ross et al., 2008). Wang used incremental learning to two

* Corresponding author. Tel.: +61 3 99259793; fax: +61 2 62688580.
E-mail address: J.Hu@adfa.edu.au (J. Hu).

1084-8045/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2012.06.005

dimensional linear discriminant analysis (Wang et al., 2010). Therefore, the subspace incremental learning (SIL) algorithm is efficiently adapting online to the changes of the target appearance. The SIL algorithm is a well known method to track objects in non-stationary environment. However, the rectangular window inevitably contains some background pixels, the SIL algorithm treat background and target pixel equally. So it is lack of the spatial information due to the flattening (Hu et al., 2011). In order to address this problem, we proposed weighted subspace incremental learning (WSIL) to treat target and background pixels.

The main research problem of this paper is to estimate the object contour with the aim of achieving good tracking accuracy as well as the tracking rate in the complex environment, such as undergoing partially occlusion and shape deforming. We assume that the global object moves as affine transform in a larger scale space, and the local shape varies in a multi-dimensional but smaller scale space. The full space of contour deformation is theoretically infinite, and it is impractical to track contour in large dimensional space using particle filter. Therefore, we descript an efficient way to approximate the contour deformation and global motion as follows. The proposed algorithm of tracking object can be separated into two steps: (1) tracking the global motion. We use weighted SIL to track affine transformation in a six dimensional space, and (2) tracking local deformations. We propose contour model with local features (CMLF), which describes object contour, to track local deformation. In this paper, we combine the weighted SIL algorithm with shape information CMLF, and propose WSIL–CMLF (weighted subspace incremental learning and contour model with local features) algorithm. The conducted experiments show that our proposed algorithm WSIL–CMLF tracks the object contour more accurately and quickly in the complex environment with the partial occlusion and appearance variation. The WSIL–CMLF algorithm can adapt to occlusion, lighting variation and deformation, therefore it performs well when target undergoes shape deformation and environmental changes.

The outline of the remainder of the paper is as follows. We describe the proposed model and relevant details of the proposed algorithm in Section 2. Section 3 presents results on real video sequences, and compares these results with other methods, followed by conclusion in Section 4.

2. WSIL–CMLF algorithm

We define the object contour state (global transformation X_l and local deformation X_c of object contour C) is denoted by X, the contour is represented by K sub-sampled points. The observation (image I_i and edge-map I_e) is denoted by I, the observation likelihood is:

$$p(I|X) = p(I_e|X_c)p(I_i|X_l) \tag{1}$$

The $p(I_i|X_l)$ is modeled by Gaussian distribution in six dimensional affine transformation space, and the $p(I_e|X_c)$ is subjected to contour feature constraints. The algorithm estimates X_l by WSIL, and estimates X_c by CMLF.

We detailed the proposed algorithm and contrast the differences between this method and prior methods. The advantages of this algorithm are discussed in related context.

2.1. Global motion and WSIL

Subspace incremental learning is well known for locating the global motion of target and performs well in the presence of object appearance variations (Ross et al., 2008). We estimate the global motion by weighted subspace incremental learning to differentiate the weight of target object from background.

The global motion tracking problem is regarded as an inference problem in a Markov model with hidden state variables. At each time step t, we observe an image patch I_t in the tracking window, and the location of the tracking window l_t, which is treated as an unobserved state variable (Ross et al., 2004). The location of the tracking window in an image frame can be represented by an affine transformation of six parameters $l_t=(x_t, y_t, \theta_t, s_t, \alpha_t, \varphi_t)$, representing its center position (x_t,y_t), rotation angle θ_t, scale s_t, aspect ratio α_t, and skew direction φ_t at time t. This transformation warps the image coordinate system, centering the target within a unit square box.

Given l_{t-1}, the global motion is modeled by a prior distribution $p(l_t|l_{t-1})$, indicating the probability of the object appearing at l_t. Particles representing possible target window locations at time t are sampled according to $p(l_t|l_{t-1})$. Each parameter in l_t is modeled by a Gaussian distribution around its counterpart in l_{t-1}.

$$\begin{aligned} &p(Y_i|X_t) = p(l_t|l_{t-1}) \\ &p(l_t|l_{t-1}) = \mathcal{N}(l_t; l_{t-1}, \boldsymbol{\psi}), \end{aligned} \tag{2}$$

where $\boldsymbol{\psi}$ is a diagonal covariance matrix including the corresponding variances of affine parameters, i.e., $\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_s^2, \sigma_\alpha^2, \sigma_\phi^2$.

Given $\boldsymbol{I}_t$, we model the likelihood of l_t with the distribution $p(\boldsymbol{I}_t|l_t)$. We introduce weight matrix $\boldsymbol{w}$ to the likelihood computation, and we denote weighted subspace incremental learning by WSIL. Specifically, we estimate the likelihood of the sample location l_t by negative exponential of weighted squared distance between $\boldsymbol{I}_t$ and subspace $\boldsymbol{U}$. Subspace $\boldsymbol{U}$ is used for representation of $\{I_1, I_2, \ldots, I_t\}$, and is updated by incremental PCA algorithm. The likelihood derivation process is as follows:

$$\begin{aligned} \ln p(I_t|l_t) &= \ln \mathcal{N}(I_t; \mu, UU^T + \varepsilon I) \\ &\propto -\left\| w\left((I_t-\mu) - UU^T(I_t-\mu)\right)\right\|^2 \\ &= -\|wd\|^2 \\ &= -\sum_{k=1}^{n^2} (w(k)d(k))^2 \\ &= -\sum_{k=1}^{n_c} (w_{in}d_{in}(k))^2 - \sum_{k=1}^{n^2-n_c} (w_{out}d_{out}(k))^2, \end{aligned} \tag{3}$$

$$w = \mathrm{diag}\{w(1), w(2), \ldots, w(n^2)\}_{n^2\times n^2}, \tag{4}$$

where $\boldsymbol{d}=(\boldsymbol{I}_t-\mu)-\boldsymbol{U}\boldsymbol{U}^T(\boldsymbol{I}_t-\mu)$, $\boldsymbol{w}$ is the weight matrix of $\boldsymbol{I}_t$, the diagonal elements of $\boldsymbol{w}$ are the weight of $\boldsymbol{I}_t$ elements, n is the width of unit square, μ is sample image mean, $\boldsymbol{I}$ is an identity matrix, the $\varepsilon\boldsymbol{I}$ term corresponds to the noise, n_c is the number of pixels in contour. The element of $\boldsymbol{w}$ is set to $\boldsymbol{w}_{in}$(0.5–0.8) if the point is inside the contour, whereas the element is set to $\boldsymbol{w}_{out}$(1.0), and the contour is depicted by CMLF. The same pixel in $\boldsymbol{I}_t$ has different probability value depending on whether in or out of the contour, where the inside probability is larger than outside. It is apparent that the WSIL algorithm changes the impact of target and background to improve the subspace algorithm.

Using Bayes' rule to incorporate our observation with prior distribution, we conclude that the most probable a posterior object location is at the maximum $p(l_t|I_t, l_{t-1})$. We select the sample location l_t^* with the largest posterior to be our approximate of target window location at time t.

$$\begin{aligned} &p(l_t|I_t, l_{t-1}) \propto p(I_t|l_t)p(l_t|l_{t-1}), \\ &l_t^* = \arg\max p(l_t|I_t, l_{t-1}^*). \end{aligned} \tag{5}$$

SIL algorithm is adaptive to environment changes, but the appearance model does not distinguish background and target.

The flattened vector in subspace method is lack of spatial correlations between neighbor pixels, whereas contour model