Image and Vision Computing 26 (2008) 1530–1539

Contents lists available at ScienceDirect

Image and Vision Computing

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

Bilateral learning for color-based tracking

Ying Ren, Chin Seng Chua *

School of Electrical and Electronic Engineering, Nanyang Technological University, Block S1, Nanyang Avenue, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 6 October 2005
Received in revised form 28 March 2008
Accepted 24 April 2008

Keywords:
Color-based tracking
Color model adaptation
Bilateral Learning (BL)

ABSTRACT

This paper addresses the issue of color model adaptation and color-based object tracking in a dynamic scene. Under different environmental conditions such as illumination changes, a static color model is inadequate for the purpose of color-based object detection and tracking. Color model adaptation is required and this has to be integrated into the tracking procedure within the spatial domain. To track a target in both the color and spatial domains, a *Bilateral Learning* (BL) approach is proposed in this paper. Formulated as an unsupervised learning problem, the adaptations of the color and spatial models are fitted into an EM framework by updating in the color and image spaces alternately. This results in the adaptation of the color model and the localization of the target along the image sequence. Experimental results show the effectiveness and efficacy of the proposed approach for color model adaptation and object tracking under illumination changes and environmental noise in real time.

1. Introduction

Efficient and robust visual tracking of objects is a prerequisite for automatic visual surveillance systems [7,2,19], Human–Computer Interaction (HCI) [23,25], robot navigation [18,28], motion analysis [4,17], video compression and coding [12] and others. Although tracking has been studied for several decades and a considerable number of methods have been proposed, robust and accurate tracking with real-time performance has largely remained unsolved.

According to the features which are employed in the tracking procedure, tracking can be classified into two categories, low-level feature-based tracking and high-level feature-based tracking. In the high-level feature-based tracking, high-level features, such as contours [9], appearance templates [7] or geometric models [8,10,21] of the targets, are employed. The extraction of these high-level features depends not only on the results of image processing, but also on the perception of human beings about the targets. For example, a contour while being perceived as a higher-level profile description of the target, is more difficult to extract than a simple edge, which can be detected by low-level image processing. A prior model (or constraint) of the target plays an important role in the procedure for contour extraction. High-level features contain rich information about the targets and the high-level feature-based tracking is able to localize and track the object robustly and accurately. However, high-level features are of high dimensions. Tracking based on high-level features is usually carried out by hypotheses-and-test. Hypotheses generation and verification of the feature state are more complicated in cluttered environments. Computational consumption of this method is the main obstacle when it is applied in real-time tracking.

On the other hand, the low-level feature-based tracking can achieve high computational speed. Among the low-level features, geometric features are often employed especially when tracking rigid objects, *e.g.*, man-made objects. The geometric features include corners [18], edges/line segments [28], regions [15,16,11] and others. Tracking based on the low-level geometric features is usually carried out by establishing the coherent relation of each feature in successive images. Those features with similar motion parameters are grouped to represent the moving objects and background. When occlusion happens, some of the geometric features cannot be found. Low robustness under occlusion and in cluttered environments are the disadvantages when geometric features are employed for tracking. In this regard, non-geometric features, such as color, may help in the reliability of tracking. Color is a spectrum feature. The color model characterizes the appearance of the target. Tracking based on the color model can be considered as a procedure of object recognition and localization along the image sequence. The coherence of the track is maintained by two regions possessing the same color distribution. Compared with the geometric features and the high-level features, color features are more robust under the situations of partial occlusion, scale changes, non-rigidity and non-stationary viewing sensors [20]. Moreover, its fast computational speed is also attractive for most tracking systems.

Color has also become one of the important cues for human tracking. Due to the distinguishing feature of the skin color of human beings, the skin color model is usually applied for face tracking [27,5], hand localization [24] in the Human–Computer Interaction (HCI) and gesture recognition systems. The skin color

* Corresponding author. Fax: +65 6793 3318.
E-mail address: ecschua@ntu.edu.sg (C.S. Chua).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.04.023

model facilitates the segmentation of faces and/or hands from the background. Color has also been employed to track human body [23,13,3,14] and moving objects [6,1]. Furthermore, color is more identifiable and usually utilized as a feature to discriminate one object from others when tracking of multiple objects.

One problem usually encountered in the color-based tracking is that the color distribution of the target is not stable under illumination changes. A static color model is inadequate to capture the color changes over time; color model adaptation [26,13,24] is required. Among the color model adaptation methods, parametric (Gaussian and Mixture of Gaussians) approaches are usually utilized. In [26], a multivariate Gaussian is used to describe the skin color, while a Mixture of Gaussians model is employed in [13] to represent the color distribution of the target. In both of the above methods, linear parameter transition models based on *Maximum Likelihood* (ML) and *Expectation-Maximization* (EM) are used. The parameters of the previous n frames and the parameters learnt from the training samples in the current frame are integrated by the linear transition models to obtain the updated parameters. Due to background noise and illumination changes, the training samples used for the adaptation are not reliable before obtaining the adapted model [24]. The adaptation is carried out solely in the color space with the pre-adapted color model. Hence, some background data may be wrongly selected as training samples, while some data from the target may be ignored. The parameter estimation of the color model would thus be biased, being based on the unreliable sample set. Besides, the linear adaptation approach cannot work well for non-linear color changes. In [24], the color model adaptation does not rely on the linear adaptation model. Instead, it is formulated as a transductive learning problem. The learning of the non-stationary color distribution is carried out by classifying the data in the current frame with the *Expectation–Discrimination-Maximization* (D-EM) algorithm in a dimensionally reduced color space. Since this method is also carried out solely in the color space, the risk of mislabelling the training samples cannot be avoided. In the above approaches, even though the tracking is processed in the image space, spatial information is not utilized to help the selection of the training samples which are used to update the color model.

In this paper, a *Bilateral Learning* (BL) approach is proposed for the purpose of color-based object tracking. Similar with [24], the color model adaptation for tracking can be formulated as an unsupervised (or semi-supervised) learning problem. The algorithm is able to select reliable training samples to update the color model automatically. These samples are from the target instead of the background, even though they are of the same distribution in the color space. In the *Bilateral Learning* approach, the spatial information and color information of the target are combined in the color model adaptation and color-based object tracking procedure. The color model and spatial model are adapted in the color space and image space alternatively, which results in an adaptation of the color model and localization of the target along the image sequence. Compared to multiple-cue integration methods, the spatial information in the *Bilateral Learning* approach does not require an accurate shape model, such as the contour or edge of the target. It is probabilistic, and is approximately learnt in the tracking procedure. The spatial information helps to discriminate between the data which should be exploited to adapt the color model and which belong to the background. Accordingly, the color model maintains continuous tracking of the color changes over time. Furthermore, the color model adaptation is not sensitive to background attractions, with the spatial model being adapted simultaneously.

In the remainder of this paper, Section 2 introduces the color space and color model which are employed in this paper. Section 3 describes the *Bilateral Leaning* (BL) approach for the color-based object tracking. Some experimental results of the proposed method are presented in Section 4. In Section 5, comparisons of the proposed approach and some related works are discussed. Finally, conclusions are drawn in Section 6.

2. Color space and color model for tracking

The selection of the color space for the object tracking is not trivial. Some color spaces, such as RGB, are more sensitive to illumination changes, while other color spaces, such as HSV and YUV, may be more stable. Reflections and shadows are other considerations when selecting the color space. Shadows often cause significant changes in intensity but has minimum effect on chromaticity [11]. Although many approaches of color model adaptation are investigated to cope with the problems caused by illumination changes, a proper selection of the color space facilitate the color model adaptation.

The HSV (hue, saturation and value) color space is exploited in this chapter, where hue represents the chromaticity, saturation describes the degree of concentration of the color and value stands for the intensity. In the HSV color space, the influence of the illumination changes, reflections and shadow can be reduced to some extent and the requirement of the color model adaptation can be relaxed.

While the choice of color space is important, the modelling of the color distribution within the color space is just as important. The approaches for color modelling can be categorized into non-parametric [20,3] and parametric [26,13,24,23,6] approaches. Histogram is a typical non-parametric approach. A good example of utilizing color histogram for object tracking by mean shift is presented in [3]. However, there are two disadvantages with this approach. Firstly, in order to ensure the convergence of the histogram modelling, the training sample set needs to be large enough. Secondly, the quality of the histogram modelling is affected by the quantization of the color space. Typically, the stronger the model that is applied, the fewer examples required to build the model. Parametric approaches are more suitable for the unsupervised color model learning and adaptation. Mixture of Gaussians (MOG) model is a typical parametric approach and is employed to model the color distribution of the target in this study. With this parametric model, moderate training samples are enough for color modelling.

Assume that the color distribution of the target $\mathcal{O}$ consists of n components $\{\mathscr{C}_k, k = 1, \ldots n\}$, and each component can be represented by a single Gaussian. Let c stands for the color vector. The probability density function of the object's color is described as

$$p(\mathbf{c}) = \sum_{k=1}^{n} p(\mathbf{c}|\mathscr{C}_k)P(\mathscr{C}_k), \tag{1}$$

where, $\sum_{k=1}^{n} P(\mathscr{C}_k) = 1$ and $p(\mathbf{c}|\mathscr{C}_k) = \mathcal{N}(\mathbf{c}; \mu_k^c, \sum_k^c)$.[1] μ_k^c and Σ_k^c are the mean and covariance matrix of the kth component of the Mixture of Gaussians. The component number should be estimated when a Mixture of Gaussians is employed to model the object's color distribution. In [13,24], different estimation methods are proposed. For simplicity, we assume that the number of components for the Mixture of Gaussians has been obtained using some *ad hoc* methods [13] in advance. Pixels with large and small intensities are not included in the training data set because hue and saturation can become unstable in this range.

[1] For the simplification and clearness, we write $p(\mathbf{c})$ instead of $p(\mathbf{C} = \mathbf{c})$, and the same is the conditional density description.