Real time object tracking via a mixture model*

Dongxu Gao¹, Zhaojie Ju¹, Jiangtao Cao² and Honghai Liu¹

Abstract—Object tracking has been applied in many fields such as intelligent surveillance and computer vision. Although much progress has been made, there are still many puzzles which pose a huge challenge to object tracking. Currently, the problems are mainly caused by appearance model as well as real-time performance. A novel method was been proposed in this paper to handle both of these problems. Locally dense contexts feature and image information (i.e. the relationship between the object and its surrounding regions) are combined in a Bayes framework. Then the tracking problem can be seen as a prediction question which need to compute the posterior probability. Both scale variations and temple updating are considered in the proposed algorithm to assure the effectiveness. To make the algorithm runs in a real time system, a Fourier Transform (FT) is used when solving the Bayes equation. Therefore, the MMOT (Mixture model for object tracking) runs in real-time and performs better than state-of-the-art algorithms on some challenging image sequences in terms of accuracy, quickness and robustness.

I. INTRODUCTION

Object tracking is known as locating positions of interest area over time in every frame of video. This can be applied in many areas especially for human computer interaction. Although many different kinds of object tracking algorithms has been studied for several decades, and much progress has been made in recent years [1], there are still many challenging problems such as fast movement, illumination variation, occlusion, background clutters, proceeding time and so on.

Generally speaking, current tracking algorithms are categorized as two methods: generative tracker and discriminative tracker. There is a noticeable difference on them is how to build the appearance model [2]. A generative tracker focuses on the appearance of moving object and tries to find a model to represent it. It is unnecessary to consider the background information which makes the tracker works faster. However, the changes of appearance caused by some factors such as occlusion and pose variation makes it more difficult for modelling. Discriminative method emphasizes how to separate the target from backgrounds of video scenes. This tracker works well even though dramatic changes of object, but it needs more sophisticated calculation which makes it fail when higher process speed was needed. Therefore, various representative models become an important research topic. Like the tracking frame of state-of-the-art methods [3,4], the Bayes function will be used as a basis framework in this paper, but a novel appearance model and the solution method for model are quite different compared others. A MMOT (mixture model of object tracking) has been adopted in this paper which combine the colour information and context information when model the appearance of object. In addition, Flourier transform would be also called to compute object model for fast.

II. RELATED WORK

A typical tracking algorithm consists of four steps: object representation, search mechanism, model solving and model update. For recent generate trackers and discriminative trackers, both of their key step are how to acquire a better appearance of object. There are many papers focus on object information to get the target. Recently, there are several methods utilized context information to handle object tracking which locate the target through finding consistent information of object. To do so, related data mining method should be introduced for extracting both object and its surrounding region as supplement information, although satisfied results have been obtained, computationally cost are still needed. Not only that, templates and subspace models are also contribute to robust performance. Dong et al [5]utilized the subspace model which can handle appearance change while online learned model can learn appearance model in IVT methods. To solve this kind of model, many optimized algorithms [6] have been proposed to meet the real time such as proximal gradient approach and l1-norm related minimization method [7]. These methods seems sensitive to partial occlusion according to many experiments.

Although some algorithms [8] are proposed to manage occlusion while drift might be occur as the offline update of template or offline subspace model. Many researchers have developed the online update model which can deal with drift well. However, scale of object sometimes will change which pose another challenge for these trackers. For different trackers [9], scale updates should be considered separately. Compressive tracking [10] method cannot handle scale variance well but introduce a multi-scale information in fast compressive tracking (FCT) [11]. However, there is no colour information included for FCT which might fail when the colour of object and background are similar.

Lots of researches which exploited colour information provided excellent performance for object detection. This method not only handle similar colour problem but can also locate the first location of object [12]. Martin et al. [13]

*This work is supported by Dream project of the EU Seventh Framework Programme(611391).
¹Dongxu Gao, Zhaojie ju and Honghai Liu are with the Intelligent Systems and Biomedical Robotics group, School of Computing, University of Portsmouth, Portsmouth, England PO1 3HE, UK zhaojie.ju@port.ac.uk
²Jiangtao Cao is with the Department of Information and Control Engineering, Liaoning Shihua University, Fushun, Liaoning, China cigroup@126.com
a Gaussian weighted function

\[ p(x) = \sum_{f(z) \in X^t} p(x, f(z)) \]

\[ = \sum_{f(z) \in X^t} p(x|f(z))o \cdot p(f(z)) \tag{2} \]

Then, the problem can be transferred to compute the joint probability, represents the context feature. Where \( f(z) \) denotes image information including the location and the feature of target which can be represented as eq. (3).

\[ M(z) = (V(z), z) \tag{3} \]

M(z) denotes the colour information which adopted the HSV (Hue, Saturation, Value) colour space at location z(m, n), especially the value of V channel, z belongs to the neighbourhood of location X that includes target object. For more details of the colour information, we refer to [16].

To solve the eq. (2), two conditional probability should be computed separately.

\[ p(x|f(z), o) = h(x - z) \tag{4} \]

Where h can be seen as a kernel function with respect to the relationship between the centre location of object and its surrounding region. The result of h will be computed next. For another prior probability which can be modelled through a Gaussian weighted function

\[ \omega_o(z - x^*) = ae^{-\frac{|x-z|^2}{\sigma^2}} \tag{5} \]

And \( P(f(z)|o) \) can be represented as:

Then, take (2)(4)(5)(6) into account, that is

\[ p(x) = \sum_{f(z) \in X^t} h(x - z)V(z)\omega_o(z - x^*) \]

\[ = h(x) \bigodot V(x)\omega_o(x - x^*) \tag{6} \]

The location of object can be determined by the maximum value of \( p(x) \) of \((t+1)\)th frame, which can be represented as:

\[ F(bel|x-x^*|b) = F(h(x)) \bigodot F(I(x)\omega_o(x - x^*)) \tag{7} \]

In addition, the appearance model can be obtained by:

\[ h(x) = F^{-1}\left( \frac{F(bel|x-x^*|b)}{F(I(x)\omega_o(x - x^*))} \right) \tag{8} \]

and the updated model can be formulated as

\[ H_{i+1}^{mmot} = (1 - \rho)H_{i}^{mmot} + \rho h_i^{mmot} \tag{9} \]

IV. EXPERIMENTS

Our algorithm is implemented in Matlab and achieves about average 25 frames per second with on a PC with Intel E7500 CPU (2.93GHz).

In order to illustrate the qualitative comparison more clearly, some methods are described briefly here. The Visual Tracking Decomposition (VTD) [17] method used the observation model which is decomposed into multiple basic observation models that are constructed by sparse principal component analysis (SPCA) [5] of a set of feature templates. The MIL [18] method put all ambiguous positive and negative samples into bags to learn a discriminative model for tracking. The l1 method [19] adopted the holistic representation of the object as the appearance model and then tracks the object by solving the l1 minimization problem. The assessment of several methods above in different situations are shown as below:

![Fig. 1. The identification of different methods](image)

Although these methods can track the object both for the sequences Davidindoors and Occlusion1 as some occlusion occurs, if there are some rotation for the Occlusion2 sequences occurs, only our method performs well. L1 and VTD could not handle the severe occlusion for like Girl sequences. While only VTD method could not track the object again if there is a drift when tracking, other methods could keep tracking after temporally drift. For the sequences of Cliffbar, the color of moving object is nearly same with its surrounding region, only our method can keep tracking over the time, as both color information and context information were adopted when modeling appearance.

In addition, as both center error evaluation and overlap evaluation which is defined by the PASCAL VOC have been used to evaluate performance of the proposed algorithm.
Fig. 2. The sequences of Cliffbar

Fig. 3. The sequences of DavidIndoor

Fig. 4. The sequences of DavidOutdoor

Fig. 5. The sequences of Girl
Fig. 6. The sequences of Occlusion1

Fig. 7. The sequences of Occlusion2

Fig. 8. The center error result

Fig. 9. The overlap result
Through these pictures, the average center error is lowest among these methods, and the overlap rate is higher than others, so it is clear that our methods can handle the current challenges and performs best compares with other methods.

V. CONCLUSION

The method in this paper combines the colour information and context feature, which makes it has robustness for appearance change of object. It can works well even though occlusion and similar colour occurs. Not only that, scale update information and online update are considered to make it performs better. In addition, it can runs in a real time system as the algorithm computed in frequency domain through Flourier Transform. Qualitative and quantitative experiments prove the effectiveness and efficiency of MMOT algorithm compared with state-of-the-art methods. Therefore, the MMOT algorithm can applied in a real tracking system.

REFERENCES