# Real time object tracking via a mixture model*

Dongxu Gao[1] , Zhaojie Ju[1] , Jiangtao Cao[2] and Honghai Liu[1]

*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,

[1]Dongxu 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`

[2]Jiangtao Cao is with the Department of Information and Control Engineering, Liaoning Shihua University, Fushun, Liaoning, China `cigroup@126.com`

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]

analyses how colour information contribute the performance of tracking and the experiments proves the effectiveness compared with CSK tracker and VTD tracker[14,15].

In this paper, advantages of both colour information and Fourier transform are utilized for effectiveness and efficiency. The structure of this paper is organized as follows: MMOT is introduced in section 3; experimental results and discussion are given as follows, finally, conclusion will be given in section5.

## III. PROPOSED ALGORITHM

Recently, object tracking problem has been treated as a predictive problem which can be solved by particle filter framework based on Bayes function. The main difference compared with previous traditional particle filter framework is the number of particles are not needed for solving the model while using a kernel function to obtain the probability needed. When estimating the object location, the object location likelihood is used which is shown as follows:

$$p(x) = p(x|o) \tag{1}$$

x is the output vector which including the predicting object information and represents the current object feature in an image sequence. p(x) can be computed according to Bayes theory

$$
\begin{aligned}
p(x) = p(x|o) &= \sum_{f(\mathbf{z}) \in X^f} p(x, f(\mathbf{z}|o)) \\
&= \sum_{f(\mathbf{z}) \in X^f} p(x|f(\mathbf{z})|o)p(f(\mathbf{z})|o)
\end{aligned} \tag{2}
$$

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).

$$M(z) = (V(\mathbf{z}), \mathbf{z}) \tag{3}$$

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.

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

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

$$\omega_\sigma(z - x^*) = ae^{\frac{|z - x^*|^2}{\sigma^2}} \tag{5}$$

And $P(f(z)|o)$ can be represented as:
Then, take (2)(4)(5)(6) into account, that is

$$
\begin{aligned}
p(x) &= \sum_{f(\mathbf{z}) \in X^f} h(x - \mathbf{z})V(\mathbf{z})\omega_\sigma(\mathbf{z} - x^*) \\
&= h(x) \bigodot V(x)\omega_\sigma(x - x^*)
\end{aligned} \tag{6}
$$

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

$$F(be^{|\frac{(x - x^*)}{\alpha}|^\beta}) = F(h(x)) \bigodot F(I(x)\omega_\sigma(x - x^*)) \tag{7}$$

In addition, the appearance model can be obtained by:

$$h(x) = F^{-1}\left(\frac{F(be^{-|\frac{x - x^*}{\alpha}|^\beta})}{F(I(x)\omega_\sigma(x - x^*))}\right) \tag{8}$$

and the updated model can be formulated as

$$H_{t+1}^{mmot} = (1 - \rho)H_t^{mmot} + \rho h_t^{mmot} \tag{9}$$

## 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:
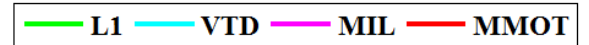


Fig. 1. The identification of different methods

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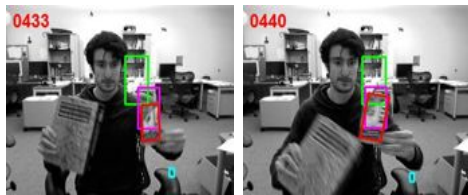
In addition,as both center error evaluation and overlap evaluation which is defined by the PASCAL VOC have been used to evaluate performance of the proposed algorithm.
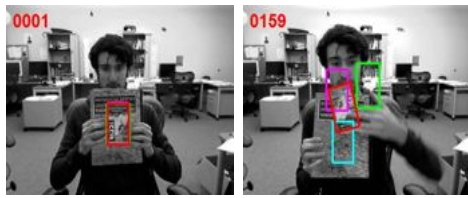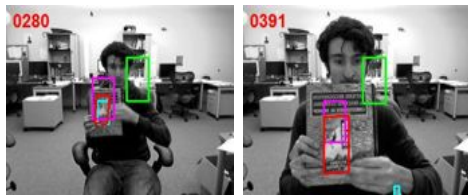
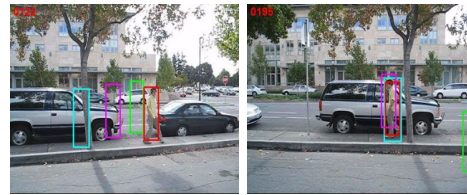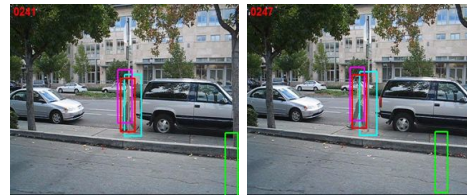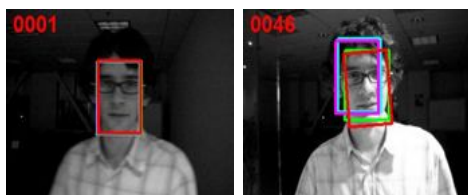Fig. 2. The sequences of Cliffbar



Fig. 4. The sequences of DavidOutdoor



Fig. 3. The sequences of DavidIndoor
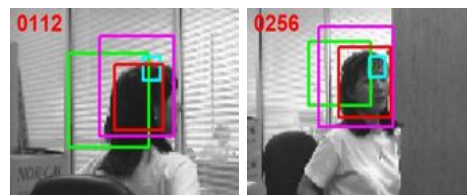


Fig. 5. The sequences of Girl
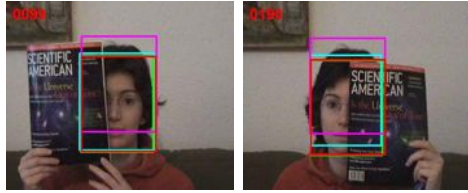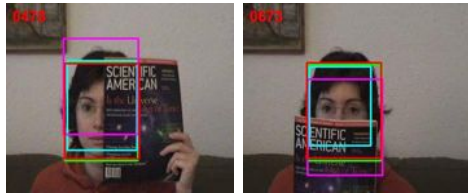
Fig. 6. The sequences of Occlusion1
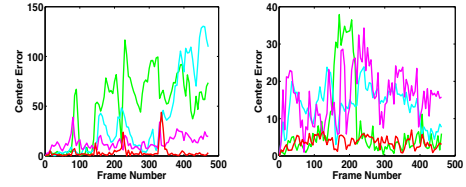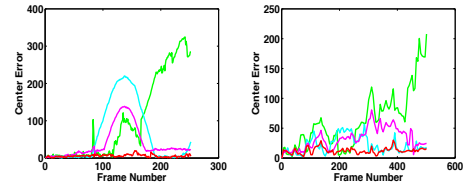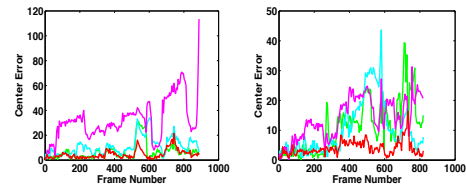


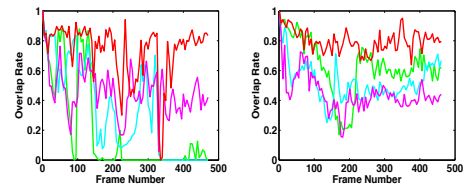Fig. 8. The center error result
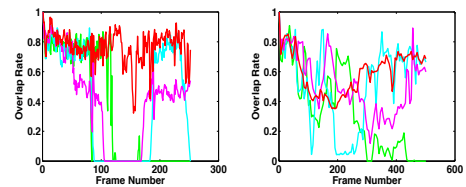


Fig. 7. The sequences of Occlusion2
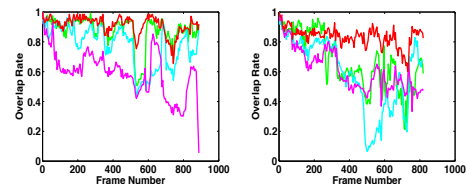


Fig. 9. The overlap result

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

[1] Y. Yuanlong, G. K. I. Mann and R. G. Gosine, "A single-object tracking method for robots using object-based visual attention," *International Journal of Humanoid Robotics* **4**(2012) pp. 1-36.

[2] M. Heber, M. Godec, M. Ruther, P.M. Roth, H.Bischof, "Segmentation based tracking by support fusion," *Computer Vision and Image Understanding* **117**(2013) pp.573-586

[3] W. Brendel and S. Todorovic, "Video Object Segmentation by Tracking Regions,"*IEEE Int. Conf. on Computer Vision (ICCV)*,(IEEE Press, Kyoto, USA, 2009), pp. 833–840.

[4] S.Avidan, "Ensemble tracking," *IEEE Transaction on pattern analysis and machine intelligence***29**(2)2007 261–271.

[5] D. Wang, H. Lu, and Y. W. Chen, "Incremental MPCA for color object tracking,"*International Conference on Pattern Recognition (ICPR)*, (IEEE Press, Istanbul, USA, 2010)pp. 1751–1754.

[6] D.Ross, J.Lim, R.S. Lin, and M.H. Yang, "Incremental Learning for robust visual tracking,"*International Journal of Computer Vision (IJCV)*, (Springer Press, USA, 2008) pp. 125–141.

[7] R. Yao, Q. Shi, C. Shen, Y. Zhang and A. Hengel, "Part-based Visual Tracking with Online Latent Structural Learning," *Computer Vision and Pattern Recognition (CVPR)*, (IEEE Press, Portland, USA, 2013), pp. 2363–2370.

[8] Y. Wu, H. Ling, J. Yu, F. Li, X. Mei and E. Cheng, "Blurred Target Tracking by Blur-driven Tracker," *Computer Vision (ICCV)*, (IEEE Press, Barcelona, USA, 2011), pp. 1100–1107.

[9] J. A. Lasserre, C. M. Bishop, and T. P. Minka, "Principled hybrids of generative and discriminative models,"*Computer Vision Pattern Recognition (CVPR)*, (IEEE Press, New York, USA, 2006), pp. 87–94.

[10] Zhang, K.H., L. Zhang, and M.H. Yang, "PReal-Time Compressive Tracking. Computer Vision" - Eccv 2012, Pt Iii, 2012. 7574: p. 864–877.

[11] K.H., Zhang, L,Zhang, Q.S. Liu, D,Zhang, and M.H. Yang. "Fast Visual Tracking via DenseSpatio-Temporal Context Learning,"*(ECCV)*, pp. 127–141.

[12] B. Liu, J. Huang, L. Yang, and C. A. Kulikowski. "Robust tracking using local sarse appearance model and k-selection," *Computer Vision and Pattern Recognition (CVPR)*, (IEEE Press, Providence, RI, USA, 2011), pp. 1313–1320.

[13] Martin Danelljan, Fahad Shahbaz Khan, Michal Felsberg and Joost van de Weijer. "Adaptive color Attributes for real-time Visual Tracking."*CVPR 2014*

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions Pattern Analysis and Machine Intelligence*, **31**(1) (2009) pp. 210–227.

[15] J. Kwon, K. M. Lee, "Visual Tracking Decomposition,"*Computer Vision and Pattern Recognition (CVPR)*, (IEEE Press, San Francisco, USA, 2010), pp. 1269–1276.

[16] M. Danelljan, et al. "Adaptive color attributes for real-time visual tracking." *2014 IEEE Conference on Coputer Vision and Pattern Recognition (CVPR)*, 2014, pp.1090–1097

[17] Kwon, Junseok, and Kyoung Mu Lee. "Visual tracking decomposition." *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,2010,pp.1269–1276

[18] B. Babenko, M.H. Yang,and S. Belongie. Visual tracking with online multiple instance learning." *2009 IEEE Conference on Coputer Vision and Pattern Recognition (CVPR)*. 2009, pp. 983–990.

[19] X. Mei, H. Ling,Y.Wu,E.Blasch, and L.Bai. "Minimum errorbounded efficient L1 tracker with occlusion detection."*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,2011.