Mean shift algorithm based on fusion model for head tracking

An Guocheng[1, 2] Gao Jianpo[1] Wu Zhenyang[1]

([1]School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

([2]Intelligence Engineering Lab, Institute of Software Chinese Academy of Sciences, Beijing 100190, China)

Abstract: To solve the mismatch between the candidate model and the reference model caused by the time change of the tracked head, a novel mean shift algorithm based on a fusion model is provided. A fusion model is employed to describe the tracked head by sampling the models of the fore-head and the back-head under different situations. Thus the fusion head reference model is represented by the color distribution estimated from both the fore-head and the back-head. The proposed tracking system is efficient and it is easy to realize the goal of continual tracking of the head by using the fusion model. The results show that the new tracker is robust up to a 360° rotation of the head on a cluttered background and the tracking precision is improved.

Key words: mean shift; head tracking; kernel density estimate; fusion model

The efficient tracking of a head in a complex environment is one of the most fundamental tasks in computer vision applications such as visual surveillance, video conferences, and human computer interaction, etc. However, developing a head tracking algorithm that is robust under a wide variety of conditions such as partial occlusion, clutter, and changes in the head appearance remains a challenging task for the visual domain. The main challenge may come from the head appearance changing during the whole tracking process. Typically a head tracker must handle a large degree of head rotation. At the same time, it is critical that the tracking algorithm should have low computational complexity because high-level tasks such as face recognition may be the main task of the system. Although combining multiple features may be an attractive option, the burden of computation increases at the same time.

Recently, mean shift has become a new important image processing algorithm which was originally developed by Fukunaga and Hostetler[1] and was further developed by Cheng[2]. Its efficacy has been demonstrated in solving many computer vision problems[3-4], such as image segmentation, smoothing, and texture classification, etc. It is a simple iterative statistical method and has been proved to be an excellent and real-time algorithm in video object tracking. Peng et al. [5] integrated an adaptive model update mechanism into the mean-shift-based tracking system. Shan et al. [6] proposed to integrate advantages of the particle filter and mean shift for improved hand tracking, using mean shift to drive particles to local peaks in the likelihood function, thus improving the sampling efficiency. And they realized real-time hand tracking in dynamic environments of the wheelchair. Yang et al. [7] presented a tracking algorithm using a new simple symmetric similarity function for kernel density estimation in a joint spatial-feature space.

A crucial component in the mean shift framework is how to represent the object[8]. In head tracking, the key challenge is to capture the variability of the head model and its color can vary over time dependent on the visual angle. In this paper, we are interested in the best way to use the type of prior knowledge for head representation; specifically, how to encode the head color variability. So we present a new approach for head tracking based on a fusion color model in the mean shift procedure, whose statistical distribution characterizes the head variation automatically. Before tracking, the fusion model is constructed. Compared with an ordinary reference model, our fusion model is more robust and simpler. It is also very convenient for integration of the shape information and generalization to other suitable target trackings.

1 Mean Shift

Mean shift algorithms are successful approaches in visual object tracking and have become popular due to their simplicity and robustness. In mean shift trackers, a color histogram is usually used to describe the target region of interest. The Bhattacharyya coefficient is used to measure the similarity between the reference model and the current candidate model. Tracking is accomplished by iteratively finding the local minima of a similarity measure between the kernel density estimates of the reference model and the target image.

Here we review some basic theories of the mean shift algorithms. Let x_k denote the k-th pixel coordinate of the object image in a frame, centered at position x_0. The function $b: R^2 \rightarrow \{1, 2, \dots, m\}$ maps a pixel x_k to its bin and m is the number of the bins. The reference model of the target of interest is represented in its feature space by the function calculated using kernel function k defined by

$$\hat{q}_u = C\sum_{k=1}^{n} k\left(\left\|\frac{x_0 - x_k}{h}\right\|^2\right)\delta[b(x_k) - u] \qquad u = 1, 2, \dots, m \tag{1}$$

where h is the bandwidth of the kernel and n is the total number of points in the kernel. The constant C is derived by imposing the condition $\sum_{u=1}^{m}\hat{q}_u = 1$. Similarly, the target candidate centered at location y is calculated by

$$\hat{p}_u(y) = C_h\sum_{k=1}^{n} k\left(\left\|\frac{y - x_k}{h}\right\|^2\right)\delta[b(x_k) - u]$$

Received 2008-11-13.

Biographies: An Guocheng (1979—), male, doctor; Wu Zhenyang (corresponding author), male, professor, zhenyang@ seu. edu. cn.

Foundation items: The National Natural Science Foundation of China (No. 60672094, 60673188, U0735004), the National High Technology Research and Development Program of China (863 Program) (No. 2008AA01Z303), the National Basic Research Program of China (973 Program) (No. 2009CB320804).

Citation: An Guocheng, Gao Jianpo, Wu Zhenyang. Mean shift algorithm based on fusion model for head tracking[J]. Journal of Southeast University (English Edition), 2009, 25(3): 299 − 302.

$$u = 1, 2, \ldots, m \qquad (2)$$

where $\sum_{u=1}^{m} \hat{p}_u(y) = 1$ and C_h is a normalization constant. The Bhattacharyya coefficient is a popular measure between the reference model and the candidate model. The distance between the two model distributions is computed as

$$d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \qquad (3)$$

The sample estimate $\hat{\rho}(y)$ of the Bhattacharyya coefficient is given by

$$\hat{\rho}(y) \equiv \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y)\hat{q}_u} \qquad (4)$$

The mean shift algorithm aims to recursively minimize the distance by shifting $\hat{y}_0$ to a new centre location $\hat{y}_1$ where the similarity of the reference model and the candidate model shows an increase in the Bhattacharyya coefficient. The new target location $\hat{y}_1$ is obtained from its initial location $\hat{y}_0$ using the following relationship:

$$\hat{y}_1 = \frac{\sum_{k=1}^{n} x_k \omega_k g\left(\left\|\frac{\hat{y}_0 - x_k}{h}\right\|^2\right)}{\sum_{k=1}^{n} \omega_k g\left(\left\|\frac{\hat{y}_0 - x_k}{h}\right\|^2\right)} \qquad (5)$$

where $g(x) = -k'(x)$ and the weights ω_k are computed as

$$\omega_k = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \delta[b(x_k) - u] \qquad (6)$$

2 Fused Model

One major component of mean shift object tracking algorithms is a proper target representation which is used to estimate the similarity. The head is generally represented by various features such as color, contour and texture. In general, color is one of the most attractive features, which is commonly used in literature due to its robustness against non-rigidity, rotation, and partial occlusion. In the current color model, the head is usually represented by the color distribution estimated from the face region. Although many successful examples have been reported by using the above color model, they usually fail in the scenarios when the head feature varies seriously. In order to handle this problem, many researchers adopt updating the reference model in some periodic fashion with the constraints of some specified threshold mechanisms. But the situation when the head histogram is updated is a very difficult problem, because it requires one to detect whether a changing appearance is due to the head rotation or a temporary occlusion. The updated model methods are also easily influenced by the background, because the background information is easily contained in the reference model during the updating process.

In order to solve these problems, a novel color-based head representation solution is proposed in this paper. In the proposed color model, the head as a whole color distribution is estimated from both the pixels in fore-head and back-head regions. The fusion model in an 11 × 11 synthetically spatial space from two models to one is shown in Fig. 1. The detail of the fusing procedure is as follows: Select model pixels alternately from Figs. 1(a)and(b), then rearrange the two pixels set as Fig. 1(c). The effect of the real head fusion model forms model A and model B as is shown in Fig. 2.

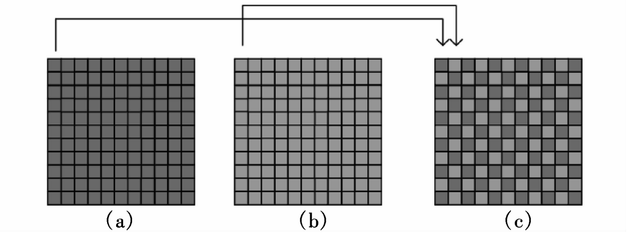


Fig. 1 Fused model forms two different models

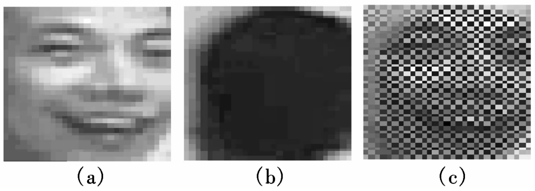


Fig. 2 Fused head model forms two different head appearances. (a) Model A; (b) Model B; (c) Fused model

In order to interpret the rationality of the fusion model by the mean shift theory, we now derive the parameter ω_k. The Bhattacharya coefficient can be approximated as

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(y)\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} = \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{y}_0)\hat{q}_u} + \frac{C_h}{2}\sum_{k=1}^{n_h} \omega_k k\left(\left\|\frac{y - x_k}{h}\right\|^2\right) \qquad (7)$$

where the first term is a constant and

$$\omega_k = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \delta[b(x_k) - u] \qquad (8)$$

For qualitatively explaining the validity of the fusion model in the head tracking process, we assume that model A, model B, and backgrounds do not have the same non-0 bins mutually. Then, without loss of generality, the candidate model $\hat{p}(y_0)$ is assumed to be composed of part of model A and the background. When computing the coefficient ω_k, the ingredient of model B in the fusion model does not have any effect. Because the corresponding candidate model bins are 0, the correspondence ω_k is equal to 0. So the non-0 coefficient ω_k mainly comes from the ingredients of model A in the fusion model. From the fusion model construction process, the bins' proportions hardly change. In other words, the non-0 coefficients ω_k of the fusion model are not changed relatively. Thus the updating location $\hat{y}_1$ is not influenced. We also can look at it from another angle so that, the so-called fusion model in fact uses two low resolution models to carry on the matching process in a more precise image space, and the needed model is full-automatically completed. This is also the biggest superiority of this novel tracker. The whole mean shift procedure is as follows:

1) Compute the kernel histogram $\hat{p}_u(y_0)$ in the reference frames and it can evaluate the Bhattacharyya coefficient $\rho[\hat{p}(\hat{y}_0), \hat{q}]$;

2) Compute the weights ω_k according to Eq. (6);

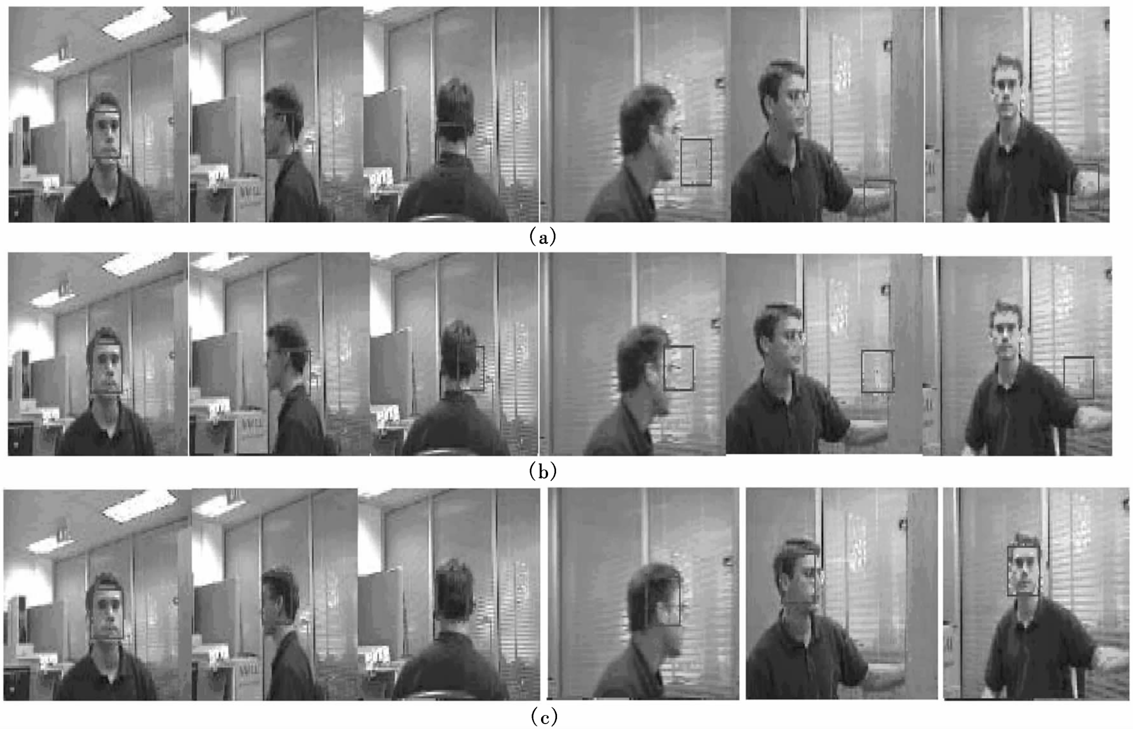
3) The new recursive location $\hat{y}_1$ is calculated by Eq. (5) and it can evaluate $\rho[\hat{p}(\hat{y}_1), \hat{q}]$;

4) While $\rho[\hat{p}(y_1),\hat{q}] < \rho[\hat{p}(y_0),\hat{q}]$, do $\hat{y}_1 \leftarrow \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$;

5) Stop if $|\hat{y}_1 - \hat{y}_0| < 1$; otherwise, if $\hat{y}_1 \rightarrow \hat{y}_0$, return to step 1).

3 Experimental Results

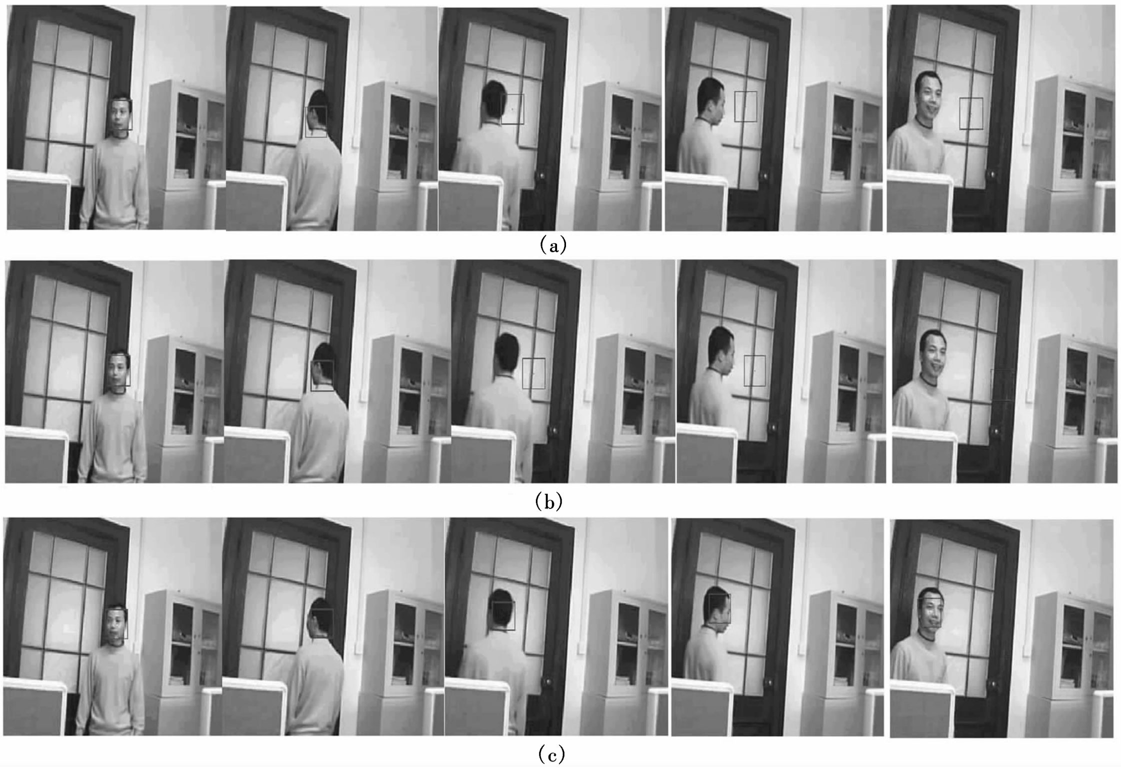
We compare the performance of the proposed fusion model-based mean shift algorithm with the standard mean shift algorithm and the adaptive model mean shift algorithm on real video sequences (see Fig. 3 and Fig. 4). Results are presented for two image sequences depicting a 360° rotation of the head on different clutter backgrounds. Here, we just present some representative results in the following figures. In all experiments, the RGB color space is quantized into 16


(a)

(b)

(c)

Fig. 3 The office sequences tracking outcomes. (a) The single reference model mean shift procedure; (b) The adaptive model mean shift procedure; (c) The fusion model-based mean shift


(a)

(b)

(c)

Fig. 4 Other sequences tracking outcomes. (a) The single reference model mean shift procedure; (b) The adaptive model mean shift procedure; (c) The fusion model-based mean shift

×16 ×16 bins. The Gaussian kernel is selected.

The office sequences come from Stanford Vision Laboratory and the frame resolution is 128 ×96 pixels. The target is initialized with a hand-drawn rectangular region of size 23 × 27. The rectangle shows the position of a person's head. For comparison, we show the office tracking results using a general model-based mean shift algorithm in Fig. 3 (a). From left to right, frames 15, 26, 32, 43, 49, and 56 are displayed. The tracking is not good in frame 32 and fails in frame 43. The other sequences are screened. The resolution is 320 ×240 pixels. Fig. 4(a) shows results of tracking a head in an indoor environment with general mean shift. From left to right, frames 18, 24, 29, 32, and 43 are displayed. Fig. 4(b) shows the results of the tracking head with the adaptive model mean shift. Fig. 4(c) gives the results of the new algorithm. Serious problems such as heavy rotation and changing background are correctly handled.

4 Conclusion

Our main contribution in this paper is to extend the mean shift framework for tracking heads with large varying features by introducing the fusion model. We have shown how it is possible to make use of head features from the mean shift algorithm. The fusion model is better than the ordinary color feature model in the tracking process. The method is validated on many real-complexity video sequences. The results are compared with the general color model and the adaptive model mean shift algorithm and show the superiority of the improved algorithm. Moreover, the fusion model-based mean shift head tracking introduced in this paper shows excellent efficiency and straightforward implementation.

References

[1] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, **21**(1): 32 − 40.

[2] Cheng Yizong. Mean shift, mode seeking, and clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(8): 790 − 799.

[3] An Guocheng, Chen Jianjun, Wu Zhenyang. A fast external force model for snake-based image segmentation[C]//*IEEE International Conference on Signal Processing*. Beijing, China, 2008: 1128 − 1131.

[4] Zhou Huiyu, Yuan Yuan, Shi Chunmei. Object tracking using SIFT features and mean shift[J]. *Computer Vision and Image Understanding*, 2009, **113**(3): 345 − 352.

[5] Peng Ningsong, Yang Jie, Liu Zhi. Mean shift blob tracking with kernel histogram filtering and hypothesis testing[J]. *Pattern Recognition Letters*, 2005, **26**(5): 605 − 614.

[6] Shan Caifeng, Tan Tieniu, Wei Yucheng. Real-time hand tracking using a mean shift embedded particle filter[J]. *Pattern Recognition*, 2007, **40**(7): 1958 − 1970.

[7] Yang Changjiang, Duraiswami R, Davis L. Efficient mean-shift tracking via a new similarity measure[C]//*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, 2005: 176 − 183.

[8] Hu Jwu-Sheng, Juan Chung-Wei, Wang Jyun-Ji. A spatial-color mean-shift object tracking algorithm with scale and orientation estimation[J]. *Pattern Recognition Letters*, 2008, **29**(16): 2165 − 2173.

基于融合模板的均值移动头部跟踪算法

安国成[1,2]　高建坡[1]　吴镇扬[1]

([1] 东南大学信息科学与工程学院,南京 210096)
([2] 中国科学院软件研究所人机交互技术与智能信息处理实验室,北京 100190)

摘要:针对被跟踪头部目标特征状态随时间变化而与参考模板不匹配的问题,提出一种基于融合参考模板的均值移动算法,即将被跟踪目标在不同状态下所呈现出的不同特征使用采样的方法进行融合,如将头部跟踪过程中正面的肤色信息和后面的发色信息进行融合,从而形成一个包含不同特征的参考模板. 在跟踪过程中,使用该融合模板可以有效地克服由被跟踪目标特征变化导致跟踪失败而不能实现头部连续跟踪的问题. 通过头部跟踪实验可以看出,该算法实现了复杂环境下的具有360°旋转的头部跟踪,并且在一定程度上提高了跟踪精度.

关键词:均值移动;头部跟踪;核密度估计;融合模板

中图分类号:TP391