Head Tracking Based on Histogram and Shape Model

Qingshan Liu, Songde Ma, Hanqing Lu
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, P.O.Box 2728, Beijing 100080, China
Emails:{qsliu, masd, luhq}@nlpr.ia.ac.cn

Abstract

A new model-based method is presented for real-time tracking a person's head in gray image sequences. Multidimensional receptive field histogram and ellipse shape model is used to describe the tracking object (head). A robust nonparametric technique called mean shift algorithm is adopted to estimate the most probable head location in the current frame, in which histogram matching is used during mean shift iteration. In order to locate the head more accurately and obtain the best scale size of the head, a local search for maximizing the gradient magnitude around the boundary of the elliptical head is preformed after mean shift estimation. It is demonstrated to be a real-time tracker and robustness to scale variation, arbitrary camera movement, part-occlusion and so on, for several image sequences.

1.Introduction

Automatic visual tracking of a person is a promising goal for computer vision research, it can be used in video conference, automatic surveillance, human computer interface and so on. There were many developing systems that were capable of performing this task [3][7][8][10]. But most of these systems didn't work well under the skin color background or gray image sequences, because the skin color feature was employed in these systems.

In this paper, the Multidimensional receptive field histogram [2] (MRFH) model is used to describe a person's head. Multidimensional receptive field histogram is a technique using local descriptions of an object's shape provided by a vector of linear neighborhood operators (receptive fields), and it can be used to identify objects in a scene, independent of the object's position, image-plane orientation and scale. Because it is irrelevant to skin color, our method can track the head in gray image sequences. Another advantage of our method presented in this paper is no prior motion model. We adopt a robust nonparametric technique called mean shift algorithm [1][4][9] to estimate the most probable head location in the current frame, in which histogram matching is used during mean shift iteration. In order to locate the head more accurately and obtain the best scale s size of the head, we combine with the elliptical head model[3], a local search for the maximizing gradient magnitude around the boundary of the elliptical head after the mean shift estimation.

Our method is similar to reference [1], in which mean shift algorithm is used to track objects. But RGB color model is replaced by MRFH model to describe the object in our method, so our method can track the object in gray image sequences. In addition, we just use the simple histogram matching l^2 function to be similarity measurement during mean shift iterations while [1] took model matching to be the problem of classification using Bhattacharyya coefficient metric. Our tracking method also combine with reference [3]. The method is demonstrated to be insensitive to partial occlusion, arbitrary camera movement and complex background, and robustness to light source intensity, changes in viewing angle and so on, for several sequences.

The paper is organized as follows. Mean shift analysis is introduced in Section 2 and multi-dimensional receptive field histogram is introduced in Section 3. The tracking algorithm is developed and analyzed in Section 4. Experiments are given in Section 5. In finally, we will give our conclusion.

2. Mean Shift Analysis

Let $\{x_i\}_{i=1\ldots n}$ be an arbitrary set of n points in the d-dimensional space R^d. The multivariate kernel density estimation obtained with kernel $K(x)$ and window radius (band-width) h, computed in the point x is defined as:

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{x - x_i}{h}) \qquad (1)$$

The kernel $K(x)$ is a scalar function that must satisfy some conditions, details in reference [4]. There are two kinds of commonly used kernel: Epanechnikov kernel and Multivariate normal kernel. The latter is used in our experiments:

$$K_N(x) = (2p)^{-d/2}\exp(-\frac{1}{2}\|x\|^2) \qquad (2)$$

Define the profile[9] of a kernel K as a function $k:[0,\infty)\to R$ such that $K(x) = k(\|x\|^2)$. So from (2), the normal profile is given by

$$k_N(x) = (2p)^{-d/2}\exp(-\frac{1}{2}x) \qquad (3)$$

Employing the profile notation, the density estimation(1) can be written as:

$$\hat{f}_K = \frac{1}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \qquad (4)$$

The estimation of the density gradient can be defined as the gradient of the kernel density estimation (4), and we denote $g(x) = -k'(x)$, so the density gradient estimation is given by

$$\hat{\nabla} f_K(x) \equiv \nabla \hat{f}_K(x) = \frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x - x_i)k'\left(\left\|\frac{x - x_i}{h}\right\|^2\right)$$

$$= \frac{2}{nh^{d+2}}\sum_{i=1}^{n}(x - x_i)g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)$$

$$= \frac{2}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)} - x\right] \qquad (5)$$

Because the derivative of the normal profile remains an exponential, the expression $\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)$ is nonzero.

The last bracket in (5) contains the mean shift vector:

$$M_{h,G}(x) \equiv \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)} - x$$

$$= \frac{\sum_{i=1}^{n} x_i G\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^{n} G\left(\frac{x - x_i}{h}\right)} - x \qquad (6)$$

Compute with a kernel $G(x)$ defined by $G(x) = cg(\|x\|^2)$, where c is the normalization constant. We also have the density estimation at x computed with the kernel G.

$$\hat{f}_G(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} G\left(\frac{x - x_i}{h}\right) = \frac{c}{nh^d}\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \qquad (7)$$

Combined with (6) and (7), then (5) becomes

$$\hat{\nabla} f_K(x) = \hat{f}_G(x)\frac{2/c}{h^2} M_{h,G}(x) \qquad (8)$$

From where it follows that

$$M_{h,G}(x) = \frac{h^2}{2/c}\frac{\hat{\nabla} f_K(x)}{\hat{f}_G(x)} \qquad (9)$$

The equation (9) shows that the mean shift vector obtained with kernel G is an estimation of the normalized density gradient obtained with kernel K. From (9), we also can see that the mean shift vector always points towards the direction of maximum increase in the density, and the mean shift step is large for low density regions and decreases as x approaches a high density region.

The mean shift procedure is defined recursively by computing the mean shift vector $M_{h,G}(x)$ and translating the center of kernel G by $M_{h,G}(x)$.

Define the sequence of successive locations of the kernel G as $\{y_j\}_{j=1,2...}$, where

$$y_{j+1} = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{y_j - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{y_j - x_i}{h}\right\|^2\right)}, \quad (\text{with } j=1,2\ldots)$$

is the weighted mean at y_j computed with kernel G and y_1 is the center of the initial kernel. The jth mean shift vector is given by $y_{j+1} - y_j$. The convergence of the sequences $\{y_j\}_{j=1,2...}$ was proved in reference [4].

3. Multidimensional Receptive Field Histogram

Bernt Schiele[2] presented a technique to identify object in a scene using matching Multidimensional Receptive Field Histogram which used local descriptions of an object's shape provided by a vector of linear neighborhood operators (receptive fields). It can be used to determine the most probable objects in a scene, and it is independent of the object's position, image-plane orientation and scale, and robustness to changes in viewing angle and to partial occlusion because of using histogram matching. In a sense, tracking is equal to finding the object (target) in every frame of the sequences, so it is reasonable to use this technique for tracking.

As for multidimensional receptive field histogram matching, the following problems are demanded:

- Choose local property measurements (see 3.1)
- Compare histograms(see 3.2)
- Design parameters of the histograms: number of dimensions of histogram and resolution of each axis.

3.1 The Local Characteristics

The calculation of local properties can be divided into the local linear point-spread function (formula (10)), and the normalization function used during measurement of local properties.

$$\mathrm{Im}\, g_{Mask}(x,y) = \sum_{i,j=-m,-n}^{m,n} \mathrm{Im}\, g(x+i, y+j)Mask(i,j) \qquad (10)$$

Several local operators were discussed in reference [2]. In this paper, our experiments are performed with the simple first derivative and Laplacian local operators given by:

$$M_{dx} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad M_{dy} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad M_{lap} = \begin{pmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{pmatrix}$$

Normalization by energy is able to remove the effects of variation in illumination, and it was demonstrated to be the most robust in respect to additive Guassian noise[2], so it is employed in this paper.

$$\mathrm{Im}\, g_{ene} = \frac{\sum_{i,j} \mathrm{Im}\, g(x+i, y+j)Mask(i,j)}{\sqrt{\sum_{i,j} \mathrm{Im}\, g(x+i, y+j)^2}\sqrt{\sum_{i,j} Mask(i,j)}} \qquad (11)$$

3.2 Histogram Comparison

There are many measurement functions for histogram comparison, such as the "intersection" measurement[5], the sum of squared distances commonly used in signal processing, l^2-test based on mathematical statistics and so on. l^2-test mea-surement function is adopted in this paper, for it was demonstrated to be the most reliable form of histogram comparison for multidimensional receptive field histogram [2].

$$l^2{}_{TH}(H,T) = \sum_{i,j} \frac{(H(i,j) - T(i,j))^2}{H(i,j) + T(i,j)} \qquad (12)$$

4.Tracking Algorithm

The tracking method can be divided to two steps. First, MRFH matching through mean shift iteration is used to estimate the most probable location, then elliptical head gradient model is performed to obtain the more accurate location and the best scale size.

4.1 MRFH Model Representation

Assuming the set $(x_i)_{i=1\ldots n_h}$ is the pixel locations of the head region, centered at y. We define the index of the histogram (MRFH) bin corresponding to the pixel at location x_i^* is $b(x_i^*)$. The probability of the intensity u (that is the intensity after local operation) in this region is derived by employing a convex and monotonic decreasing kernel profile k with radius h which assigns a smaller weight to the locations that are farther from the center of the head region. The weight increases the robustness of the estimation, since the peripheral pixels are the least reliable, being often affected by occlusions (clutter) or background. Hence we can represent the MRFH model of head region as equation (13).

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) d\,[b(x_i) - u] \quad {}_{(u=1\ldots m)} \qquad (13)$$

Where d is the Kronecker delta function and C_h is the normalization constant. By imposing the condition $\sum_{u=1}^{m} \hat{p}_u = 1$, we obtain $C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)}$.

During tracking, the histogram of the head candidate in all the frames is still computed using equation (13), but the radius h and the number of the pixels will be changed with the scale variation.

4.2 Maximizing Gradient Magnitude

The head's shape information is also an important feature for effective head tracker[3][10]. In this paper, we shape the head to be an ellipse, and an elliptical head gradient model[3] is used after mean shift iteration, in order to obtain the more accurate location by maximizing gradient magnitude around the perimeter of the ellipse and the best scale s size of the head through a local search around the result of mean shift iteration. The gradient model is given by:

$$f_g(s) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left| g_s(i) \right| \qquad (14)$$

Where $g_s(i)$ is the intensity gradient at perimeter pixel i of the ellipse at location s, and N_s is the number of pixels on the perimeter of ellipse with scale size s

4.3 Tracking Procedure

During mean shift iterations, assuming q_u is the known histogram model and $\hat{p}_u(y)$ is the histogram of the candidate region centered at y, then the most probable location $\hat{y}$ of the head in the current frame can be obtained by minimizing the equation (12). Because it is a reasonable assumption of no drastic change between two consecutive frames, we search for the new head location in the current frame starts at the estimated location $\hat{y}_0$ that is the head location in the previous frame. Thus, the histogram $\{\hat{p}_u(\hat{y}_0)\}_{u=1\ldots m}$ can be computed first, using Taylor expansion around the value $\hat{p}_u(\hat{y}_0)$, equation (12) can be approximated as equation (15).

$$l^2[\hat{p}(y), q] = \sum_{u=1}^{m} \frac{(\hat{p}_u(y) - q_u)^2}{\hat{p}_u(y) + q_u}$$

$$\approx f(\hat{p}(\hat{y}_0), q) + \sum_{u=1}^{m} \left(1 - \frac{4q_u^2}{(\hat{p}_u(\hat{y}_0) + q_u)^2}\right) \hat{p}_u(y) \qquad (15)$$

Where

$$f(\hat{p}(\hat{y}_0), q) = \sum_{u=1}^{m} \frac{\hat{p}_u^3(\hat{y}_0) + 4q_u^3 - 2\hat{p}_u^2(\hat{y}_0)q_u - 3\hat{p}_u(\hat{y}_0)q_u^2}{(\hat{p}_u(\hat{y}_0) + q_u)^2} \qquad (16)$$

Because $\sum_{u=1}^{m} \hat{p}_u(y) = 1$ and incorporating equation (13), equation (15) becomes

$$l^2[\hat{p}(y), q] \approx f(\hat{p}(\hat{y}_0), q) + 1 - C_h \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \qquad (17)$$

Where $w_i = \sum_{u=1}^{m} \frac{4q_u^2}{(\hat{p}_u(\hat{y}_0) + q_u)^2} d\,[b(x_i) - u]$ (18)

From expression (16), we know $f(\hat{p}(\hat{y}_0), q)$ is independent of y, so to minimize l^2, the last term in equation (17) has to be maximized. The last term represents the density estimation computed with kernel profile k at y in the current frame with the data being weighted by w_i (18). Based on mean shift iterations, the best histogram matching can be achieved using the following algorithm (from step 1 to step 4).

The whole tracking procedure is:

1. From previous frame, the initial location $\hat{y}_0$ and the scale s size of the head in the current frame

can be estimated, compute the distribution $\{\hat{p}_u(\hat{y}_0)\}_{u=1\ldots m}$, and evaluate $l^2(\hat{p}_u(\hat{y}_0), q_u)$.

2. Based on the mean shift vector and equation (18), derive the new location $\hat{y}_1$ of the head $\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}$. Update $\{\hat{p}_u(\hat{y}_1)\}_{u=1..m}$, and evaluate $l^2(\hat{p}_u(\hat{y}_1), q_u)$.
3. While $l^2(\hat{p}_u(\hat{y}_1), q_u) > l^2(\hat{p}_u(\hat{y}_0), q_u)$, do $\hat{y}_1 \leftarrow \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$.
4. If $\|\hat{y}_1 - \hat{y}_0\| < e$, go to step 5. Otherwise set $\hat{y}_0 \leftarrow \hat{y}_1$ and go to step 1.
5. Around $\hat{y}_1$, a local search for the more accurate location and the best scale size of the elliptical head by maximizing the normalized sum of the image gradient magnitude around the perimeter of the ellipse.

5. Experiment

We test this method using three sequences with 128 × 96 pixels per frame. The first sequence has 75 frames and the other two have 40 frames. At first, before tracking we need to compute the target (head) model representation using multidimensional receptive field histograms with 11 × 11 × 11 bins, and the initial value for mean shift iteration in the current frame is given by the tracking result of previous frame, the initiation in the first frame is given by hand. The local search range for maximizing gradient magnification is ±2 pixels centered the result of mean shift iteration. The scale s adaptation scheme in our experiments is just achieved by modifying the short axis of the ellipse with ±1 pixel deviation.

The method has been implemented on PIII 500 PC machine using Visual C++ under Microsoft Windows, and it can be tracked with about 20 frames per second, the tracking results are presented in the following figure 1.

6. Conclusion

In this paper, one effective head-tracker is presented that is able to follow the people in the gray image sequences. Multidimensional receptive field histogram and ellipse shape model is adopted to describe the head. Mean shift algorithm is used to estimate the most probable location of the head in the current frame, in which the simple histogram matching is performed during mean shift iteration, then the elliptical head gradient model is used to locate the head accurately and to obtain the best scale size. The tracker works in real time and has been demonstrated to be robust to partial occlusion, view change, scale variations and so on. But it is fragile under the condition of big rotation and big occlusion.

Figure 1

Acknowledgments:

We would like to thank Stan Birchfield for that he permit us to download these test sequences from his Homepage: http://vision.stanford.edu/~birch

References:

[1] Dorin Comaniciu and Peter Meer, "Real-time tracking of non-rigid objects using Mean shift", *In the Proc.of the IEEE CVPR*, 2000, pp.142-149.

[2] Bernt Schiele and James L.Crowley, "Object recognition using multidimensional receptive field histograms", *In the Proc of ECCV*, 1996, pp.610-619.

[3] Stan Birchfield, "Elliptical head tracking using intensity gradient and color histograms", *In the Proc of the IEEE CVPR*, 1998, pp232-237.

[4] Dorin I.Comaniciu, "Nonparametric robust methods for computer vision", *PHD thesis*, Department of Electrical and Computer Engineering, Rutgers Univ, USA, January, 2000.
[5] M.J.Swain and D.H.Ballard. "Color indexing". *IJCV*, 7(1), 1991, pp11-32.
[6] K.Waters, J.Rehg, M.Loughlin, S.B.Kang andD.Terzopoulos. "Visual sensing of humans for active public interfaces". *Technical report*, Cambridge Research Lab, Digital Equipment Corporation, March 1996.
[7] C.R.Wren,A.Azarbayejani,T.Darrell, and A.P.Pentland. "Pfinder: Real-time tracking of the human body". *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol 19, No 7, July 1997, pp730-742.
[8] Paul Fieguth and Demetri Terzopoulos, "Color-based tracking of head and other mobile objects at video frame rates", *In the Proc of the IEEE CVPR*, 1997, pp 21-27.
[9] Y.Cheng, "Mean shift, mode seek, and clustering", *IEEE Transaction on Pattern Analysis And Machine Intelligence*, Vol 17,1995, pp790-799.
[10]H.P.Graf, M. Kocheeisen and E.Petajan, "Multi-model system for locating heads and faces", *In the Proc of the second Intl. Conference on Automatic Face and Gesture Recognition*, 1996, pp88-93.