

Research on action recognition method under mobile phone visual sensor

Wang Wenbin¹, Chen Ketang², Chen Liangliang³

Qiongzhou University, College Of Electronics And Information Engineering

Hainan Sanya 572022, China

Hainan province natural sciences fund: 613170

Keywords: mobile phone; sensor; action recognition

Abstract. During action recognition process under the mobile phone visual sensor, in order to ensure the speed of recognition, generally distributing as sloppy feature points, so as to express composite action feature needs to be recognized. However, if the action is composite feature in small area, traditional action recognition method based on sparse representation, when expressing action, is not able to express feature details within small area in high accuracy and achieve action recognition under mobile phone visual sensor accurately and effectively. To this end, an action recognition method under mobile phone visual sensor based on visual optimization analysis of behavior is proposed. With the maximum similarity clustering method to segment action images obtained by mobile phone vision sensor accurately, to exclude a large number of irrelevant features. By fitting the action clustering function to calculate principal curvature of action edge position, so as to acquire the position and scale of action. By adopting the action gray level co-occurrence matrix to apply adaptive classification for action, quickly complete the accurate identification of the action under mobile phone visual sensor. The simulation results show that, this algorithm has high stability and accuracy.

Introduction

With the development of computer hardware technology, human-computer interaction is developing toward various novel interaction ways [1,2]. Action as a way of interaction consistent with the living habits of people, depending on the characteristics like straight forward and visual, regarded as a very important part of human-computer interaction [3,4]. Therefore, action recognition gets more and more extensive attention, becomes the key topic for the related scholars to study [5].

During action recognition process under the mobile phone visual sensor, the process is vulnerable to the interference of objects with similar skin color and action, in the process of recognizing the action, in order to ensure the speed of recognition, generally distributing as sloppy feature points, so as to express composite action feature needs to be recognized[6,7]. However, if the action is composite feature in small area, traditional action recognition method based on sparse representation, when expressing action, is not able to express feature details within small area in high accuracy and achieve action recognition under mobile phone visual sensor accurately and effectively.

Aiming at the shortcoming of the traditional algorithm, an action recognition method under mobile phone visual sensor based on visual optimization analysis of behavior is proposed. With the maximum similarity clustering method to segment action images obtained by mobile phone vision sensor accurately, to exclude a large number of irrelevant features. By fitting the action clustering function to calculate principal curvature of action edge position, so as to acquire the position and scale of action. By adopting the action gray level co-occurrence matrix to apply adaptive classification for action, quickly complete the accurate identification of the action under mobile phone visual sensor. The simulation results show that, this algorithm has high stability and accuracy.

Action recognition principle based on sparse representation under mobile phone vision sensor

L1-LS algorithm is used to solve the sparse representation, this algorithm solves the l_1 norm minimization problem by least square method. Least square method have optimal statistics for normally distributed data, when the least square method is selected, usually the simplex method is adopted to complete the calculation of linear programming and the selecting weight iteration.

Assuming that training samples consisted of the action image under mobile phone visual sensor is $d_{ik} = \|k - v_i\|$, wherein, v_i , which is used to describe the input value of the training samples in the category i , subsets $\mu_{ik}^{(t)} = \frac{1}{\sum_{j=1}^c (d_{ik}^{(t)} / d_{jk}^{(t)})^{2/(m-1)}}$, $\forall i, k$ is used to linearly describe the new sample

$v_i^{(t+1)} = \frac{\sum_{k=0}^{L-1} (\mu_{ik}^{(t)})^m His(k) k}{\sum_{k=0}^{L-1} (\mu_{ik}^{(t)})^m His(k)}$, $\forall i$, consistent with the coefficient vector, then:

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (1)$$

However, in practical application, α which satisfies the above requirement may not exist, the coefficient β can be obtained by least square method:

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \end{aligned} \quad (2)$$

Among them, $\alpha = \gamma\beta$ is used to describe the parameter for adjusting sparse degree, if the above formula is met, then $\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r}$ can be approximately described by vector group $p(g_1, g_2) = \frac{p(g_1, g_2)}{R}$, $R = \begin{cases} N(N-1), \theta = 0, \text{ or } \theta = 90 \\ (N-1)^2, \theta = 45, \text{ or } \theta = 135 \end{cases}$ linear combination at the moment in the square error v . Then the optimal coefficient $ASM = \sum_{i=1}^k \sum_{j=1}^k (G(i, j))^2$ is calculated, after expansion, the above formula is:

$$CON = \sum_{n=0}^{k-1} n^2 \left\{ \sum_{|i-j|=n} G(i, j) \right\} \quad (3)$$

In the formula, $IDM = \sum_{i=1}^k \sum_{j=1}^k \frac{G(i, j)}{1 + (i-j)^2}$; $ENT = -\sum_{i=1}^k \sum_{j=1}^k G(i, j) \log G(i, j)$, Supposing that the derivative

$$\begin{aligned} u_i &= \sum_{i=1}^k \sum_{j=1}^k i \cdot G(i, j) \\ u_j &= \sum_{i=1}^k \sum_{j=1}^k j \cdot G(i, j) \\ s_i^2 &= \sum_{i=1}^k \sum_{j=1}^k G(i, j) (i - u_i)^2 \\ s_j^2 &= \sum_{i=1}^k \sum_{j=1}^k G(i, j) (j - u_j)^2 \end{aligned} \quad :$$

$$\begin{aligned} \eta^* &= K_{mm}^{-1} k_m(x) \\ \delta^* &= k(x, x) - k_m(x)^T \eta^* \end{aligned} \quad (4)$$

For the sample (x, y) , if $\delta^* > \gamma$, then the sample is unable to be linearly described through the current vector, the subset $B = B \cup \{(x, y)\}$ at the moment should be updated.

During action recognition process under the mobile phone visual sensor, the process is vulnerable to the interference of objects with similar skin color and action, resulting in proposed action recognition method based on sparse representation, cannot accurately and effectively implement the action recognition under mobile phone visual due to eliminate interference unsuccessfully.

The action recognition method adding behavior visual analysis under mobile phone vision sensor

Segmentation of action image collected under mobile phone vision sensor. Before recognizing the action under mobile phone visual sensor, to start with, the maximum similarity clustering algorithm is needed to segment the action in the action image, the action image is mapped in histogram space of gray.

The gray value of pixels (m,n) in the action image under mobile phone visual sensor is $f(m,n) \in \{0,1,\dots,L-1\}$, L is the gray level of the action image, by using the following formula can describe the clustering function of action:

$$J(U,V,k) = \sum_{k=1}^{L-1} \sum_{i=1}^c \mu_{ik}^m (d_{ik})^2 His(k) P_{k,j} \quad (5)$$

$$= (cx(j) - \alpha_{k,j} r(j), cy(j) - \beta_{k,j} r(j))$$

Among them, $k \in \{0,1,\dots,L-1\}$, $His(k)$ is used to describe the gray histogram of action image under mobile phone vision sensor, μ_{ik} is used to describe the metric value of k -th gray level belonging to the i -th cluster center, d_{ik} is utilized to describe the distance between k -th gray level and the i -th cluster center, $d_{ik} = \|k - v_i\|$. v_i is used to describe the gray level of i -th clustering center. The action object function under mobile phone vision sensor is minimized, can obtain the fuzzy membership matrix and the cluster center matrix of action:

$$\mu_{ik}^{(t)} = \frac{1}{\sum_{j=1}^c (d_{ik}^{(t)} / d_{jk}^{(t)})^{2/(m-1)}}, \forall i, k \quad (6)$$

$$v_i^{(t+1)} = \frac{\sum_{k=0}^{L-1} (\mu_{ik}^{(t)})^m His(k) k}{\sum_{k=0}^{L-1} (\mu_{ik}^{(t)})^m His(k)}, \forall i \quad (7)$$

According to the method described above, using the maximum similarity algorithm can efficiently segment the action in the action image under mobile phone vision sensor, the action is separated from the background image, thereby providing accurate data base for the identification of action under mobile phone vision sensor.

Determine the position and scale of the action under mobile phone vision sensor. Extreme of Gauss operators has larger principal curvature at the position in the action image where parallel to the edge of action under mobile phone vision sensor, and has the smaller principal curvature at the position which perpendicular to the edge position of action. The principal curvature of action can be calculated using the following formula:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (8)$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta$$

The principal curvature and characteristics value of action under mobile phone vision sensor are proportional, if α is the biggest characteristic value of action, β is the smallest characteristic value of action, the following formula can be obtained:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (9)$$

$$Det(H) = D_{xx} D_{yy} - (D_{xy})^2 = \alpha\beta$$

Setting $\alpha = r\beta$, then it has:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (10)$$

The above formula indicates that the formula is monotone increasing function, so the range of action under mobile phone vision sensor can be calculated based on the value of the formula, which can eliminate the influence caused by the interference point of action edge.

Through the above method to fit the action clustering function under mobile phone vision sensor, by calculating the principal curvature of edge position of action to realize the accurate calculation of

location and scale of action under mobile phone vision sensor.

Adaptive classification method of action under mobile phone vision sensor. Characteristics of the action under the mobile phone vision sensor can describe the inherent attribute of the action, so the different action characteristics can reflect the statistical characteristics of action image collected by mobile phone vision sensor collected from various aspects, which has become a hot topic in the research of action recognition. Gray co-occurrence matrix is an accurate method for image classification, during the process of action recognition, the ASM energy, contrast, entropy and correlation degree, IDM feature of the action can be adopted to extract and classify the action in the image collected by mobile phone effective vision sensor.

When $a=1, b=0$, the action in the action image acquired by mobile phone vision sensor is horizontal; when $a=0, b=1$, the action is vertical; when $a=1, b=1$, action is right diagonal; when $a=-1, b=1$, action is the left diagonal. In this way, the probability of the gray level of two action happens at the same time, will be able to translate space coordinate (x, y) into the description of (g_1, g_2) from the gray value of the action, and constitute the gray co-occurrence matrix of action.

With the following formula to normalize gray level co-occurrence matrix of action:

$$p(g_1, g_2) = \frac{p(g_1, g_2)}{R}, R = \begin{cases} N(N-1), \theta = 0, \text{ or } \theta = 90 \\ (N-1)^2, \theta = 45, \text{ or } \theta = 135 \end{cases} \quad (11)$$

The following formula is used to calculate the ASM energy of the action:

$$ASM = \sum_{i=1}^k \sum_{j=1}^k (G(i, j))^2 \quad (12)$$

The following formula is used to calculate the contrast of the action::

$$CON = \sum_{n=0}^{k-1} n^2 \left\{ \sum_{|i-j|=n} G(i, j) \right\} \quad (13)$$

The following formula is used to calculate the IDM features of the action::

$$IDM = \sum_{i=1}^k \sum_{j=1}^k \frac{G(i, j)}{1 + (i - j)^2} \quad (14)$$

The following formula is used to calculate the entropy of the action image::

$$ENT = - \sum_{i=1}^k \sum_{j=1}^k G(i, j) \log G(i, j) \quad (15)$$

The following formula is used to calculate the correlation between actions:

$$COR = \sum_{i=1}^k \sum_{j=1}^k \frac{(ij)G(i, j) - u_i u_j}{s_i s_j} \quad (16)$$

Among them:

$$\begin{aligned} u_i &= \sum_{i=1}^k \sum_{j=1}^k i \cdot G(i, j) \\ u_j &= \sum_{i=1}^k \sum_{j=1}^k j \cdot G(i, j) \\ s_i^2 &= \sum_{i=1}^k \sum_{j=1}^k G(i, j)(i - u_i)^2 \\ s_j^2 &= \sum_{i=1}^k \sum_{j=1}^k G(i, j)(j - u_j)^2 \end{aligned}$$

According to the method described above, gray level co-occurrence matrix of action can accurately classify the action in the image collected by mobile phone vision sensor, so as to complete the identification of the action under mobile phone vision sensor.

Analysis of simulation experiments

In order to verify the validity of the method in this paper, there is the need for related experiments. With the proposed method and the traditional method to make 100 times recognition for 6 kinds of actions, the obtained recognition rate are described in Table 1.

Table 1 Comparison of recognition rate obtained with two kinds of methods

Action number	recognition rate of the method in this paper (%)	recognition rate of the traditional method (%)
1	90	88
2	92	86
3	91	88
4	92	89
5	90	87
6	93	88

It can be seen from Table 1 that the action recognition rate under mobile phone visual sensor obtained with the proposed method was significantly higher than with the traditional method, the effectiveness of the proposed method is testified.

Based on the recognition rate obtained with two methods, accuracy of the recognition of these method were counted, the results described as below:

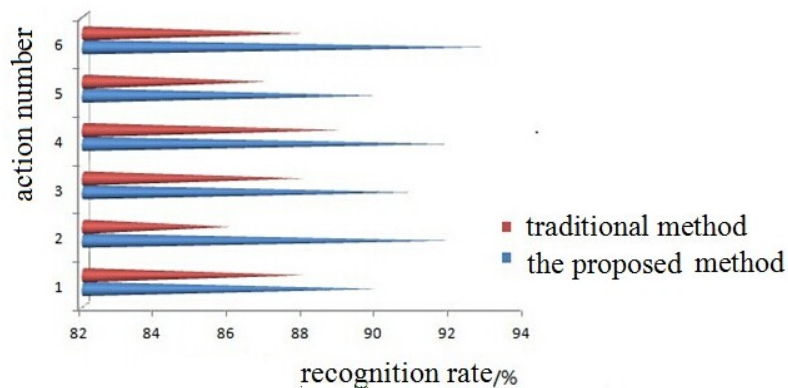


Figure 1 the accuracy of identification obtained with the proposed method and the traditional method

It can be seen that the proposed method have higher accuracy of identification than the traditional method, and the wave is more stable, which verify the stability and accuracy of the method proposed in this paper.

Conclusions

An action recognition method under mobile phone visual sensor based on visual optimization analysis of behavior is proposed in this paper. With the maximum similarity clustering method to segment action images obtained by mobile phone vision sensor accurately, to exclude a large number of irrelevant features. By fitting the action clustering function to calculate principal curvature of action edge position, so as to acquire the position and scale of action. By adopting the action gray level co-occurrence matrix to apply adaptive classification for action, quickly complete the accurate identification of the action under mobile phone visual sensor. The simulation results show that, this algorithm has high stability and accuracy.

References

- [1] Ren Haibing, Zhu Yuanxin, Xu Guangyou, et al. The gesture recognition segmentation and recognition in complex background [J]. *Acta automatica Sinica*, 2002.28 (2): 256-261.
- [2] Wang Changxi, Yang Xianjun, Xu Qiang, et al. Upper limb movement recognition system based on 3D acceleration sensors [J]. *Chinese Optics Letters*, 2010.6:816-819.

- [3] Chen Shenglei, Chen Geng, Xue Hui. Rarefaction method research based on least squares support vector machine classification [J]. Computer Engineering, 2011, 11 (22): 147-150.
- [4] Xiong Juntao, Liu Zijian, Sun Baoxia, et al. The gesture tracking and action recognition algorithm based on visual technology [J]. Computer and modernization, 2014.7:75-79
- [5] Huang Yuefei, Xu Guangyou. Independent action recognition of angle of view [J]. Journal of software, 2008.7:1623-1634.
- [6] Fu Sihua, Zhang Xiaohu. Real time detection method of moving target based on sequential image [J]. Optical technique, 2004, 30 (2): 215-217.
- [7] Xiao Ling, Li Renfa, Luo Juan. A human action recognition method based on compressed sensing in body area network [J]. Journal of electronics and information technology, 2013.1:119-125.