The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

Simultaneous Recognition of Multiple Actions using CHLAC features

Kyosuke Masumitsu, Takayasu Fuchida

Kagoshima University, Japan
(Tel: 099-285-3408, Fax: 099-285-3408)
sc108067@ics.kagoshima-u.ac.jp

Abstract: In this study, we propose simultaneous recognition of multiple actions in moving image using sub-space method and CHLAC features. As the CHLAC features are integral value of entire the image, it has the additivity. Thus, when multiple objects exist, it is possible to use the added-feature which is made by adding the each action of the object. But if you use such an added-feature, it becomes difficult to distinguish each action respectively. Then we introduce a method to trace the region of each object in the image. From the computer simulation by using these 8 kinds of moving images, we show that our proposed method can recognize these 8 kinds of actions.

Keywords: CHLAC, Color-HLAC, CLAFIC method, Projection distance method.

1 Introduction

Recently, as the performance of computers improves, the demand of action recognition using features of moving image has risen and different method has been proposed. In recent features descriptor methods, without descripting the features from one only image in the moving image, many methods that utilize time and space that consider the current frame, the previous and the next frame is proposed[1][2][3]. CHLAC(Cubic Higher-order Local Auto Correlation)[3] is one of the amount of image feature. This is what Higher-order Local Auto Correlation[4] is extended and can simultaneously evaluate features of both moving and shape reflected in the camera. This features extraction method is, in general, performed by calculating the frequency of matching between the pixels which has moved in the three binary images obtained by the frame difference from four successive frames in the moving image and the mask patterns prepared in advance. Consequently, the general CHLAC feature has 251 dimensions.

As the CHLAC features have the additivity, when multiple objects exist, it is possible to use the added-feature which is made by adding the each action of the object. It can be simultaneously multiple objects using the added-features, but when the number of object increases, variation becomes enormous and configuring the discriminator is difficult. In previous research which resolves the problem[5], the method using multiple regression analysis has been proposed and it enables simultaneous recognition of multiple moving objects. But in this method, it isn't able to current location of the object and to track the object. So, in this study, we propose the method which traces the region of each object in the image and extract CHLAC features from each object and distinguish the moving.

We used color-information for tracking the objects. We adopted the features which combined Color-HLAC[6] and HLAC for this information.

We used sub-space method to the discrimination of motion. This is one of the pattern recognition methods which utilizes the bias of input vectors and takes relatively small amount of calculation at the time of the identification. So we think that it suits the motion recognition which requires the real-time processing. In this research, we adopt the sub-space method which uses the bias of distribution of the CHLAC features which are extracted for each motion of objects in the movies.

And, we validate the effectiveness of our method through the computer simulations using some movies which include multiple objects in them.

2 extraction of CHLAC features

We will easily explain how to extract CHLAC.

2.1 HLAC feature

HLAC(Higher-order Local Auto Correlation)[4] is the amount of statistical features based on Higher-order correlation the N dimensional autocorrelation for $f(r)$ is expressed as following equation.

$$\mathrm{x}(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_N) = \int f(\boldsymbol{r}) f(\boldsymbol{r} + \boldsymbol{a}_1) ... f(\boldsymbol{r} + \boldsymbol{a}_N) dr \quad (1)$$

Here, f is the image which is retrieved in time series and r is the reference position and $\boldsymbol{a}_i$ is the vector that indicates the displacement. Since the correlation between the pixel of interest and its neighboring regions is expected to be high, we usually will not calculate the correlation for all areas, but for only neighbor pixels, this feature is usually set as degree－$N = 2$. Consequently, it has 25 dimensions.

2.2 Extraction of CHLAC features

We will extract the features which are used to recognize motion from moving image. First, we convert the moving image to YCbCr color-space in order to remove the illumination variation, and then remove component Y of luminance signal. After this, we perform smoothing processing in order to reduce the noise in the image. Next, we make frame differences from moving image and create the subtraction binary images and then calculate CHLAC feature from its subtraction binary images.

HLAC which we described in sub-section 2.1 is the correlation in the two-dimensional local region, but CHLAC is feature value which introduces the time axis to the local area and then calculates correlation in the local region of $3 \times 3 \times 3$.

In general, degree N is set {0, 1, 2} and the feature has 251 dimensional vector.

3 Tracing the region of the object

In this chapter, we will explain about tracing the region of the object and tracking using color-information.

3.1 Tracing the region of object and tracking

In order to simultaneously recognize each location of multiple objects and motion, we trace the each region of objects. First, we prepare the background image in advance. Next, we extract the subtraction image between the current frame and background image. We divide the whole of subtraction image into the many vertical narrow rectangles and then check the all pixels in each segmented region whether there are any differences. If the different pixel is found in the segmented region, its region becomes a candidate for the presence of the object region and we will check the adjacent pixels of this region.

If the candidate of the object is found in the rectangle which divides the image lengthwise, we divide the candidate region horizontally and explore the region with the subtraction as same as the vertical segmentation.

Moreover, since we use the difference image between two sequential frames in calculating CHLAC feature, we are able to make a more accurate segmentation of the object by examining whether there is a pixel which has motion in the subtraction image within the candidate region of the object. If there is an area of noise, we update the previous pixel value to the current pixel value in order to denoising.

If the presence of the object is confirmed, we register the color-information of the object in order to distinguish between each object of the next frame, and in order to track the object. We have been utilized color-histogram and color-HLAC to the color-feature. Then distinction between objects is performed by Euclidean distance and then we calculate the subtraction between the features quantity of registered object and the features quantity obtained from the current object region. The region which has smallest difference becomes the updated position of the object, and we update the previous feature value.

3.2 Features amount by combining the color-HLAC and HLAC

Color-HLAC is the feature which is extended higher-order local auto correlation to color image. Similarly HLAC, we will consider mask pattern of 3×3 sizes. This is expressed as the combination of each color component in the local region of pixel of interest. In this study, we used 45 dimensional features that have degree {0, 1} only.

In addition, we also added HLAC considering the luminance. Consequently, the feature amount of a combination of the HLAC and color-HLAC becomes 70 dimensions.

4 Motion determination using sub-space method

In this research, we use sub-space method which is the multi-class classification for the input patterns.

4.1 CLAFIC method and projection distance method

Sub-space method is to calculate sub-space of each learning class using bias of distribution of the feature and recognize input pattern by projecting it onto the sub-space.

CLAFIC method is one of the sub-space methods. CLAFIC is, at the time of learning, a method which puts the origin of sub-space on the origin of original feature space and performs KL expansion for each class and then determines sub-space of each class. Then, when an unknown pattern is input, its pattern class is determined as the sub-space class which the length of the projection vector for each class is maximum.

We will describe how to configure the sub-space by this method.

We calculate autocorrelation matrix R_i using CHLAC feature vector $\boldsymbol{x}_{i,j} \in \mathrm{C}_i (1 \leq \mathrm{j} \leq \mathrm{N}_i)$ which belongs to the class C_i as the learning data.

The feature vector is normalized so that its norm becomes 1 using $\boldsymbol{x}_{i,j}/\|\boldsymbol{x}_{N_i}\|$.

$$R_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{x}_{i,j} \tag{2}$$

Sub-space is composed using eigenvalues $\lambda_{i,j}(\lambda_{i,1} \cdots \lambda_{i,251})$ obtained from the eigenvalue decomposition of the autocorrelation matrix R_i and the corresponding part of the eigenvectors $\boldsymbol{u}_{i,1}, \cdots, \boldsymbol{u}_{i,251}$.

In order to compose the sub-space, we calculate principal component of the consisted sub-space.

To set the number of principal components, we use a cumulative contribution ratio $a(d_i)$ of the dimension d_i shown in equation (3).

$$a(d_i) = \frac{\sum_{j=1}^{d_i} \lambda_{i,j}}{\sum_{j=1}^{251} \lambda_{i,j}} \quad (3)$$

d_i is selected to satisfy the following conditions,

$$a(d_i) < k_d < a(d_i + 1) \quad (4)$$

where k_d is the threshold and in this study, we use $k_d = 0.998$. After determining d_i, we select the d_i eigenvectors from the top of the value. Then, we calculate projection matrix P_i of sub-space L_i using equation (5).

$$P_i = [\boldsymbol{u}_{i,1}, \cdots, \boldsymbol{u}_{i,di}] \quad (5)$$

In contrast to CLAFIC method, projection distance method is the method which the center of each class is the origin.

While CLAFIC method perform eigenvalue decomposition of autocorrelation matrix obtained by the equation (2), projection distance method calculate average of feature vectors of each class and calculate the variance-covariance matrix R_i according to the following formulas. And sub-space S_i using projection distance method is configured by eigenvalue decomposition of this matrix.

$$\boldsymbol{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{x}_{i,j} \quad (6)$$

$$R_i = \frac{1}{N_i} \sum_{j=1}^{N_i} (\boldsymbol{x}_{i,j} - \boldsymbol{m}_i)^t (\boldsymbol{x}_{i,j} - \boldsymbol{m}_i) \quad (7)$$

$$S_i = [\boldsymbol{u}_{i,1}, \cdots, \boldsymbol{u}_{i,di}] \quad (8)$$

4.2 Recognition

In this section, we will explain how to identify the unknown pattern $\boldsymbol{q}$ according to each method of CLAFIC method and projection distance method.

Recognition of unknown pattern $\boldsymbol{q}$ using CLAFIC method is executed by calculating the square of the length of $P_i\boldsymbol{q}$ as the degree of similarity between the unknown pattern $\boldsymbol{q}$ and the class C_i by using projection matrix P_i which is determined for each class. Sub-space which projection length is greatest becomes unknown pattern that determines the class of.

$$argmax_i\{D_i\} = \|P_i\boldsymbol{q}\|^2 = \boldsymbol{q}^t P_i \boldsymbol{q} = \boldsymbol{q}^t\left(\sum_{j=1}^{d_i}(\boldsymbol{u}_{ij}^t\boldsymbol{u}_{ij})\right)\boldsymbol{q} = \sum_{j=1}^{d_i}\left(\boldsymbol{q}^t\boldsymbol{u}_{ij}\right)^2 \quad (9)$$

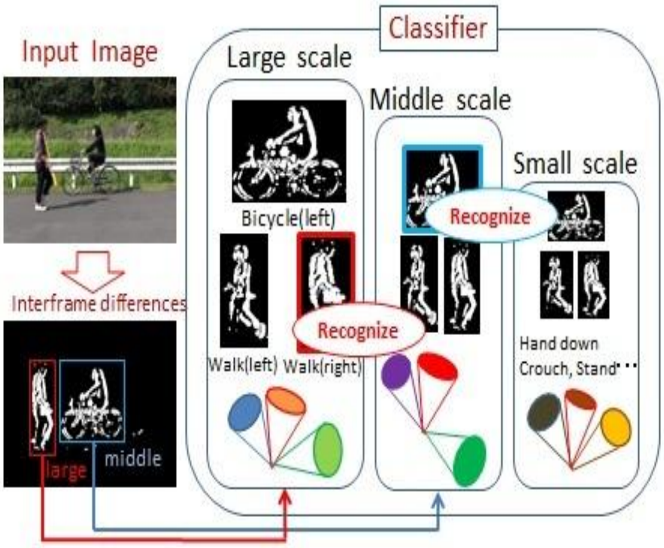

Figure 1. Corresponding to the scaling

In recognition of projection distance method, the total of projection distance to the sub-space is calculated by following:

$$argmin_i\{D_i\} = \sum_{i=1}^{n}\left\{\|\boldsymbol{q} - \boldsymbol{m}_i\| - \left\|{S_i}^t(\boldsymbol{q} - \boldsymbol{m}_i)\right\|^2\right\} \quad (10).$$

The class which has smallest D_i is the decision class of the unknown pattern.

4.3 Configuring the discriminator corresponding to the scaling

CHLAC isn't invariant feature quantity to scaling.

Therefore, we configured the sub-space for each scale of object at the time of learning. We configured to determine whether any portion of the scale corresponding to the space according to the size of extracted region of the object (Figure 1).

5 Experiments

To evaluate the effectiveness of our proposed method, we have conducted some experiments.

5.1 Conditions of the experiments

We have set up a camera on the street and have taken the video which was assumed to detect each motion and location of the multiple entering objects. The video was taken with the state of the objects entering the lateral of camera.

We prepared 8 kinds of moving images, such like "bicycle left", "bicycle right", "walk left", "walk right", "crouch", "stand up", "raise hand" and "hand down".

In addition, preparing video such that maximum three objects exist, we have experimented using them.

The experiments were conducted with six subjects. The motion at each scale was learned two per person. Moving images were taken with 30fps and size of them was set to 320×240 pixels.

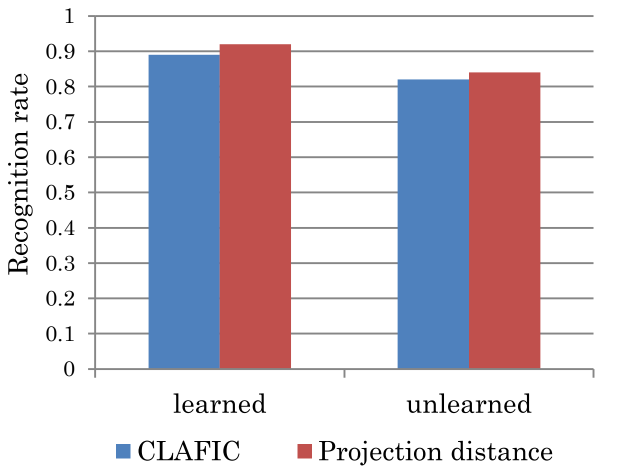

Figure 2. Action recognition rate by each classifier

5.2 Experiment using only objects who were learned

We have carried out an experiment using two videos in which only learned persons exist. Each video has the content that at most three persons are entering.

5.3 Experiment using only objects who were not learned

We have taken two videos which were consists of only persons who were not learned, and conducted an experiment. Each video has the content that at most three persons are entering.

5.4 Result

The success rate of segmentation and object tracking was 95.4%. There was a slight mistake when an object enters and when the object goes out the screen. However, at the time of occlusion between the objects occurred, we have confirmed that detecting the regions of the objects were performed well and also tracking was performed correctly after the occlusion.

The action recognition success rate using each classifier shows at the Figure 2. On the process of identification, mainly erroneous identification occurred when the object entered and went out. In addition, we confirmed that incorrect identification rate didn't change much if an occlusion occurs or not.

6 Conclusion

In this research, we proposed the method for recognizing the location and the motion of multiple objects in moving images captured from a camera which was set on the street. In our proposed method, first, we extract a difference image between one frame image and background image that has been prepared beforehand, and extract regions of objects in the image. Then, we performed the motion recognition by sub-space method that matches the size of each object region. As the feature quantity of the motion recognition, we used CHLAC obtained from the binary image sequence by frame difference. Then, we used CLAFIC method and projection distance method as the classifier. And we utilized a combination of feature quantity color-CHLAC and HLAC.

We have performed two evaluation experiments for videos that have eight types of behavior and at most three people who consist of person and bicycle. For the video of learned objects, the result of CLAFIC method was 89%, and the result of projection distance method was 92%. In addition, for the video of unlearned objects, the result of CLAFIC method was 82%, and the result of projection distance method was 84%.

In this study, we proposed a method that corresponds to the scaling and have confirmed its effectiveness, but we have not been able to respond to changes in speed yet. So, for the future work, we also need to propose a method that corresponds to the action speed of the object. And many type of improved sub-space method are proposed, we are able to consider the implement of variant of sub-space method.

And we think that the accuracy of motion recognition will be improved using the variant of CHLAC[7] which is extended to cross correlation which takes the timing of the motion of an object into consideration.

REFERENCES

[1] Yeffet, L., Wolf, L., Local trinary patterns for human action recognition. In: ICCV, 2009
[2] Nazli, I., Stan, S., Object, Scene and Actions: Combining Multiple Features for Human Action Recognition, ECCV, 2010
[3] T.Kobayashi, and N.Otsu, Three-way auto-correlation approach to motion recognition, Pattern Recognition Letters, Volume 30, Issue 3, pp.212 221, 2009.
[4] N.Otsu and T.Kurita, "A new scheme for practical, flexible and intelligent vision systems", Proc.IAPR Workshop on Computer Vision, pp431-435, 1988.
[5] Y.Simohata, N.Otsu, "Simultaneous Recognition of Multiple Moving Objects Using Cubic Higher-Order Local Autocorrelation", PRMU, 2006.
[6] T.Kurita, T.Kato, I.Fukuda, A.Sakakura, "Sense Retrieval on a Image Database of Full Color Paintings", Information Processing Society of Japan Vol.33, 1992.
[7] T.Matsukawa, T.Kurita, "Action recognition using cu bic cross-correlations features of local motion attribute s", PRMU2009-285, 2010.