Journal of Information Processing Vol. 19 190–200 (Apr. 2011)

Regular Paper

New Techniques of Foreground Detection, Segmentation and Density Estimation for Crowded Objects Motion Analysis

Wei Li,[†1] Xiaojuan Wu[†1] and Hua-An Zhao[†2]

Now video surveillance systems are being widely used, the capability of extracting moving objects and estimating moving object density from video sequences is indispensable for these systems. This paper proposes some new techniques of crowded objects motion analysis (COMA) to deal with crowded objects scenes, which consist of three parts: background removal, foreground segmentation, and crowded objects density estimation. To obtain optimal foregrounds, a combination approach of Lucas-Kanade optical flow and Gaussian background subtraction is proposed. For foreground segmentation, we put forward an optical flow clustering approach, which segments different crowded object flows, and then a block absorption approach to deal with the small blocks produced during clustering. Finally, we extract a set of 15 features from the foreground flows and estimate the density of each foreground flow. We employ self organizing maps to reduce the dimensions of the feature vector and to be a final classifier. Some experimental results prove that the proposed technique is useful and efficient.

1. Introduction

At present, video-surveillance systems are widely used in banks, highways, hotels and other places. With the development of computer vision and pattern recognition, more and more video-surveillance systems employ crowded objects motion analysis (COMA) techniques. COMA techniques include object detection, foreground segmentation, abnormal behavior detections, and object density estimation. COMA is attracting more and more interest [1].

There exist several challenges in the field of COMA. (1) How to detect an optimal foreground which contains all of the targets without any noise. Because some slow moving objects are easily classified as background objects, and crowded objects scenes always contain much noise, such as shaking leaves, weather changing, windows, and so on, getting an optimal foreground is very difficult. (2) How to segment a crowded objects scene into several parts with objects moving in different directions. It is hard to find out the exact watershed of moving targets in a complex crowded objects scene. (3) How to extract available features from a crowded objects scene for detecting abnormal behavior or estimating crowd density.

In this paper, we propose three new techniques to solve these three challenges respectively. First, we propose a background removal approach which combines dense optical flow with a Gaussian background model. Using this approach, most noise is eliminated and an optimal foreground can be obtained. Second, we propose a clustering approach based on optical flow, which is employed to segment the foreground into different flows. This clustering is similar to K-means clustering, but our approach involves supervised learning classification and can be used to segment both regular moving videos and irregular moving videos. Because the optical flow is not exact and some misclassification blocks appear during clustering, we propose a block-absorption approach to solve this problem. Third, according to texture analysis and moment analysis on foreground images, we extract a set of features to estimate the flow density and classify the flows into different groups.

This paper is organized as follows. Section 2 discusses some current works on COMA. Section 3 describes the proposed techniques in detail. Some experimental results and discussion are shown in Section 4. Finally, the paper is concluded in Section 5.

2. Related Works

Currently, in the area of background removal, there are three common methods. The first is frame differencing which calculates difference in gray of each pixel using the two, or more, adjacent frames, and then determines the foreground area by setting a threshold. Based on this idea, some improved methods have been proposed [2),3)]. Frame differencing is effective for dynamic backgrounds, but when the objects are moving fast or slowly, the detection results are not good. The

†1 School of Information Science and Engineering, Shandong University, China
†2 Department of Computer Science and Electrical Engineering, Kumamoto University, Japan

second is background subtraction which calculates difference in gray between the current image and the background image, then detects the foreground area by setting a threshold. Generally, there are two methods to obtain a background image. (1) Manually appointing one image as the background. (2) Training background using images, such as the Gaussian background model and so on [4),5)]. These methods are good at dealing with complex backgrounds, but are sensitive to illumination changes and always produce much noise. The third is an optical flow method which calculates the optical flow field instead of the velocity field [6)]. However, the calculation of optical flow is complex, and this method is sensitive to noise. There are several kinds of optical flow methods, of which the Lucas-Kanade method is widely used to detect crowded objects. The weakness of this method is that it produces much noise because it is sensitive to illumination changes.

In general, foreground segmentation can be divided into two categories: static image segmentation and video segmentation. We only consider the later in this paper. Lauer employed spectral clustering of linear subspaces for motion segmentation [7)]. This method tracks feature points of motion area and segments different objects through spectral embedding and clustering of linear subspace. Brendel carried out video object segmentation by tracking regions [8)], the author assumed that moving objects occupied the same size of region in each frame. Huang proposed hypergraph cut method, which initially over-segment image into small patches, then a hypergraph structure is built to represent the complex spatio-temporal neighborhood relationship among the patches, finally, the patches with the same attribute value are combined together [9)].

The research on COMA mainly concerns two fields, abnormal behavior detection and crowded objects density estimation. In the first field, the general approach consists of modeling normal behaviors, then estimating the deviations between the normal behavior model and the observed behaviors. If the deviation is larger than threshold, that means abnormal behavior appears [10)–12)]. In the second field, because crowded objects density is an important feature and crowded objects of high density should receive more attention. Therefore, density estimation is attracting more and more interests. Recently, Chan adopted the mixture of dynamic texture to segment the crowds moving in different directions and extracted a set of 28 features to estimate the number of pedestrians in each flow [13),14)]. Ma derived a mathematical relation for geometric correction for the ground plane [15)]. A linear relation between the number of pixels and number of people was derived by applying the geometric correction. Both the methods assumed that the number of foreground pixels are proportional to the number of targets, which is only true when there are no serious occlusions among the targets. Marana proposed a real-time crowd density estimation approach using texture analysis in which a set of features could be extracted from gray-level co-occurrence matrix (GLCM) [16)]. H. Rahmalan proposed translation invariant orthonoraml Chebyshev moments (TIOCM) which is improved from orthonormal Chebyshev moments (OCM), and then the author employed TIOCM in crowd density estimation and made a comparison between GLCM and TIOCM [17)]. A useful evaluation parameter is the rate of true classification (RoTC), the average RoTC of GLCM and TIOCM are 70% and 80%. There are two reasons for the low RoTC, one is background noise, the other is some features have little relationship with crowded objects density.

3. Proposed COMA Technique

Our approach is composed of three parts, background removal, foreground segmentation, density estimation and flow classification. In this section, we will illustrate each part in detail.

3.1 Background Removal

Background removal or foreground detection is the basis of our framework, the result of foreground detection will directly decide the performance of foreground segmentation and density estimation. Because crowded objects movement is wide-area, it is necessary to detect the change on every pixels, so Lucas-Kanade (LK) optical flow is the best choice. If $I(x, y, t)$ is the intensity of pixel $m(x, y)$ at time t, $v_m = [v_x, v_y]$ is the velocity vector of pixel $m(x, y)$, then after a short time interval Δt, the optical flow constrain equation is

$$\nabla I \cdot v_m + \frac{\partial I}{\partial t} = 0 \tag{1}$$

where $\nabla I = [\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}]^T$ is the spatial intensity gradient vector. Because v_m is 2-dimension variable, we need more constraints to settle this question. So far there are a lot of methods to solve Eq. (1) where LK optical flow is a quite well

solution. It estimates v_m by v expressed in Eq. (2) on the assumption that v_m is a constant in a small spatial neighborhood Ω.

$$\sum_{m\in\Omega} W^2(m)\left(\nabla I\cdot v+\frac{\partial I}{\partial t}\right)^2. \tag{2}$$

In Eq. (2), $W^2(m)$ is a window function making the central part of the neighborhood has greater weight than the peripheral part. For the pixels m_i $(i=1,2,\cdots,n)$ in Ω, the solution v of Eq. (2) can be obtained by

$$v=(A^TW^2A)^{-1}A^TW^2b \tag{3}$$

where

$$A=(\nabla I(m_1),\cdots,\nabla I(m_n))^T,$$
$$W=diag(W(m_1),\cdots,W(m_n))$$

and

$$b=-\left(\frac{\partial I(m_1)}{\partial t},\cdots,\frac{\partial I(m_n)}{\partial t}\right)^T.$$

As mentioned, because LK method calculates optical flow on every pixels, it can detect all the changes between adjacent images. However, optical flow methods are very sensitive to illumination change, it is difficult to find a proper threshold to segment foreground and background by LK method. In fact, no matter how to make a choice, the detection result may either lose some foreground area or contain some background noises. Obviously we can not obtain an optimal foreground by LK method. Therefore, we present a new approach by combining LK optical flow with Gaussian background subtract (GBS) method to get an optimal foreground [18], we name this approach optical flow and background model (OFBM).

The proposed background removal approach OFBM is shown in **Fig. 1**, in which we apply LK optical flow and GBS in parallel. On the one hand, we firstly use the two adjacent images $f(x,y,t-1)$ and $f(x,y,t)$ to calculate the LK optical flow field, then median filter and Gaussian filter are used to eliminate high-frequency noises and salt and pepper noises respectively. After that we use a threshold T_{lk} to segment optical flow field to get LK foreground mask $flk(x,y,t)$. Since the optical flow vector on each pixel has magnitude and phase values, we only make

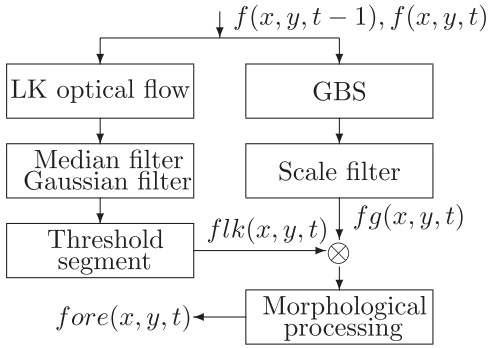


Fig. 1 Outline of proposed background removal approach OFBM.

use of magnitude information to segment foreground, all pixels which have lower magnitude value than T_{lk} will be classified as foreground. Our test results show the range of T_{lk} is $[0.05, 0.20]$, choosing smaller T_{lk} will produce larger foreground area including background noises, while choosing bigger threshold may lose some foreground area. In order to detect all the movement area we select the smallest value 0.05, then we try to eliminate the noises in the foreground mask $flk(x,y,t)$. On the other hand, GBS method is used to get another foreground mask where the scale filter is employed for segmenting foreground and background. In the scale filter, we set another threshold T_g which means block area. For an obtained foreground image, if a pixel block has smaller size than T_g, it will be classified as background, otherwise it is kept as foreground. Hence, we can get another foreground mask $fg(x,y,t)$. The value of T_g should be smaller than any single object, so that $fg(x,y,t)$ can keep all foreground area. As like LK method, we choose the smallest T_g to obtain the largest foreground mask $fg(x,y,t)$.

Finally, we take logical conjunction on the two masks and operate morphological processing to joint the adjacent areas and exclude small blocks in the foreground [19], then an optimal foreground $fore(x,y,t)$ can be obtained as shown in Fig. 1. Noted that though both $flk(x,y,t)$ and $fg(x,y,t)$ contain noises, the noise in $flk(x,y,t)$ is caused by brightness alteration and randomly appears on the profiles of objects, the noise in $fg(x,y,t)$ occurs on the edge of objects and appears at the same place. Because the two noises appear at different place, we

can eliminate most background noises by using logical conjunction processing.

3.2 Foreground Segmentation

Generally, objects in foreground move toward different directions, in another word, foreground area consists many flows. As shown in **Fig. 2**, (a) shows people in a Marathon game and (b) shows pedestrians at crosswalk. It can be seen that Fig. 2 (a) contains two flows, top-to-bottom flow by red color and bottom-to-top flow by blue color. Figure 2 (b) contains three flows, people crossing the pavement in opposite directions and people walking along the street. To analyze the movement in different flows, it is necessary to segment foreground. Since we use LK optical flow to detect foreground and optical flow field implies movement on every pixels, we propose a new approach to segment optical flow field.

K-means clustering is widely used in image segmentation and data analysis, which is useful tool to classify samples with different properties into different groups [20]. After the initial setting, this method can be used to deal with large amount of data in short time. In our framework, we apply K-means method to segment foreground optical flow field. We denote K is the initial number of clustering centers, $\mathbf{v} = (u, v)^T$ is the velocity vector and θ is the angle of $\mathbf{v}$, $0° \le \theta < 360°$, N is the number of foreground pixels. Initial clustering centers are $\{0°, 360°/K, 2 \times 360°/K, \cdots, (K-1) \times 360°/K\}$, threshold of clustering $\alpha = 360°/(2K)$, threshold of clustering combination $\beta = 360°/K$, times of clustering $T = 20$. The algorithm is as follows:

Step 1 Initial clustering.

for $1 \le i \le N$

if $|\theta_i - c_j| \le \alpha, c_j \in \{0°, 360°/K, 2 \times 360°/K, \cdots, (K-1) \times 360°/K\}$

$\mathbf{v}_i(u, v)^T$ is classified to clustering j;

the number of samples of clustering j n_j is added 1;

Step 2 New clustering centers calculation.

for $1 \le j \le K$

$c'_j = \sum_{i=1}^{n_j} \theta_i / n_j$;

Step 3 Combining similar clusterings.

if $|c'_p - c'_q| \le \beta, 1 \le p, q \le K$

$c'_q = \sum_{i=1}^{n_p+n_q} \theta_i / (n_p + n_q)$;

$c'_p = c'_q$;

else

end of clustering.

Step 4 Return to Step 1 and restart clustering.

This approach can be used to segment all kinds of images, for images with complicated movement, objects moving towards every directions, we assign K a large value, such as 16 or 32. If objects move regularly, like Fig. 2, we assign K a small value, such as 4. In our test, we can segment most images with $K = 4$. **Figure 3** shows the framework of proposed approach, it can be seen that, many

(a) (b)

Fig. 2 Examples of crowds with different flows.

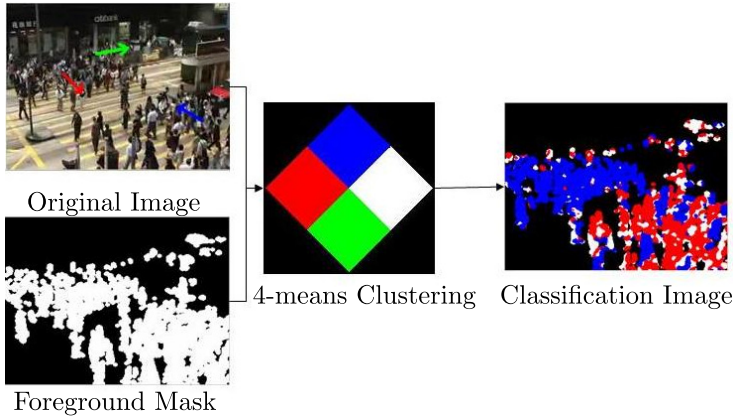


Fig. 3 A framework of K-means clustering on foreground image.

small patches appear after segmentation. There are three reasons. (1) LK method calculates optical flow on every pixels, if some object was so large that the moving distance between adjacent frames was less than the size of the object, optical flow of some pixels on the object may be not same as the movement. (2) When two neighboring objects moving toward different directions, optical flow on the edge of objects may be chaotic. (3) There exists calculation errors in optical flow field.

To solve this problem, we propose a block-absorb approach, which is shown in **Fig. 4**. It can be seen that after clustering the foreground is segmented into three parts, red, blue and white, but the small blocks make the clustering result disordered. The block-absorb approach is proposed to remove these small blocks. It is easy to understand two conjoint blocks with the same color become one bigger block, so we try to absorb the blocks by changing the colors. Considering the size of single object, we set a block area threshold $L = 200$, all the blocks smaller than L will be absorbed, except isolated blocks that are surrounded by background. Then we arrange and count the blocks, and denote M as the number of blocks. Next, we change the color of one block and count the number of blocks again. If

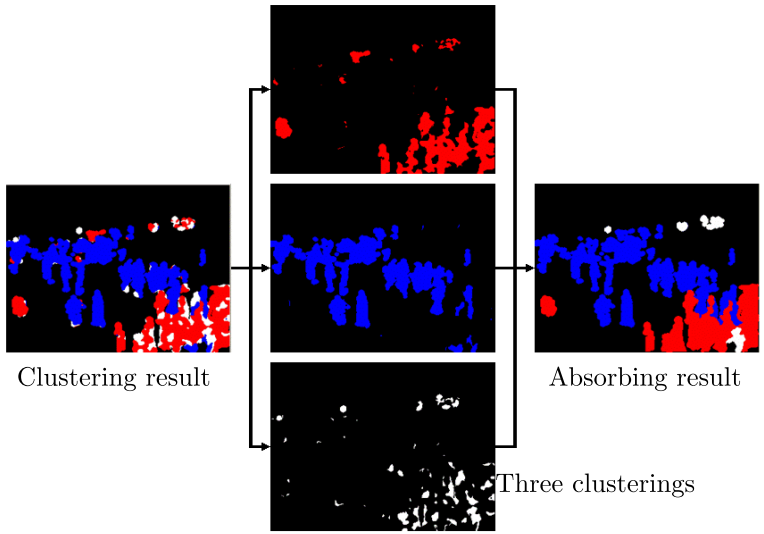


Fig. 4 A framework of block-absorb approach.

we found M changing to $M - 1$, that means this block is absorbed successfully, if not, we change to another color. Since at most there are three colors in the foreground, so we can absorb one block by twice color-change. Then the next block is going to be absorbed, until $M = 0$. After 5 times block-absorb, all the small patches disappear. The result of this approach can be seen in Fig. 4.

3.3 Density Estimation

The final goal of COMA is to extract features from foreground and analyze objects movement [1]. As above mentioned, abnormal detection and density estimation are two approaches of motion analysis. In this paper, we focus on the second approach, crowded objects density estimation and flow classification. Objects density is special characteristic of foreground image, based on the result of foreground segmentation, we estimate objects density on each flow respectively.

3.3.1 Definition of Classification

To carry out COMA exactly, crowded objects with different densities should be classified and received different level of attention. Noted that for different size of targets, we should define different classifications. For example, human is smaller than car but larger than fish, it is improper to use the same standards to classify human flow, car flow and fish flow. In this paper we take human videos and car videos for test, so we define two kinds of classification. For human videos, we use the same standard as Rahmalan proposed, which is shown in **Table 1** [16]. **Table 2** shows the classification definition for car videos. It can be seen that,

Table 1 Classification of human flows.

Level	Range of people	Group
A	<20	Very Low Density
B	20–40	Low Density
C	40–60	Moderate
D	60–80	High Density
E	>80	Very High Density

Table 2 Classification of car flows.

Level	Range of cars	Group
A	<10	Low Density
B	10–20	Moderate
C	>20	High Density

human flows are classified into five groups and car flows are classified into three groups.

3.3.2 Feature Extraction

To estimate crowded objects density, we should exact efficient features from each flow. The existing methods extracted texture feature [16] or moment feature [17] for estimation and the results are not satisfactory, because each of the features cannot indicate density exactly. For improving accuracy of density estimation, we propose a new approach that from each flow, as many as four kinds of features are extracted and all of them are closely related to crowded objects density. These features are illustrated as follows:

Area feature: *Area* means total number of pixels in each flow, large foreground area implies high density objects.

Edge feature: *Edge* means total number of edge pixels in each flow. It is easy to understand that high density crowds contain complex edge information.

Moment feature: Moment feature $M_{00} = \sum_{i=1}^{m} \sum_{j=1}^{n} f(i,j)/\sqrt{m \times n}$ is the zeroth order orthonormal Chebyshev moment of each flow, where $m \times n$ is the size of image, $f(i,j)$ is the value of pixel (x,y). Rahmalan has proved that low order orthonormal Chebyshev moment is useful in crowd density estimation [17], the author employed zeroth order, first order and second order moments to carry out estimation and the average rate of true classification is about 80%, which is not high. According to a survey paper [21], the first order moments of image are used to locate the center of mass; the second order moments are used to determine the principle axes of the object in image, obviously both first and second order moments have little relationship with crowd density. So we only extract zeroth order moment as feature in our approach.

Twelve texture features: Texture analysis has been proved very useful for crowd motion analysis in different papers [16),22),23]. A set of texture features can be extracted from the gray-level co-occurrence matrix (GLCM), in which the Contrast, the Homogeneity, the Energy and the Entropy are always used for estimating density. However, when extracting these features for estimating, the rate of true classification is about 70%, which is not a good result. We propose new features from GLCM which can represent crowd density more exactly. The definition is as follows:

$$TF = -\ln \sum_{i,j} M(i,j)^2 - \sum_{i,j} M(i,j) \ln M(i,j), \tag{4}$$

in which M is the GLCM of each flow. In fact, this feature contains the Energy and Entropy of GLCM, where the Energy implies complexity of image, high complex image has small value of the Energy, the Entropy shows distributing of elements in GLCM, a GLCM with even distribution produces large Entropy [24]. It means an image with high crowded objects should have small Energy, large Entropy and accordingly large value of TF. We extract TF by the following steps:

Step 1 The RGB foreground image is changed into three channels of gray images.

Step 2 The 256-level gray images are changed into 64-level gray images.

Step 3 Each gray image is used to calculated 4 GLCM with different θ: $0°, 45°, 90°$ and $135°$.

Step 4 Extracting TF from each GLCM.

So we extract 12 features from texture analysis, totally, 15 features are used for estimating flow density. Since both Marana and Rahmalan used self organizing maps (SOM) [25] as a final classifier, we also use it in our framework. The set of 15 features is used as feature vector by SOM neural network to classify flows into different levels.

4. Experiments and Discussion

We have verified our approach on 10 videos shown in **Fig. 5** where 8 of them are human videos and 2 videos are car videos. We selected 1,776 images as our dataset, and the size of each image is 320×240. Since our framework is made up of three parts, our experiments are carried out in three steps. The first step is to evaluate the performance of background removal, the second step is to test the result of foreground segmentation and the last step is to evaluate crowded objects density estimation and classification, which are illustrated respectively as follows.

4.1 Background Removal Result

We randomly picked up 100 images from dataset to test OFBM. We firstly marked foreground area on each image, which is considered as "ground truth" of

Fig. 5 10 different videos used in experiments.

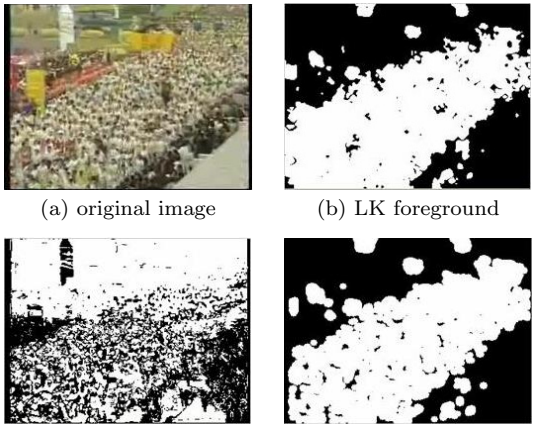
foreground, then we compared the experimental foreground result and the "true value", finally we calculated the error rate ER_{br} of background removal approach as follows,

$$ER_{br} = \frac{|A_{real} - A|}{A_{real}} \times 100\% \tag{5}$$

where A_{real} is the "ground truth" of foreground, A is the foreground area obtained by our approach. After testing the 100 images, the average error rate ER_{br} is 4.64%. As comparison, we also calculated the error rate of LK optical flow and GBS methods and the results turned out 22.3% and 12.7%. Obviously, our approach is much better and foreground detection is more accurate. Figures 6 to 8 show three groups of image result, each group contains four images, (a) the original image, (b) the LK foreground image, (c) the GBS foreground image and (d) OFBM foreground image. **Figure 6** (a) is a picture of marathon game, which contains a large number of people and there are a lot of movement in the background. It can be seen that the result of GBS method is very bad, but by using our approach OFBM, most noise is eliminated. **Figure 7** (a) is a picture of students, though GBS result is good, LK result includes much noise caused by illumination change. The same situation happens in **Fig. 8**, where Fig. 8 (a) is a picture of highway cars. From Figs. 7 (d) and 8 (d) we can see that OFBM approach is robust to illumination change.

4.2 Foreground Segmentation Result

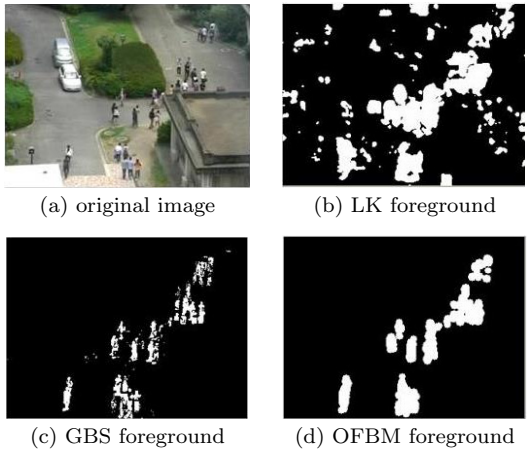
We selected another 100 images for foreground segmentation test, after using optical flow clustering, we employed the block-absorb approach to remove small



(a) original image (b) LK foreground

(c) GBS foreground (d) OFBM foreground

Fig. 6 Background removal result of marathon image.



(a) original image (b) LK foreground

(c) GBS foreground (d) OFBM foreground

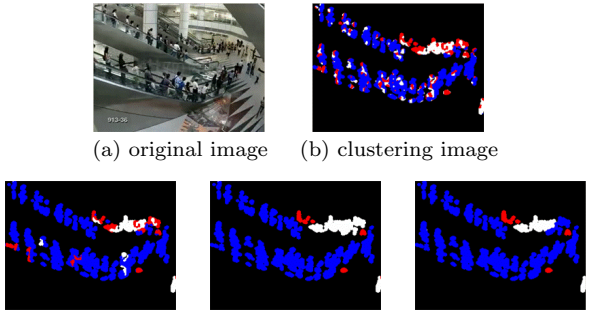
Fig. 7 Background removal result of students image.

blocks. As mentioned, a threshold L is employed to determine which block will be absorbed, where $L = 200$ is a suitable value verified by many experimental results. As comparison, a clustering image is disposed using three threshold

(a) original image (b) LK foreground

(c) GBS foreground (d) OFBM foreground

Fig. 8 Background removal result of highway cars image.



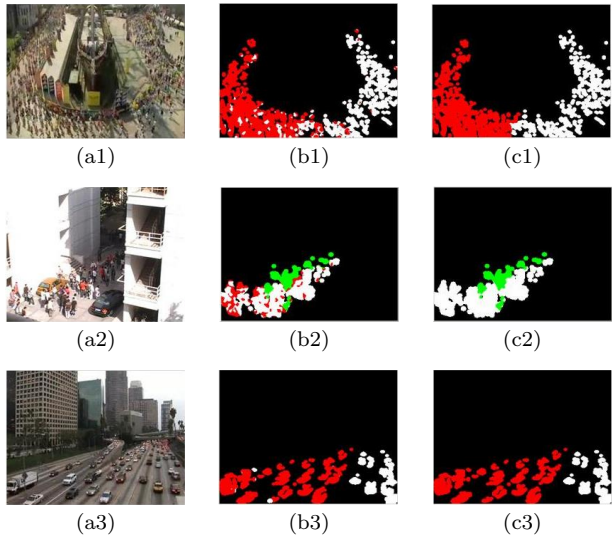
(a) original image (b) clustering image

(c) $L = 50$ (d) $L = 200$ (e) $L = 350$

Fig. 9 Absorbing result with different thresholds.

separately (50, 200, 350), and the result image is shown in **Fig. 9**. As can be seen, when using small threshold ($L = 50$), the blocks can not be absorbed completely (Fig. 9 (c)). If using large threshold ($L = 350$), some foreground regions will be misclassification (Fig. 9 (e)). $L = 200$ is the proper selection.

In addition, we show 3 groups of image results in **Fig. 10**, Fig. 10 (a1) is a picture of marathon game, Fig. 10 (a2) is a picture of students and Fig. 10 (a3)



(a1) (b1) (c1)

(a2) (b2) (c2)

(a3) (b3) (c3)

Fig. 10 3 groups of foreground segmentation results. Each group contains three pictures, the original image, the initial segmentation result image and the result of block-absorb.

is a picture of highway cars. Our foreground segmentation consists of two steps, initial foreground segmentation and block-absorb. After initial segmentation, the foregrounds of marathon image and car image are separated into two parts Fig. 10 (b1)(b3), which is consistent with the observation. The foreground of students image is divided into three parts, red, green and white Fig. 10 (b2), because some objects are big and move slowly so that the optical flow is disordered. When using block-absorb approach, the red color is absorbed successfully and the foreground is segmented into two parts Fig. 10 (c2).

We have compared our approach (**Fig. 11** (a)) with spectral clustering [26] (Fig. 11 (b)) and particle dynamics segmentation [10] (Fig. 11 (c)). From Fig. 11 (c) we can see that the image is segmented into three parts colored in blue, red and yellow. The blue region shows the background, the red part and yellow part are two flows with opposite moving directions. In fact, there are small flows in the red and yellow parts, but this method can not detect, which implies that the performance of dynamic segmentation should be improved. Spectral clustering is

(a) Our approach (b) Spectral clustering (c) Dynamic segmentation

Fig. 11 Comparison result of three foreground segmentation methods.

Table 3 Numerical result of human flows classification.

Level	A: Very Low	B: Low	C: Mid	D: High	E: Very High
Truth Value	93	176	238	251	225
Test Value	88	191	203	299	202
Rate of Truth	94.6%	92.1%	85.3%	83.9%	89.8%

similar with our approach, but it carries out segmentation on sparse optical flow field. So the result image (Fig. 11 (b)) only shows the approximate segmentation, we can not find clear boundaries between flows. Compared to these two methods, our approach can detect large and small flows, and by using block-absorb method, we can find the exact boundaries, as shown in Fig. 11 (a).

4.3 Density Estimation Result

We picked up 800 human images to test crowded objects density estimation, 400 images were used for training and the other 400 were used for classification. Because each image contains two or three flows, totally there are 983 flows in 400 images. After manual estimation, these flows are classified into five levels according to Table 1. Then we employed our approach to classify all the 983 flows, the numerical result is listed in **Table 3**. It can be seen that the average true classification rate is 86.2%, compared with two methods proposed by Marana [16] and Rahamalan [17], in the former 73.89% of the test images are correctly classified using real-time crowd density estimation and in the latter the true classification rate of TIOCM is about 80%. There are two reasons for the low rates of the methods, one is that both the methods estimated density on whole image and the background noise (floor texture, background movement and so on) led to misclassification. The other is the extracted features could not efficiently reflect crowd density, such as second-order moments and the Homogeneity.

Table 4 Numerical result of car flows classification.

Level	A: Low	B: Mid	C: High
Truth Value	32	130	38
Test Value	46	110	44
Rate of Truth	69.5%	84.6%	86.3%

We carried out the same experiment on 200 car images, in which 100 images for training and 100 images for test, which contain 200 flows. The numerical result is shown in **Table 4**, the average true rate is 80.1%, which is not high because of the various sizes of cars and occlusions. Since no one estimated cars density before, our approach is just a start to induce others to come forward with valuable contributions.

As experiment results, when we estimated crowded objects density on every frame, it took about 35 seconds to deal with 10-second video sequence on a PC of 2.4 GHz CPU and 2.0 GB memory. However, for practical application, we need not to estimate every image. When carrying out density estimation at intervals of 10 frames, it can be implemented on real time.

5. Conclusion

This paper proposes some new techniques of COMA, which consists of background removal, foreground segmentation and crowded objects density estimation. To obtain an optimal foreground image, we propose a background removal approach OFBM, which can eliminate most background noises caused in LK and GBS methods and detect all foreground. In the part of foreground segmentation, we propose optical flow clustering and block-absorb approach. Since the optical flow field implies all the changes in original image, we segment optical flow field. After initial segmentation, there are some small blocks in the flows, so we employ block-absorb approach to deal with them. Finally, based on the result of foreground segmentation, crowded objects density estimation on each flow is carried out.

Experimental results show that the proposed approach is efficient. First, the detected foreground contains less noise than that of LK and GBS methods. Second, optical flow clustering and block-absorb are useful in foreground segmentation. With proper K, the proposed algorithm can be used to segment different videos.

Third, the proposed density estimation approach is more accurate than former methods, because there is no background influence and the extracted features are more efficient.

However, there are two factors that influence the result of proposed framework. The first one is the optical flow. In our framework, both background removal and foreground segmentation employ LK optical flow, but LK optical flow is not exactly, which will influence the performance of our framework. The other factor is occlusions in crowd scene. When too many objects gather together, there will be much occlusions in the crowd. The occlusions make it very difficult to estimate crowd density, because the texture of foreground becomes chaotic and the edge is hard to find out. For future work, we want to solve these two problems and improve the performance of our approach.

Acknowledgments The work was supported by JASSO (Japan Student Service Organization) financial support and Beijing Key Laboratory of Advanced Information Science Network Technology and Railway Key Laboratory of Information Science Engineering (No.XDXX1010).

References

1) Zhan, B., Monekosso, D. and Xu, L.: Crowd Analysis: A Survey, *Machine Vision and Application*, pp.345–357 (2008).
2) Lipton, A., Fujiyoshi, H. and Patil, R.: Moving target classification and tracking from real-time video, *Proc. IEEE Workshop on Applications of Computer Vision*, pp.8–14 (1998).
3) Collins, R.: A system for video surveillance and monitoring: VSAM final report, Carnegie Mellon University, Technical Report: CMU-RI-TR-00-12 (2000).
4) Haritaoglu, I., Harwood, D. and Davis, L.: Ghost: A human body part labeling system using silhouettes, *14th International Conference on Pattern Recognition*, Vol.1, pp.77–82 (1998).
5) Winter, A.: *The Biomechanics and Motor Control of Human Movement*, 2nd ed., John Wiley and Sons (1990).
6) Cheriyadat, A. and Radke, R.: Detecting Dominant Motions in Dense Crowds, *Signal Processing*, Vol.2, No.4 (2008).
7) Lauer, A. and Schnorr, C.: Spectral Clustering of Linear Subspaces for Motion Segmentation, *IEEE Conference on Computer Vision 2009* (2009).
8) Brendel, W. and Todorovic, S.: Video Object Segmentation by Tracking Regions, *IEEE Conference on Computer Vision 2009* (2009).
9) Huang, Y., Liu, Q. and Metaxas, D.: Video Object Segmentation by Hypergraph Cut, *IEEE Conference on Computer Vision and Pattern Recognition 2009* (2009).
10) Ali, S. and Shah, M.: A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis, *IEEE Confence on Computer Vision and Pattern Recognition 2007* (*CVPR2007*), pp.1–6 (2007).
11) Ihaddadene, N. and Djeraba, C.: Real-Time Crowd Motion Analysis, *19th International Conference on Pattern Recognition 2008* (*ICPR2008*), pp.1–4 (2008).
12) Andrade, E., Blunsden, S. and Fisher, R.: Hidden Markov models for optical flow analysis in crowds, *18th International Conference on Pattern Recognition 2006* (*ICPR2006*), Vol.1, pp.460–463 (2006).
13) Chan, A.: Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking, *IEEE Confence on Computer Vision and Pattern Recognition 2008* (*CVPR2008*) (2008).
14) Chan, A. and Vasconcelos, N.: Modeling, Clustering, and Segmenting Video with Mixtures of Dynamic Textures, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.30, pp.909–926 (2008).
15) Ma, R., Li, L., Huang, W. and Tian, Q.: On Pixels Count Based Crowd Density Estimation for Visual Surveillance, *IEEE Confence on Cybernerics and Intelligent Systems*, pp.170–173 (2004).
16) Marana, A. and Cavenaghi, M.: Real-Time Crowd Density Estiamtion Using Images, *International Symposium on Visual Computing 2005* (*ISVC2005*), pp.355–362 (2005).
17) Rahmalan, H., Nixon, M. and Carter, J.: On Crowd Density Estimation for Surveillance, *Crime and Security*, pp.540–545 (2006).
18) Stauffer, C. and Grimson, W.: Adaptive background mixture models for real-time tracking, *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.246–252 (1999).
19) Gonzalez, R., Woods, R. and Eddins, S.: *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ (2002).
20) MacKay, D.J.C.: *Chapter 20: An Example Inference Task: Clustering*, Information Theory, Inference and Learning Algorithms, pp.284–292, Cambridge University Press (2003).
21) Prokop, R. and Reeves, A.: A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition, *Graphical Models and Image Processing*, pp.438–460 (1992).
22) Marana, A., Velastin, S. and Lotufo, R.: Automatic Estimation of Crowd Density Using Texture, *Safety Science*, Vol.28, No.3, pp.165–175 (1998).
23) Li, W., Wu, X.j., Matsumoto, K. and Zhao, H.A.: Crowd Density Estimation on Real Scenes, *Proc. IEEE International Symposium on Industrial Electronics* (*ISIE 2010*), pp.1595–1600 (2010).
24) Zhang, J. and Tan, T.: Brief review of invariant texture analysis methods, *Pattern Recognition*, Vol.35, pp.735–747 (2002).

25) Kohonen, T.: The Self Organizing Map, *Proc. IEEE*, Vol.78, No.9, pp.1464–1480 (1990).
26) Eibl, G. and Brandle, N.: Evaluation of Clustering Methods for Finding Dominant Optical Flow Fields in Crowded Scenes, *The 19th International Confence on Pattern Recognition*, pp.1–4 (2008).

(Received August 23, 2010)
(Accepted January 14, 2011)
(Released April 6, 2011)

Wei Li was born in 1982. He is a candidate Ph.D. student of the School of Information Science and Engineering, Shandong University, China. From October 2009 to September 2010, he joined Department of Computer Science and Eletrical Engineering, Kumamoto University as an exchange researcher. His current research interests include the areas of digital image processing and computer vision.

Xiaojuan Wu received her M.S. degree from Shandong University, China in 1968. Following her graduation, she worked as a Chief Technician at Shandong Yangxin Electricity Generating Station. Since 1977, she has worked at Shandong University. From 2000, she is a professor at Shandong University. Her current research interests include the areas of computer vision, image processing and pattern recognition.

Hua-An Zhao received his B.S. and M.S. degrees in Electrical Engineering from Anhui University, China in 1982 and 1986, respectively. He also received his Ph.D. degree from Hiroshima University, Japan in 1993. During 1993–2006, he joined the Faculty of Engineering, Kyushu Kyoritsu University. From 2007, he is a professor at Kumamoto University. His current research interests include the areas of communications, signal processing, graph theory and its applications and VLSI layout design.