Foreground segmentation under sudden illumination changes by feature fusion

Gong Weiguo, Liu Runyao, Zhang Rui

(Key Lab of Optoelectronic Technology & Systems of Education Ministry, Chongqing University, Chongqing 400044, China)

Abstract: To address the challenging problem of robust background subtraction under sudden illumination changes, a novel foreground segmentation method based on feature fusion was proposed. The method consisted of three stages. First, a scene model through integrating the global illumination function into the framework of Gaussian mixture models was built. Second, three kinds of illumination invariant features, i.e. zero mean normalized cross-correlation (ZNCC), textures, and contours, were extracted from the current frame image. Third, the illumination invariant features were combined for foreground segmentation in two steps. Specifically, the ZNCC and textures were combined in the first step, and the contour was integrated in the second step. The experimental results show that the proposed method can effectively improve the accuracy and robustness of foreground segmentation.

Key words: feature fusion; Gaussian mixture models; global illumination function; main structure extraction of image; sudden illumination changes

CLC number: TP391.4 **Document code**: A **Article ID**: 1007-2276(2014)12-4164-06

光照突变下融合多类特征的场景分割方法

龚卫国,刘润瑶, 张 睿

(重庆大学 光电技术及系统教育部重点实验室,重庆 400044)

摘 要:为解决场景模型在快速光照变化下失效的问题,提出了一种新的前景目标分割方法。该方法共包括三个步骤。首先,利用全局光照函数建立高斯混合模型;其次,提取当前帧中的纹理、ZNCC 及轮廓特征;最后,将提取到的特征分两阶段与高斯混合模型进行融合(第一阶段:融合纹理及 ZNCC 特征;第二阶段:融合轮廓特征),得到最终的场景分割结果。实验结果表明:该算法具有较好的鲁棒性,并且相较于基于全局光照建模的方法具有更高的精度值及召回值。

关键词:特征融合; 高斯混合模型; 全局光照函数; 图像主结构提取; 光照突变

收稿日期:2014-04-10; **修订日期**:2014-05-13

基金项目:重庆市科技攻关计划(CSTC2012GG-YYJSB40001);重庆市集成示范计划(CSTC2013JCSF40009);中央高校基本科研业务费科研专项“研究生科技创新基金”(CDJXS11122216);国家自然科学基金(61105093)

作者简介:龚卫国(1957-),男,教授,博士生导师,博士,主要从事模式识别及智能化信息处理方面的研究。Email:wggong@cqu.edu.cn

通讯作者:刘润瑶(1988-),女,硕士生,主要从事信息获取和处理方面的研究。Email:Liurunyao@cqu.edu.cn

0 Introduction

Foreground segmentation is an important research area in video surveillance, and background subtraction is the most popular approach for the foreground segmentation. The essence of the background subtraction is known as scene modeling, which was divided into six different aspects by Toyama et al.[1], according to its research issues and difficulties. Standard Gaussian Mixture Models(GMM) can resolve parts of those difficulties, but it still cannot adequately address the misclassification problems caused by sudden illumination changes. Li[2] proposed an approach that adopted a fixed adaptation rate to deal with global illumination changes. This method requires that the adaptation speed of the adaptation function must be slower than the speed of the detected illumination changes. Xue et al.[3] proposed to use the illumination invariant features to reduce the effect of sudden illumination changes to foreground segmentation. But these methods couldn't resolve the problems caused by global illumination changes. Pilet[4] et al. proposed a kind of scene modeling methods for the global illumination scenario. However, since the illumination invariant features they used are texture features, the number of the textures would affect the effectiveness of the foreground segmentation.

To solve the problem of foreground segmentation under sudden illumination changes, in this paper, we propose a foreground segmentation method that integrates with Zero-mean Normalized Cross Correlation (ZNCC), textures, and contours. First, we build the GMM based on global illumination information. Then, we combine the ZNCC and textures for an initial foreground segmentation. Subsequently, the contour features are integrated to correct the initial segmentation results. To eliminate the effect of details on contour extraction, our method introduces an image main structure extraction technology, which makes the extracted contour more concise and easier to be segmented. Experiments on the video sequences with different illumination changes demonstrate that the proposed method can effectively improve not only the precision and recall of the foreground segmentation, but also the robustness under sudden illumination changes.

1 Scene modeling

Since the illumination can dramatically affect a pixel's intensity, the standard GMM method (that models the pixel's intensity) is easy to produce a large area of mismatches. To resolve the above problem caused by the illumination, this paper combines a global illumination function with the GMM.

1.1 Global illumination function

In this paper, we employ Lambertian reflection models to build the global illumination function: Given the stored background is $\mathbf{m}$, the background in the current frame is $\mathbf{m}_i$, and each pixel in the current frame is $\mathbf{u}_i$. If there is no foreground in the current frame, then each pixel's value is determined by the irradiance and reflectivity on the surface of the objects in the scene. According to Lambertian assumption: $\mathbf{u}_i=\mathbf{e}_i\mathbf{a}_i$, we can have a similar assumption:

$$\mathbf{m}_i=\mathbf{e}_m\mathbf{a}_i \tag{1}$$

where $\mathbf{e}_m$ is the surface constant of the scene in the current frame. From Eq.(1), we can obtain the global illumination function as Eq.(2). The Eq.(2) shows that the global illumination function is relevant to the reflectivity, and only relevant to the current frame and background images. The global illumination function is determined by the current scene:

$$l_i=\frac{\mathbf{u}_i}{\mathbf{m}_i}=\frac{\mathbf{e}_i}{\mathbf{e}_m} \tag{2}$$

The Eq.(2) can be generalized to color images with three channels:

$$l_i=\left[\frac{\mathbf{u}_{i,r}}{\mathbf{m}_{i,r}},\frac{\mathbf{u}_{i,g}}{\mathbf{m}_{i,g}},\frac{\mathbf{u}_{i,b}}{\mathbf{m}_{i,b}}\right]^T \tag{3}$$

As shown in Eq.(3), the global illumination function, l_i, is the ratio of the current pxiel's value to the background model. Without foreground occlusion, this function can eliminate the effect of the global illumination changes to the pixels of the background.

1.2 Gaussian mixture modeling

This paper adopts the modeling method proposed by Pilet et al[4-6]. Given a pixel $\mathbf{u}_i$ in the current frame, the corresponding global illumination function is l_i. We use the K-component GMM to model the previous records of l_i. As a result, the entire scene can be described by $\mathbf{K}$ different illumination functions. Similar to the standard GMM [7]. The probability density function (PDF) of the current pixel is the weighted sum of the $\mathbf{K}$ Gaussian PDF:

$$\mathbf{p}(l_i|\varepsilon_i,\mu,\sigma)=\prod_{k=1}^{\mathbf{K}}\omega_{\mathbf{k},t}^{\varepsilon_i^{(k)}}\mathbf{N}(l_i|\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})^{\varepsilon_i^{(k)}} \tag{4}$$

The Eq. (4) represents the background model, where μ denotes the mean, σ denotes the variance, and $\omega_{\mathbf{k},t}$ is the weight of the $\mathbf{K}^{th}$ Gaussian component, $\mathbf{N}$ is the PDF of the i^{th} Gaussian distribution.

In the established model, the foreground is composed of $\mathbf{K}_f$ Gaussian distributions, $\mathbf{N}(\mathbf{u}_i|\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})$, which are based on illumination changes. Due to the existence of the potential pixel variations, the distributions of these pixels will not comply the normal distribution. The probability of the occurrence of these pixels is $1/256^3$. So, the distribution of pixel $\mathbf{u}_i=[\mathbf{u}_i^{r},\mathbf{u}_i^{g},\mathbf{u}_i^{b}]$ is presented in Eq.(5):

$$\mathbf{p}(\mathbf{u}_i|\varepsilon_i,\mu,\sigma)=\left(\frac{\omega_{\mathbf{K}_b+\mathbf{K}_f+1}}{256^3}\right)^{\varepsilon_i^{\mathbf{K}_b+\mathbf{K}_f+1}}\times\prod_{k=\mathbf{K}_b+1}^{\mathbf{K}_b+\mathbf{K}_f}\omega_{\mathbf{k},t}^{\varepsilon_i^{(k)}}\mathbf{N}(\mathbf{u}_i|\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})^{\varepsilon_i^{(k)}} \tag{5}$$

The background model described by Eq. (4) is based on global illumination change function. The foreground model described by Eq. (5) is based on pixel intensity values. To unify the variables, we use Jacobi determinant to transform Eq.(4):

$$\mathbf{p}(\mathbf{u}_i|\varepsilon_i,\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})=\mathbf{p}(l_i|\varepsilon_i,\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})/\mathbf{J}_i \tag{6}$$

Since Eq. (4) and (5) are used to describe background models, the value of $\mathbf{k}$ in Eq.(6) will be $\{1,2,\cdots,\mathbf{K}_b\}$; is the Jacobi determinant of the illumination function $l_i(\mathbf{u}_i)$.

To reduce the amount of calculation, we assume that all the pixels of the current scene are independent and identically distributed. Then, the joint distribution of all the pixels in the current frame can be figured out from Eq.(7):

$$\mathbf{p}(\mathbf{u},\varepsilon|\mu,\sigma,\omega)=\prod_{i,k}\mathbf{p}(\mathbf{u}_i|\varepsilon_i^{(k)},\mu_{\mathbf{k},t},\sigma_{\mathbf{k},t})\,\omega_{\mathbf{k}} \tag{7}$$

The preliminarily built scene model is presented by Eq.(7). As the presupposition of the model is that all the pixels are independent and identically distributed, we introduce the method of multi-feature fusion in Section 2.

2 Illumination invariant features extraction

To optimize the scene model of the current frame, we extract ZNCC, textures, and contour features, respectively. Then, we integrate these features in twice, in order to describe the special relationships between pixels, and more accurately segment the scene.

2.1 ZNCC features

Given an arbitrary pixel in the current frame, by mean removal we transform the grayscale vector into a zero-mean grayscale vector $\mathbf{u}'=(\mathbf{u}_1',\mathbf{u}_i',\cdots,\mathbf{u}_n')$. Given before and after mean removal, the grayscale vectors in the window $\mathbf{w}_{ib}/\mathbf{w}_i$ of the background/current frame are $\mathbf{u}_b/\mathbf{u}_i$ and $\mathbf{u}_b'/\mathbf{u}_i'$, respectively. Then the normalized feature is represented by Eq.(8):

$$f=\frac{\mathbf{E}(\mathbf{u}_b'\times\mathbf{u}_i')-\mathbf{E}(\mathbf{u}_b')\mathbf{E}(\mathbf{u}_i')}{\mathbf{D}(\mathbf{u}_b')\times\mathbf{D}(\mathbf{u}_i')}=\frac{1}{\mathbf{n}}\left(\frac{\mathbf{I}_b'}{\mathbf{D}(\mathbf{I}_b')}\right)\times\left(\frac{\mathbf{I}_i'}{\mathbf{D}(\mathbf{I}_i')}\right)^{T} \tag{8}$$

where $\mathbf{E}$ and $\mathbf{D}$ denote the operations of getting the average and the variance, respectively.

Given $j=(\mathbf{x},\mathbf{y})$ is the corresponding pixel position in the background frame and the current frame, then:

$$\mathbf{u}_{bj}=\frac{\sigma_b}{\sigma_i}\mathbf{u}_{ij}+\left(\mu_b-\frac{\sigma_b\mu_i}{\sigma_i}\right)=\lambda\mathbf{u}_{ij}+\eta \tag{9}$$

where λ and η are constants, Eq. (9) satisfies light affine model: $f'=\lambda f+\eta$. Therefore, ZNCC can describe the spatial relationship between the current frame and the background frame. It satisfies light affine model and has illumination invariant features.

The ZNCC features extracted in our method are obtained from Eq.(10).

$$f_i^A=\frac{n\sum_{j\in W_i} u_{bj}'u_{ij}'-(\sum_{j\in W_i} u_{ij}')\cdot(\sum_{j\in W_i} u_{bj}')}{\sqrt{n\sum_{j\in W_i}(u_{ij}')^2-\sum_{j\in W_i}(u_{ij}')^2}\sqrt{n\sum_{j\in W_i}(u_{bj}')^2-\sum_{j\in W_i}(u_{bj}')^2}} \quad (10)$$

2.2 Texture features

Textures also have a certain degree of robustness to illumination changes. Texture features can describe the spatial color distributions and light intensity distributions of an image or a small region of the image. Therefore, texture features can be utilized as illumination invariant features. The texture features in the current frame are calculated as shown in Eq.(11):

$$f_i^B=\sqrt{n\sum_{j\in W_i}(u_{ij})^2-(\sum_{j\in W_i}u_{ij})^2}+\sqrt{n\sum_{j\in W_i}(u_{bj})^2-(\sum_{j\in W_i}u_{bj})^2} \quad (11)$$

The proposed method manually segments a large number of texture images and conducts off-line training. The data obtained from this training do not need to be updated online, and the current frame image can be directly conducted the statistics of its texture features.

2.3 Contour features

If uniform areas appear in the scenes, the texture features/ZNCC information of these regions will dramatically decrease. This will lead our model to obtain an inaccurate foreground segmentation in these regions. To solve the segmentation problems in uniform regions, the proposed method introduces the contour features to the further correct the scene modes.

In the process of contour extraction to the scenes, the extracted contour information will contain a lot of details because of the existence of the textures. However, the traditional image filtering, such as Gaussian filtering, can only accomplish a simple denoising of images, but cannot eliminate the effect of the inherent textures in the original images. In order to obtain concise and clear contour of the scene, the main image structures of the input image are extracted to obtain $\mathbf{m}_i'$ with prominent main contours and less texture details[8]. Subsequently, the contours are extracted to obtain a more concise contour feature.

To obtain the main contour features, we construct the following filters:

$$\mathbf{H}_1=\begin{vmatrix}1 & -1\\ 1 & -1\end{vmatrix}\ \mathbf{H}_2=\begin{vmatrix}-1 & -1\\ 1 & 1\end{vmatrix}\ \mathbf{H}_3=\begin{vmatrix}1 & 0\\ 0 & -1\end{vmatrix}\ \mathbf{H}_4=\begin{vmatrix}0 & 1\\ -1 & 0\end{vmatrix}$$

The above model ensures the symmetry and non-offset properties of the extracted image. The extracted main contour features are calculated from Eq.(12):

$$f_i^C=\sqrt{\sum_{z\in 1}^{4}[n\sum_{j\in W_i}(u_{ij}*\mathbf{H}_z)^2]}+\sqrt{\sum_{z\in 1}^{4}[n\sum_{j\in W_i}(u_{bj}*\mathbf{H}_z)^2]} \quad (12)$$

3 Feature fusion for foreground segmentation

3.1 Initial foreground segmentation

To accurately segment the scene, we obtain the statistics of the spatial information between pixels of the background and foreground, respectively. The statistical information is represented by the histograms of the background and foreground, which are stored in $\mathbf{h}_f(f_i|\mathbf{u}_i)$ and $\mathbf{h}_b(f_i|\mathbf{u}_i)$, respectively. The obtained spatial information between pixels is obtained as follows:

$$\mathbf{P}_s(f_i^A,f_i^B|\mathbf{k})=\begin{cases}\mathbf{P}_b(f_i^A,f_i^B),\ \mathbf{k}=1,\cdots,\mathbf{k}_b\\ \mathbf{P}_f(f_i^A,f_i^B),\ \mathbf{k}=\mathbf{k}_{b+1},\cdots,\mathbf{k}_{b+f}\end{cases} \quad (13)$$

To combine the scene model based on independent pixel built by Eq.(7) with the spatial information scene between pixels obtained by Eq.(13), the joint probability distribution of the preliminary foreground segmentation can be obtained as follows:

$$\mathbf{p}(\mathbf{u},\varepsilon,\mathbf{f_A},\mathbf{f_B}|\mu,\sigma,\omega)=\prod_{i,k}\omega_k\cdot\mathbf{p}(\mathbf{u}_i|\varepsilon_i^{(k)},\mu_{k,t},\sigma_{k,t})\cdot \mathbf{P}_s(f_i^A,f_i^B|\mathbf{k}) \quad (14)$$

where the optimal parameters are obtained by Expectation Maximization (EM) algorithm.

3.2 Further foreground segmentation

After the initial foreground segmentation, parts of the foreground regions(e.g., the face, hands, and other more balanced regions of a person in the foreground) are often misclassified into the background because of the lack of ZNCC and texture features. To improve the segmentation accuracy, we conduct further foreground segmentation by combining the contour features. By a lot of experiments, we can obtain the following empirical results:

(1) The ideal segmented contour information of the scene is a subset of the extracted contour features;

(2) The foreground contour information, including the edges of the frame image, is closed irregular shapes.

Based on the above two points, the proposed method combines the contour features of the frame image edges: $f_i^{c0}=f_i^{c}+f_z^{c}$ (f_z^{c} is the frame image edge). To calculate the pixels within the closed edges f_i^{c0}:

$$\beta(\mathbf{u}_i)=\frac{\sum_{i\in\psi_t}\mathbf{u}_{if}}{\sum_{i\in\psi_t}\mathbf{u}_{if}+\sum_{i\in\psi_t}\mathbf{u}_{ib}} \tag{15}$$

where ψ_t is a window naturally formed by the closure of the contours. According to the experimental results, we set the threshold of $\beta(\mathbf{u}_i)$ as 0.4. If $\beta(\mathbf{u}_i)\geqslant 0.4$, then the window is considered as foreground window; otherwise, it is considered as background window. In this way, the scene can be segmented into foreground and background according to the windows by traversing every window in the scene. It can further correct the misclassified pixels of the foreground affected by the uniform areas, thus improving the accuracy of foreground segmentation.

4 Experiments

To verify the accuracy and robustness of the proposed method, the experiments are implemented on PC with an Intel 3 GHz Pentium 4 CPU and 1.50 GB memory by using OpenCV library and Matlab2012b.

To objectively evaluate the foreground segmentation results, we will make quantitative comparisons and analysis of the experimental results by computing the precision and recall. Precision and recall reflect the accuracy of the foreground segmentation algorithm at the same time. The closer the above two metrics approach to 100% simultaneously, the better the performance of the method. We tested several videos containing illumination changes. To simulate the rapid illumination changes, we created three new video sequences from three original video sequences, i.e., Ford, Bank and TimeOfDay. Specifically, from each video sequence, we selected a frame from it in every 5-20 frames (varying dependent on different videos) and combined the selected frame into a new video sequence. As a result, all of these new video sequences contain significant rapid illumination changes. We used different methods to test the above videos. The segmentation results are shown in Fig.1

Fig.1 Segmentation results of different methods for different scenes with rapid illumination changes

and the corresponding precision and recall values are given in Tab.1. In Fig.1, the first column shows the scenes of the test video sequences. The second column shows the current frames. The segmentation results of the standard GMM/illumination method are presented in the third/forth column, respectively. The segmentation results of the proposed method are shown in the fifth column. The last column gives the corresponding ground truth. From the first row to the fourth row, the test scenes are LightSwitch, TimeOfDay, Bank, and Ford, respectively. In Tab.1, **V**1, **V**2, **V**3 represent the standard GMM, the method based on the global illumination model, and the proposed method, respectively. Note that, in the Ford and Bank scenes, the GMM will segment most part of the frame image into foreground and lead **Fn** to a very small value and the precision of almost 100%. Since these precisions cannot be used for evaluation, we will not list them in Tab.1.

Tab.1 Precision and recall of different methods for different video sequences (Unit: %)

		Lightswitch	Ford	Bank	Time of day
Precision	$\mathbf{V}_1$	44.98%	-	-	49.31%
	$\mathbf{V}_2$	75.30%	60.77%	93.67%	75.32%
	$\mathbf{V}_3$	94.63%	85.87%	98.09%	78.36%
Recall	$\mathbf{V}_1$	10.91%	7.17%	18.1%	5.26%
	$\mathbf{V}_2$	51.92%	37.62%	72.15%	23.01%
	$\mathbf{V}_3$	78.70%	77.02%	93.88%	34.4%

According to the comparison of the results in Tab.1, the proposed detection method can achieve better results than other methods in the environments with illumination changes. Compared to the standard GMM method, our method is much better in both precision and recall. Compared to the method based on the global illumination model, the recall of our method has an increase of 24.83% in average and 39.4% in the highest; and our precision has an increase of 12.97% in average and 25.10% in the highest. To the challenging scenes, i.e., TimeOfDay and Ford, the proposed method can still remain good segmentation results. According to the above results from different sets of the test videos, it indicates that our method has higher accuracy than the standard GMM method and the method based on the global illumination change for foreground detection.

5 Conclusions

We proposed a novel foreground segmentation method to tackle the sudden illumination change problem. We first built Gaussian mixture models (GMM) with a global illumination function. Then we extracted three kinds of illumination invariant features, i.e. ZNCC, textures, and contours. Combined with the GMM, these features are used to segment the foreground in two steps, from coarse to fine. The experimental results have shown that our method can effectively improve the accuracy and robustness of the foreground segmentation under sudden illumination changes.

References:

[1] Toyama K, Krumm J, Brumitt B, et al. Wallflower: principles and practice of background maintenance [C]// ICCV, 1999, 1: 255-261.

[2] Li Y, Xu R, Nie G, et al. Moving object detection method for change in scene light [J]. Computer Technology and Development, 2011, 21(2): 140-146. (in Chinese)

[3] Xue G, Sun J, Song L. Background subtraction based on phase and distance transform under sudden illumination change[C]//ICIP 2010, 2010: 3465-3468.

[4] Strecha Pilet C, Fua P. Making background subtraction robust to sudden illumination changes [C]//Proc European Conf on Computer Vision, 2008: 567-580.

[5] Osters L P J V, Caifeng Shan, Tommaso Gritti. Background subtraction under sudden illumination changes[C]//Advanced Video and Signal Based Surveillance (AVSS), 2010: 384-391.

[6] Luc Vosters, Caifeng Shan, Tommaso Gritti. Real-time robust background subtraction under rapidly changing illumination conditions [J]. Image and Vision Computing, 2012, 30(12): 1004-1015.

[7] Yuan C, Wang C, Zhang X, et al. Video segmentation of illuminance abrupt variation based on MOGs and gradient information [J]. Journal of Image and Graphics, 2007, 11 (12): 2068-2072. (in Chinese)

[8] Xu Li, Yan Qiong, Xia Yang, et al. Structure extraction from texture via relative total variation [J]. ACM Transactions on Graphics, 2012, 31(6): 139.