Background Reconstruction via Low Rank Tensor Factorization

Guiping Shen[1,2,3], Zhi Han[1,2], Xi'ai Chen[1,2,3], Yandong Tang[1,2], Yang Zhang[1,2,3]
[1]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, China
[2]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, China
[3]University of Chinese Academy of Sciences, China

ABSTRACT

This paper introduces a new method for background reconstruction. Background reconstruction from video sequences captured by a static camera can be regarded as a low rank factorization problem. Background is the low dimensional subspace restored from the higher dimensional visual data, and foreground is treated as sparse noise of unknown distribution. The existing algorithm could not deal with noise of unknown distribution effectively. Due to the limitation of the matrix decomposition which would lost space structure information, we process video data directly as higher order tensor based on low rank tensor factorization (LRTF). We put forward a new model of foreground by using Mixture of Gaussians (MoG) and Markov Random Field (MRF). Extensive experiments show that our method can effectively construct the background.

Keywords: background reconstruction, low rank, Tensor factorization, MoG, MRF.

1. INTRODUCTION

Background reconstruction is a key technic for video analysis, especially for video surveillance and target tracking. At present, many analytical approaches are based on the background extraction method. However, background extraction can not recover and reconstruct the occluded background regions effectively. Thus, it has certain limitations for video analysis. The general approach to background extraction is to use background models to distinguish between foreground objects and background regions in a video sequence.

Currently, many background extraction methods have their own advantages and disadvantages. The mixed Gaussian model (GMM) [1,2] is one of the classic background modeling methods. Chris Stauffer [3] uses several Gaussian models to represent the features of each pixel in the image, and updates it based on a new frame. Then match each pixel updated with previous ones. The pixels matching successfully are considered as the background regions, otherwise it is the foreground objects. The main parameters are the mean and variance of the mixed Gauss, which directly affect the stability, accuracy and convergence of the model. The disadvantage is that the amount of calculation is relatively large, the speed is slow, and it is sensitive to illumination. On this basis, Lee [4] proposes a new update method for the parameters of the mixed Gaussian model. Zoran Zivkovic [5] proposes an adaptive mixed Gauss. Pilet [6] improves the background separation method aiming at the problem that misunderstandings when light changes. Kim [7] uses codebook to achieve separation of foreground and background in 2004. The basic idea of codebook algorithm is to obtain a time series model of each pixel. This model can handle time fluctuations well. Hoverer, the disadvantage is that it requires a lot of memory. The codebook algorithm creates a codebook (CB) structure for each pixel of the current image, and each CB structure is composed of multiple codewords (CW). When the light changes a lot, the video frame can not match the codebook, therefore the background is misidentified as the foreground. [8,9,10] Later, the codebook algorithm was improved to enhance the algorithm robustness. Barnich proposed the ViBe method [11] in 2011, storing a sample set for each pixel. The sample value in the sample set is the pixel value of the pixel before and the pixel value of its neighbor pixels, and then each new pixel value is compared to the values in the sample set to determine if it is a background point. All of the above are typical methods based on pixel value modeling. In 2006, Marko models LPB histograms in a circular region for each pixel based on texture modeling [12]. Shengcai Liao proposes a new texture representation method SLITP [13] in 2010. In 2012, Li Feng merges the codebook and texture into a two-layer video background model [14].

In recent years, the low rank analysis problem has attracted a lot of attention. The restoration of low-dimensional linear subspace from high-dimensional visual data is widely used in face recognition, motion segmentation, video

surveillance, and background restoration. The static background of the video is the potential low-rank component of the video data, and the motion foreground can be seen as the sparse noise of the unknown distribution. Some existing methods deals with video data in the form of high-order tensors. For example, LRMF, RPCA, matrize the tensor to process tensor problem by means of matrix. Such matrization does not effectively utilize the tensor intrinsic structure and often results in local optimal. An effective method is to use low rank tensor decomposition (LRTF) to extract potential low rank subspace. It is generally obtained by minimizing the loss function between the observed data and the low rank decomposition data. The loss function has different forms under different noise distributions. For example, the Laplacian distribution is repressed by 1-norm and the Gaussian distribution is repressed by 2-norm. The actual noise distribution is complex and unknown, Meng proposes MoG to establish a noise model [15]. And Chen applies MoG to the LRTF framework [16]. Since there is some certain prior knowledge of noise in practical applications, Meng applies the local continuous prior knowledge of the noise component to the matrix decomposition [17].

In this paper, a new background reconstruction method is proposed, which makes full use of the low rank property and spatial structure information of the video background sequence. Under the LRTF framework, the noise model is built by using the mixed Gaussian and Markov random fields to better utilize the prior knowledge, i.e. local continuity of the foreground target. The model parameters are finally obtained by the variational EM method.

2. ALGORITHM DESCRIPTION

2.1 Conventional Notation for Symbols and Operators

A scalar is represented by a lowercase letter (a , b ,$\cdots$), and a lowercase letter ($\mathbf{a}$, $\mathbf{b}$,$\cdots$) in bold represents a vector whose elements are （ a_i , b_j ,$\cdots$）. Uppercase letters（ A , B ,$\cdots$）represent matrices whose elements are（ a_{ij} , b_{ij} ,$\cdots$）. The calligraphic letters （ $\mathcal{A}$, $\mathcal{B}$,$\cdots$）indicate high-order tensors. An N-order tensor is noted as $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, where $I_n (\mathrm{n} = 1, 2, \cdots \mathrm{N})$ are positive integers. Each element of the tensor is recorded as $x_{i_1 \cdots i_n \cdots i_N}$, where $1 \le i_n \le i_N$. The rank-1 tensor can be written as an outer product of N vectors:

$$\mathcal{X} = u \circ v \circ \cdots \circ t \tag{1}$$

where the elements in $\mathcal{X}$ is the product of the corresponding vector element, represented as $x_{i_1 i_2 \cdots i_N} = u_{i_1} v_{i_2} \cdots t_{i_N}$. Mode-n unfolding of a tensor is denoted as $X_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \ne n} I_i}$, and the rank of the mode-n unfloding matrix r_n is denoted as $X_{(\mathrm{n})} : r_n = rank(\mathrm{X}_{(\mathrm{n})})$. The mode-n product calculations of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and matrix $U \in \mathbb{R}^{J_n \times I_n}$ is expressed as $\mathcal{X} \times_n U \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots I_N}$. The inner product of two tensors of the same size $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} x_{i_1 \cdots i_N} y_{i_1 \cdots i_N} \tag{2}$$

$\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$ indicates the F-norm. $\|\mathcal{X}\|_0$ represents the 0-norm and counts the number of non-zero elements. $\|\mathcal{X}\|_1 = \sum_{i_1, \cdots i_N} \left| x_{i_1, \cdots i_N} \right|$ represents 1-norm. For any $1 \le n \le N$, $\|\mathcal{X}\|_F = \|X_{(\mathrm{n})}\|_F$, $\|\mathcal{X}\|_0 = \|X_{(\mathrm{n})}\|_0$, $\|\mathcal{X}\|_1 = \|X_{(\mathrm{n})}\|_1$.

2.2 Background Extraction Model

The background sequence of video data acquired from the fixed camera, constitutes a low rank tensor. And the motion foreground is regarded as sparse noise. In order to make full use of the spatial structure information of the background sequence, the motion foreground is modeled on the framework of low-rank tensor decomposition. Refer to Chen [16] and Meng [17] to use the prior knowledge of the local continuity of the foreground, we also take the Markov random field into consideration, to recover the low rank background sequence.

The foreground pixels of the motion are denoted as f_{ijk} . The elements of a CP decomposition of a 3rd order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is written as:

$$x_{ijk} = \sum_{d=1}^{r} u_i^d v_j^d t_k^d + f_{ijk} \tag{3}$$

We use MoG to model foreground moving targets, so the distribution $p(f)$ of each pixel f_{ijk} is defined as:

$$p(f) \sim \sum_{n=1}^{N} \pi_n N \quad f \left| 0, \sigma_n^2 \right. \tag{4}$$

where π_n is the mixing ratio parameter greater than 0 and $\sum_{n=1}^{N} \pi_n = 1$. Suppose a hidden variable z_{ijkn} in the model that satisfies $z_{ijkn} \in \{0,1\}$, $\sum_{n=1}^{N} z_{ijkn} = 1$, and $Z = \{ z_{ijkn} | i = 1, \cdots I; \mathrm{j} = 1, \cdots \mathrm{J}; \mathrm{k} = 1, \cdots \mathrm{K}; \mathrm{n} = 1, \cdots \mathrm{N} \}$. Given a mixture coefficient π, the distribution of z_{ijkn} satisfies:

$$p(\mathrm{z}_{ijk} | \pi) = \prod_{n=1}^{N} \pi_n^{z_{ijkn}} \tag{5}$$

Given a hidden variable and model parameters, a linear superposition of the Gaussian distribution can be written as a conditional distribution of the observed data:

$$p(\mathrm{x}_{ijk} | Z, \Delta, \Sigma) = \prod_{n=1}^{N} N(\mathrm{x}_{ijk} \left| \sum_{d=1}^{r} u_i^d v_j^d t_k^d, \sigma_n^2 \right.)^{z_{ijkn}} \tag{6}$$

where $\Delta = \sum_{d=1}^{r} u_i^d v_j^d t_k^d$ and $\Sigma = \{ \sigma_1, \sigma_2, \cdots \sigma_N \}$. The log likelihood function of $\mathcal{X}$ is:

$$\max_{U,V,T,Z,\Pi,\Sigma} L(\mathrm{X} | U, V, T, Z, \Pi, \Sigma) = \sum_{i,j,k \in \Omega, n} z_{ijkn} \left[\log \pi_n + \log N \left(x_{ijk} \left| \sum_{d=1}^{r} u_{id} v_{jd} t_{kd}, \sigma_n^2 \right. \right) \right] \tag{7}$$

Where $U = \{ \mathbf{u_1}, \mathbf{u_2}, \cdots \mathbf{u_r} \}$, $V = \{ \mathbf{v_1}, \mathbf{v_2}, \cdots \mathbf{v_r} \}$, $T = \{ \mathbf{t_1}, \mathbf{t_2}, \cdots \mathbf{t_r} \}$, $\Pi = \{ \pi_1, \pi_2, \cdots \pi_r \}$, and Ω is the index set of non-missing elements of $\mathcal{X}$. z_{ijk} decides ε_{ijk} on the clustering labels in the model. The pixels in adjacent space-time tend to similar z_{ijk}. So the distribution of z_{ijk} under the smooth prior knowledge can be written as:

$$z_{ijk} \sim M(\mathrm{z}_{ijk} | \pi) \prod_{p,q,r \in \mathbb{N}\, i,j,k} \Psi \ z_{ijk}, z_{pqr} \tag{8}$$

where $\Psi \ z_{ijk}, z_{pqr} = \frac{1}{C} \prod_n \exp\left[\tau \ 2 z_{ijkn} - 1 \ \ 2 z_{pqrn} - 1 \right]$. τ and C are a positive scalar and a regular constant respectively. And $\mathbb{N}\ i, j, k$ is the neighborhood of the element i, j, k. When the value of z_{ijkn} and z_{pqrn} are same (0 or 1), the value of $\Psi \ z_{ijk}, z_{pqr}$ is higher. This term can make better use of the prior knowledge of smoothness. After adding a smooth prior, the distribution is written as:

$$p \ Z | \pi = \frac{1}{C} \prod_{i,j,k \in \Omega, n} \pi_n^{z_{ijkn}} \prod_{i,j,k \in \Omega, n} \prod_{p,q,r \in \mathbb{N}\, i,j,k} \exp\left[\tau \ 2 z_{ijkn} - 1 \ \ 2 z_{pqrn} - 1 \right] \tag{9}$$

The log likelihood function at this time is:

$$\max_{U,V,T,Z,\Pi,\Omega} \mathrm{L}\ \mathcal{X}\left|U,V,T,Z,\Pi,\Omega\right.$$
$$= \sum_{\mathrm{i},j,k\in\Omega,n} z_{ijkn}\left[\log\pi_n + \log \mathrm{N}\left(x_{ijk}\left|\sum_{d=1}^{r} u_{id}v_{jd}t_{kd}, \sigma_n^2\right.\right)\right] \tag{10}$$
$$+\tau \sum_{i,j,k\in\Omega,n}\ \sum_{p,q,r\in\mathbb{N}\ i,j,k} 2z_{ijkn}-1\ \ 2z_{pqrn}-1\ +c1$$

3. ALGORITHM IMPLEMENTATION

We use the variational EM algorithm to solve the model with hidden variables. By maximizing the lower bound $\mathrm{L}\ q\ Z$, approximating the posterior distribution $p\ Z|\mathcal{X}$ and $p\ \mathcal{X}$, the decomposition logarithmic edge probability is expressed as:

$$lnp\ \mathcal{X} = \mathrm{L}\ q\ Z\ + KL\ q\ Z\ \left\|p\ Z|\mathcal{X}\right. \tag{11}$$

where $\mathrm{L}\ q\ Z\ = \int q\ Z\ \ln\left\{\frac{p\ X,Z}{q\ Z}\right\}dZ$, and $KL\ \cdot$ represents the Kullback-Leiber difference. The second term of equation (11) is $-\int q\ Z\ \ln\left\{\frac{p\ Z|X}{q\ Z}\right\}dZ$. When $q\ Z$ equals to $p\ Z|\mathcal{X}$, the lower bound is the largest, but in this case the model loses its meaning. So we assume that the distribution of $q\ Z$ has a constraint family and then find the number of families that minimize the KL difference. Assume $q\ Z$ meets the following distribution:

$$q\ Z\ = \prod_{i,\mathrm{j},\mathrm{k}} q\ z_{ijk}\left|\gamma_{ijk}\right. \tag{12}$$

where $q\ z_{ijk}\left|\gamma_{ijk}\right. = \prod_n \gamma_{ijkn}^{z_{ijkn}}$ and $\sum_n \gamma_{ijkn} = 1$. γ is a variational parameter. The lower bound at this time is written as:

$$\mathrm{L}\ q\ Z\ = \sum_{\mathrm{i},j,k\in\Omega,n} \gamma_{ijkn}\left[\log\pi_n - \log\sqrt{2\pi\sigma_n} - \frac{\left(x_{ijk} - \sum_{d=1}^{r} u_{id}v_{jd}t_{kd}\right)^2}{2\pi\sigma_n^2}\right] \tag{13}$$
$$+\tau \sum_{i,j,k\in\Omega,n}\ \sum_{p,q,r\in\mathbb{N}\ i,j,k} 2\gamma_{ijkn}-1\ \ 2\gamma_{pqrn}-1\ - \sum_{i,j,k\in\Omega,n} \gamma_{ijkn}\log\gamma_{ijkn} + c1$$

We use the variational EM algorithm to solve this optimization problem, alternating iterations between the variation E step (represented by VE) and the variational M step (represented by VM).

VE: Calculate variational parameters γ_{ijkn}

$$\gamma_{ijkn} = \pi_n N\left(x_{ijk}\left|\sum_{d=1}^{r} u_{id}v_{jd}t_{kd}, \sigma_n^2\right.\right)\exp\left\{\sum_{p,q,r\in\mathbb{N}\ i,j,k} \gamma_{pqrn}\right\} \tag{14}$$

VM: Updating Gaussian model parameters $U,V,T,\prod,\Sigma$ by maximizing the lower bound

$$E_z p\ \mathcal{X},Z\left|U,V,T,\prod,\Sigma\right. = \sum_{i,j,k\in\Omega}\sum_{n=1} \gamma_{ijkn}\ \log\pi_n - \log\sqrt{2\pi\sigma_n} - \frac{\left(x_{ijk} - \sum_{d=1}^{r} u_{id}v_{jd}t_{kd}\right)^2}{2\pi\sigma_n^2} \tag{15}$$

The closed solution form of $\prod,\Sigma$ is:

$$m_n = \sum_{i,j,k} \gamma_{ijkn} \quad \pi_n = \frac{m_n}{\sum_n m_n}$$
$$\sigma_n^2 = \frac{1}{m_n} \sum_{i,j,k} \gamma_{ijkn} \left(x_{ijk} - \sum_{d=1}^{r} u_{id} v_{jd} t_{kd} \right)^2 \quad (16)$$

Then we substitute the value of equation (16) into equation (15) and solve the unknown parameters U,V,T. The solving process here is referred to Chen [16].

4. EXPERIMENT

We conducted a lot of experiments on actual sports scenes. Figure 1 is a video scene with foreground motion under natural light changes. For small background disturbances (such as slow changes in water waves), the background can be reconstructed well, and the temporal and spatial information of the video data can be used to reconstruct the background which used to be occluded by foreground targets.

Figure 2 - Figure 4 are for background reconstruction when indoor scene changes (normal motion, dark light, light change). Figure 2 shows the background reconstruction with foreground motion under normal lighting. Figure 3 shows the video background reconstruction in dark light with ambiguous motion foreground and background colors. Our algorithm can distinguish the foreground from the background and reconstruct the complete background. Figure 4 shows a video scene with a strong change in light. Our method based on low-rank tensor decomposition uses mixture of Gaussian modeling for foreground, which is robust to light changes and small motion disturbances, and can effectively reconstruct the complete background.

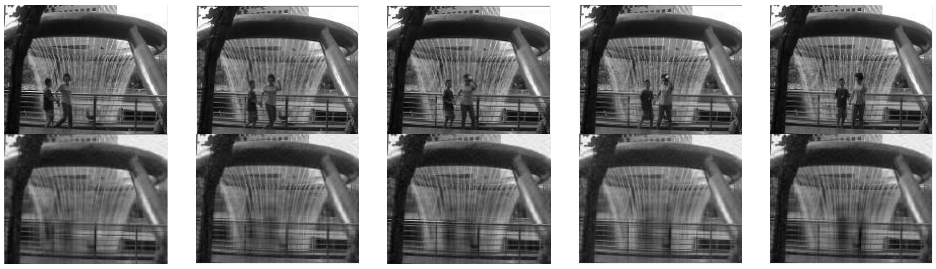

Figure 1. Actual video sequence background rerereconstruction, the first row: 5 consecutive frames, the second row: the results of background rerereconstruction

Figure 2. The indoor background rereconstruction with moving objects, the first row: 6 consecutive frames, the second row: the results of background rerereconstruction

5. CONCLUSION

The video sequence obtained by the fixed camera has a low-rank background, and the motion foreground can be regarded as sparse noise. Based on these characteristic properties, in this paper, we propose a new background reconstruction method based on the internal structural information of the video data. Based on the low-rank tensor decomposition, the mixture of Gaussian and Markov random fields are used to model the motion foreground, at the same time the prior knowledge of the local continuity of the motion foreground is utilized. Our model uses a variational EM algorithm for optimization calculations. A large number of comparative experiments prove that our algorithm is effective.

However, there are ghosting phenomena in the results of partial background reconstruction. Our next work plan is to remove ghosts and make the background better reconstructed.

Figure 3. Background rereconstruction under dark light, the first row: 6 consecutive frames, the second row: the results of background rereconstruction

Figure 4. Background rereconstruction with illumination changes, the first row: 6 consecutive frames, the second row: the results of background rereconstruction

REFERENCES

[1] Kaewtrakulpong P, Bowden R. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection[M] Video-Based Surveillance Systems. Springer US, 2002:135-144.

[2] Bouwmans T, El Baf F, Vachon B. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey[J]. Recent Patents on Computer Science, 2008, 1(3):219-237.

[3] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking[C] cvpr. IEEE Computer Society, 1999:2246.vol 2

[4] Lee D S. Effective Gaussian mixture learning for video background subtraction[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(5):827-32.

[5] Zivkovic Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction[C] International Conference on Pattern Recognition. IEEE Computer Society, 2004:28-31.

[6] Pilet J, Strecha C, Fua P. Making Background Subtraction Robust to Sudden Illumination Changes[M] Computer Vision – ECCV 2008. Springer Berlin Heidelberg, 2008:567-580.

[7] Kim K, Chalidabhongse T H, Harwood D, et al. Background modeling and subtraction by codebook rereconstruction[C] International Conference on Image Processing. IEEE, 2004，3061-3064 Vol. 5.

[8] Ilyas A, Scuturici M, Miguet S. Real Time Foreground-Background Segmentation Using a Modified Codebook Model[C] Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE Computer Society, 2009，454-459.

[9] Kyungnam Kim, Thanarat H. Chalidabhongse, David Harwood, & Larry Davis. Real-time foreground-background segmentation using codebook model[J]. Real-Time Imaging, 2005，11(3), 172-185.

[10] Guo J M, Hsu C S. Hierarchical method for foreground detection using codebook model[C]. IEEE International Conference on Image Processing. IEEE, 2010:804-815.

[11] Barnich O, Van D M. ViBe: a universal background subtraction algorithm for video sequences.[J]. IEEE Transactions on Image Processing, 2011, 20(6):1709.

[12] Heikkil M, Pietikinen M. A Texture-Based Method for Modeling the Background and Detecting Moving Objects[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(4):657.

[13] Liao S, Zhao G, Kellokumpu V, et al. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes[C] Computer Vision and Pattern Recognition. IEEE, 2010:1301-1306.

[14] F Li, H Zhou. A two layer video background modeling method that integrates codebook and texture[J] USTC, 2012,42(2):99-105.

[15] Meng D, Torre F D L. Robust Matrix Factorization with Unknown Noise[C] IEEE International Conference on Computer Vision. IEEE Computer Society, 2013:1337-1344.

[16] X Chen, Z Han, Y Wang, Q Zhao, D Meng,and Y Tang, Robust tensor factorization with unknown noise[C].in proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016, 5213-5221

[17] Cao X, Zhao Q, Meng D, et al. Robust Low-rank Matrix Factorization under General Mixture Noise Distributions[J].IEEE Transactions on Image Processing, 2016, 25(10):4677-4690.