AN ADVANCED CODEBOOK BACKGROUND MODEL USING CONFIDENCE AND MEMBERSHIP FUNCTION

Yuanyuan Cui¹, Xiaoqiang Guo², Wan Liu¹ and Aidong Men¹

¹ Beijing University of Posts and Telecommunications, Beijing, China ²Academy of Broadcasting Science, Beijing, China

ABSTRACT

This paper introduces an advanced codebook model for foreground-background segmentation. The improvement is two-fold. First, to cope with changing global lumination, we develop the conventional codebook with confidence functions. Similarities of both brightness and normalized color-vector are integrated confidence-weightedly to form overall similarity. A designed codeword-update progress also contributes to the stable performance. Besides, by introducing a membership function based measurement, the threshold of overall similarity is able to adjust itself with statistical properties of the given video. It makes the model rather robust. A thorough evaluation is performed on the Wallflower dataset. Qualitative and quantitative results and comparisons with other approaches justify the model.

Index Terms— Codebook, confidence function, membership function, background modelling

1. INTRODUCTION

The ultimate goal for a video surveillance system is to identify given objects automatically. To achieve the desired performance, background subtraction is a crucial and fundamental step. It contributes to differentiating current image from a reference background model.

There are multiple methods for background subtraction. Authors of [1] proposed to process images at pixel, region, and frame levels. Mixture of Gaussians (MOG) algorithm [2] and Kernel algorithm [14] show an excellent performance when the background scene is highly dynamic. Codebook (CB) [4] is developed by Kim et al. in 2005.

In [5], the CB and cache CB methods are both modified by bringing in coefficients of frequency. Authors of [6] proposed a multi-scale multi-feature CB for challenging surveillance environments. Another extension easing memory burden and avoiding high computational complexity is introduced in [7].In [8], the update process is selective and blind. Both a short-term model and a long-term model are maintained. Authors of [4] have developed a temporary layer of short-term background over the permanent layer of longterm background. Ref. [9] dealt with segmentation task with an approach using hue clustering and took the special cyclic property of the hue cell into consideration.

In this paper, a novel codebook model is proposed to cope with the two problems mentioned above. We develop the conventional codebook with confidence functions. Similarities of both brightness and normalized color-vector are integrated confidence-weightedly to form overall similarity. Then, a background membership function is constructed from former segmentations. It may help approximate the threshold of overall threshold.

The remainder of this paper is organized as follows. Section 2 introduces how the codebook model works. In Section 3, we describe the proposed approach, which combined brightness and normalized color-vector to calculate a overall similarity. Section 4 mainly introduces how the background membership is constructed and playing a role. Section 5 compares our results on a suite of videos with other approaches. Section 6 and 7 are conclusions and references respectively.

2. REVIEW OF THE CODEBOOK MODEL

The codebook model can effectively quantizes sample background values at each pixel into codebooks.For each pixel, the largest quantity of codewords is set as n_c . In a training stage of $t_{training}$ seconds, codewords will not be deleted until the limit is reached. At last, the smallest observation gap t_{min} for codewords in a certain position is computed. If a codeword doesn't appear whithin $t_{min} + 0.2t_{training}$ seconds, it will be deleted from the list.

If a codeword doesn't repeat during $t_{updating}$ seconds, it will be deleted. When all the n_c codeword slots are occupied, a new set of features may be rejected. In this case, codeword with the largest observation gap is removed to make room for the newcome. From begining to end, color distortion and brightness are used to detect a background pixel. For a codeword c, the main features are a mean vector $\overline{v} = (\overline{R}, \overline{G}, \overline{B})$, I and \hat{I} (the *min* and *max* brightness respectively, of all pixels assigned to the codeword). For a observed pixel f, the standard for brightness is

$$brightness(I_f, \langle \check{I}, \hat{I} \rangle) = \begin{cases} true & I_{low} \le ||f|| \le I_{high} \\ false & otherwise \end{cases}$$
(1)

$$I_{low} = \alpha \cdot \hat{I}, \ I_{high} = min\{\beta \hat{I}, \frac{\hat{I}}{\alpha}\}$$
(2)

where α and β are pre-defined parameters. A empirical combination is $0.7 \le \alpha \le 0.8, 1.15 \le \beta \le 1.25$ [10]. Then, the color distortion is formulated as

$$colordist(c, f) = \sqrt{\left\|f\right\|^2 - \gamma^2}$$
 (3)

$$\gamma^{2} = \frac{\langle f, \overline{v} \rangle^{2}}{\|\overline{v}\|^{2}} = \frac{(R_{f} \cdot \overline{R} + G_{f} \cdot \overline{G} + B_{f} \cdot \overline{B})^{2}}{(\overline{R})^{2} + (\overline{G})^{2} + (\overline{B})^{2}}$$
(4)

Finally, the two conditions for matching codeword c are

$$\begin{cases} brightness(I_f, \left\langle \check{I}, \hat{I} \right\rangle) = true\\ colordist(c, f) \le \varepsilon \end{cases}$$
(5)

where ε is the detection threshold.

3. THE CONDITION-BASED CODEBOOK MODEL

The conventional codebook model measures color distortion and brightness separately. However, brightness is very sensitive to global illumination changes, which makes the judgement more difficult. So, the proposed method combined the two features into a general similarity function, and introduce confidence into the model.

For a codeword c, we can mark it by two features as $c = \{c_I, c_C\} = \{|v|, \frac{v}{|v|}\}$. Analogously, a coming pixel is noted by its observed features as $f = \{f_I, f_C\}$. Each component k in $K = \{I, C\}$ (brightness and color respectively), owns a *similarity* function and a *confidence* function. The former one is $\sin_k(c, f) \in [0, 1]$, which shows how well the current codeword matches the pixel. The latter one is $\operatorname{conf}_k(c, f) \in [0, 1]$, which is equal to the expected accuracy of the corresponding similarity. Another parameter is λ_k , acting as a sensitivity factor.

3.1. The Overall Similarity

The proposed method takes the harmonic mean of the two similarities as the overall similarity sim(c, f). Harmonic mean is susceptible to extreme values, especially the minimums. Therefore, an extreme mismatch in either the brightness feature or the color feature will dominate the overall similarity. Both the two features are weighed by confidence function and sensitivity parameter:

$$\sin(c, f) = \frac{\sum_{K}^{k} \lambda_k \operatorname{conf}_k(c, f) + 0.05}{\sum_{K}^{k} \lambda_k \operatorname{conf}_k(c, f) / \operatorname{sim}_k(c, f) + 0.1}$$
(6)

The two small constants guarantee that denominator isn't zero in any case, and drive the overall value towards 0.5 when neither of the two features is confident. A coming pixel will be compared with every codeword to compute the similarities. Only the ones meet certain demand (introduced in the next section) are valid and the corresponding codewords are viewed as matched. Among them, the one with the highest similarity will be chosen by the algorithm. A computed result over 0.95 is considered as significant enough and the search will come to an early end. If there is not a valid match, a new codeword is created. Otherwise, the selected codeword will be updated weightedly as $c = \alpha_0 c + (1 - \alpha_0) f$.

3.2. The Similarity of Brightness

Brightness is the dominant local discriminator in foregroundbackground segmentation. However, it is so sensitive to lumination changes that even a local variation may lead to misidentification. Therefore, the model copes with variant lumination separately in another subsection. The similarity of brightness is formulated as:

$$\sin_B(c,f) = \frac{\delta_B^2}{\delta_B^2 + \lambda_B(c_B - f_B)^2} \tag{7}$$

A higher λ_B means the similarity is more sensitive to brightness change. $\delta_B = 25$ is a constant scaling factor. As brightness is the foremost, its confidence is set to the maximum: $\operatorname{conf}_B(c, f) = 1.0$.

3.3. The similarity of Color

The proposed model takes normalized vector $c_C = \frac{v}{|v|}$ to describe colors, rather than origin RGB vector which contains brightness information. Thus, the feature is reasonalby robust to variant luminance and additive image noise. The similarity of color is defined in a way similar to that of brightness:

$$\sin_C(c, f) = \frac{\delta_C^2}{\delta_C^2 + \lambda_C color dist(c, f)^2}$$
(8)

where λ_C is the sensitivity to color change and δ_C is a constant scaling factor. As colordist(c, f) ranges in [0, 1], the scaling factor δ_C is set as 1.

The grayer the two colors are, the lower the color confidence is. In this case, colors with low saturation will contribute less to the overall similarity. We take gray(v) =



Fig. 1. Cumulated percentages of background and foreground in two different images.

 $1 - colordist(v, c_w)$ ($c_w = \frac{\sqrt{3}}{3}(1, 1, 1)$ is pure white) as the grayness. Confidence function for color is calculated as:

$$\operatorname{conf}_C(c, f) = 1 - \operatorname{gray}(c)\operatorname{gray}(f) \tag{9}$$

3.4. Updating with lumination changes

Sudden changes in global illumination will make large-scale "fake" foreground in the vision. Robust features alone are not enough, an appropriate codeword-updating mechanism is essential. The modified system tracks the mean intensity of illumination. Fast changes will trigger a brief relearning progress which accelerates the learning of new codewords. Specifically, the time required to learn a new codword is reduced.

4. THE MEMBERSHIP FUNCTION-BASED MEASUREMENT OF SIMILARITY THRESHOLD

Inspired by clustering techniques described in [11], this paper proposed a novel method to measure the threshold of overall similarity. It is based on fuzzy logic and employs a membership function to approximate an optimal threshold for foreground-background segmentation. For each frame $F_{x,y}$, \sin^w is the overall similarity between observed feature and the *v*th of n_c codewords. The similarity is defined as

$$\mathbf{s} = max(\sin^1, \sin^2 \dots \sin^w), \forall w \tag{10}$$

An informal experiment is carried out to facilitate the explanation. The result with different images is shown in Fig. 1. It is obvious that background pixels are inclined to cluster to higher similarity. Thus, we constructed two memberships in the schematic diagram Fig. 2. The right one is the background membership function(MB) and the other is the foreground membership function(MF). For a similarity under s_0 , the background membership function equals 0. With the similarity increasing, background membership increases



Fig. 2. Fuzzy membership functions: MB() and MF().

too. The uptrend can be modeled as a cosine curve. Then, we can calculate $s_{x,y}$, value of membership function in position (x, y). First, the algorithm computes the threshold similarity s_0 by accumulating the first q_{mb} pixels ranging in $[0, s_0]$. s_0 means the interval at which MB(s) is evaluated to 0. Then, the cardinality of the pixels at a similarity s_i is calculated as

$$Card(s_i) = Card(F_{x,y}|s_{x,y} \le s_i) \tag{11}$$

Iteratively, the algorithm searches for a proper similarity precise enough to discover small augmenters. This process may be described as computing the rise rate from similarity s_i to s_{i+1} as $m_i = \frac{Card(s_i) - Card(s_{i+1})}{s_{i+1} - s_i}$. Once gotten m_i , the period for our designed cosine curve, which is T = 2/m. Taking the range of $[s_0, s_1]$ for instance, s_1 is the third consecutive color distortion whose angle of inclination $\beta \leq \beta_0$. In this way, foreground miss-identification caused by noise is effectively avoided. Finally, the background membership function can be formulated as

$$MB(s) = \begin{cases} 0 & 0 \le s \le s_0\\ \frac{1 - \cos[m\pi(s - s_0)]}{2} & s_0 < s \le s_0 + \frac{1}{m} \\ 1 & s_0 + 1/m < s \end{cases}$$
(12)

An observed feature will be identified as background only when it meets MB(s) > 0.5.

5. PERFORMANCE EVALUATION

In this section, sufficient experiments were carried out to evaluate the effectiveness of the proposed method. Codebook algorithm is as effective as probabilistic classification, and performs better in both memory and speed [4]. Thus, the proposed model is only compared with the original codebook and a robust one published in ICASSP 2010 [14]. The latter works well with dynamic background and illumination changes thanks to a pseudo background layer and a two-step update.

Sequence	Item	original codebook	codebook from [14]	proposed codebook
Bootstrap	TPR	0.72	0.88	0.90
	FPR	0.38	0.15	0.13
Camouflage	TPR	0.79	0.86	0.88
	FPR	0.15	0.08	0.06
Foreground	TPR	0.77	0.82	0.86
Aperture	FPR	0.15	0.13	0.10
LightSwitch	TPR	0.69	0.80	0.86
	FPR	0.65	0.35	0.31
MovedObject	TPR	0.87	0.96	0.98
	FPR	0.15	0.04	0.04
TimeofDay	TPR	0.83	0.89	0.97
	FPR	0.21	0.13	0.11
MovingTrees	TPR	0.90	0.99	0.98
	FPR	0.18	0.07	0.06

Table 1. Comparison of three different methods.

Sequences for test. Wallflower benchmark is a recognized validity test, which successfully cover most of the major problems in forground-background segmentation.

Evaluation methods. Results of 20 frames per video are collected to demonstrate the improved performance. Both segmented frames and receiver operating characteristic (ROC) [10] are used. Four fundamental parameters are calculated by comparing the segmented results with manual labeled ground truths. They are true positive (TP), false positive (FP), true negative (TN) and false negative (FN). A quantitative evaluation is performed using TPR and FPR:

$$FPR = \frac{FP}{TN + FP}, \ TPR = \frac{TP}{TP + FN}$$
 (13)

Parameters. The proposed method is performed under a set of consistent parameters as $n_c = 10$, $t_{updating} = 200$, $\alpha_0 = 0.99$, $\lambda_B = 2$, $\lambda_C = 1$, $q_{mb} = 30\%$ of pixels and $\beta = 1^o$. The other two methods are performed with the parameters stated by their authors.

Results and Discussion. TPR and FPR of all the seven sequences in Wallflower benchmark are listed in Table 1. A higher TPR value reveals that more foreground pixels are correctly identified. A lower FPR value means that fewer pixels are labeled as foreground. With the background membership function, different videos no longer share the same fixed threshold value. It makes the proposed algorithm adaptive to the certain characteristics of a test video. Thus, the proposed codebook offers a more satisfactory segmentation than the original codebook and the modified version in [14]. EsOriginal Groundtruth Codebook CB in [14] Proposed



Fig. 3. Results for the Wallflower benchmark.

pecially, in LightSwitch and TimeofDay, the improvement is more significant than the other five videos. This is because the combined-feature similarity and the method for codeword updating are designed for changing global lumination, which is just the case as the two mentioned sequences. Figure 3 shows the segmented images for all the videos.

6. CONCLUSIONS

In this paper, we presented a novel codebook-based model for background subtraction. The conventional codebook approach is developed with confidence functions. Similarities of both brightness and normalized color-vector are integrated confidence-weightedly to form overall similarity. In addition, a background membership function is constructed and the threshold of overall similarity may change with different statistical properties of the video. The improvements realize a precise segmentation, especially when the global illumination is not stable. Experiments on wallflower benchmark demonstrated the better performance, with both quantitive and qualitative results are listed.

7. ACKNOWLEDGEMENT

This work is supported by the National High Technology Research and Development Program of China(863 Program) under Grant No. 2011AA01A103, and the National Science Foundation of China (61271190).

8. REFERENCES

- K. Toyama, J. Krumm, B. Brumitt, and B. Meyers,"Wallflower: Principles and practice of background maintenance," IEEE ICCV, pp. 255C261, 1999.
- [2] S. Cheung and C. Kamath, "Robust Techniques for Background Subtraction in Urban Traffic Video," in Proc.Video Communications and Image Processing, SPIE Electronic Imaging, San Jose, Calif, USA, January 2004.
- [3] A. Elgammal, D. Harwood, and L. Davis, "Nonparametric Model for Background Subtraction," European Conference on Computer Vision, 2000.
- [4] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis,"Real-time Foreground-background Segmentation Using Codebook Model," Real-Time Imaging, vol. 11, pp.172C185, June 2005.
- [5] J. Yao et al., "Multi-Layer Background Subtraction Based in Color and Texture," IEEE CVPR, pp. 1-8, 2007.
- [6] A. Zaharescu and M. Jamieson, "Multi-Scale Multi-Feature Codebook-Based Background Subtraction," IEEE ICCV, pp. 1753-1760, 2011.
- [7] L. Geng and Z. Xiao, "Real Time Foreground-Background Segmentation Using Two-Layer Codebook Model," IEEE CASE, 2011.
- [8] A.M. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance," Proceedings of the IEEE, vol. 90, pp. 1151C1163, 2002.
- [9] C. Zhang and P. Wang, "A New Method of Color Image Segmentation Based on Intensity and Hue Clustering," IEEE ICPR, vol. 3, 2000.
- [10] A. Ilyas et al., "Real-Time Foreground-Background Segmentation Using a Modified Codebook Model," 6th IEEE AVVS, pp. 454-459, 2009.
- [11] J. Rosell-Ortega, G. Garcia-Andreu, A. Rodas-Jorda and V. Atienza-Vanacloig, "A Combined Self-configuring Method for Object Tracking in Colour Video," IEEE ICPR, pp. 2081-2084, 2010.
- [12] I. Sun, S. Hsu, and C.Huang, "A Hybird Codebook Background Model for Background Subtraction," IEEE SiPS, pp. 96-101, 2011.
- [13] I. Colman, A. Rhuma, M. Yu and J. Chambers, "A Robust Technique for Person-background Segmentation in Video Sequences Based on the Codebook Method of Background Subtraction and Head Tracking," IEEE SSPD, 2010.

[14] A. Pal, G. Schaefer and M. Emre Celebi, "Robust Codebook-based Video Background Subtraction," IEEE ICASSP, 2010.