ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Fusion of Foreground Object in Domain Motion Information for Video Summarization

R. Badmapriya [1], D. Kirubha [2]

M.E Student (CSE), Sree Sowdambika College of Engineering, Chettikurichi, Aruppukottai, Virudhunagar District, Tamil Nadu , India [1]

Associate Professor (IT), Sree Sowdambika College of Engineering, Chettikurichi, Aruppukottai, Virudhunagar District, Tamil Nadu, India [2]

ABSTRACT: Surveillance video camera captures a lot of consistent video stream each day. To examine or research any huge occasions from the tremendous video information, it is arduous and exhausting occupation to distinguish these occasions. In order to solve this problem, a video summarization technique combining foreground objects and movement data in spatial and frequency domain is presented in this paper. We remove foreground object utilizing foundation demonstrating and movement data in both spatial and frequency domain. Frame transition is connected for acquiring movement data in spatial domain. For obtaining movement data in frequency domain, phase correlation (PC) strategy is connected. Then, foreground objects and movements in spatial and frequency domain are combined and key frames are separated. Experimental results discover that the proposed strategy performs superior than technique.

I. INTRODUCTION

In our routine life a tremendous measure of surveillance video is caught 24 hours all through the entire world for giving security, monitoring, preventing crime, controlling traffic etc. Generally various surveillance video cameras are set up in various distinction spots of a building, business or congested zone. These cameras are associated with a monitoring cell for storage and investigation. To store this gigantic volume of video information necessitates huge memory space. Notwithstanding this, to discover any vital occasions from the stored video for examining or performing investigation, administrators need to get to the stored videos. This procedure is extremely slow, lengthy and costly. To solve of these issues, a strategy for producing the shorter version of original video containing critical occasions is very suitable for memory administration and data recovery.

Video summarization (VS) is the method to choose the most informative frames so that it could contain all the essential occasions and reject superfluous content to build the summarized video as compact as could reasonably be expected. In this manner, a great video summarized strategy is one that has a few critical properties. To start with, it must have the capacity to incorporate incidents inside of the first video. Second, it ought to have the capacity to create a littler variant of the rendered long video. Third, it ought not to contain tedious data. The fundamental motivation behind VS is to represent to a long unique video in a consolidated version in a manner that a user can get the entire thought of the whole video in a compelled measure of time.

In a video, foreground objects usually comprise more detail data [1]. Once more, human typically focus more on the developments of items [2]. Thus, objects and in addition their movement are critical components for a video. Propelled by these discoveries, a video summarization technique is proposed in this paper based on objects and their movement in a video. To incorporate foreground object data, Gaussian mixture-based parametric background modeling (BGM) [3] has been employed.

To adopt the complete data of item movement in a video, object movement is removed not only in spatial domain as well as in frequency domain as well. To get movement data in spatial domain, successive frame contrast is employed. For accomplishing object movement in frequency, phase correlation procedure [4] is needed. To the best of our

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

knowledge, phase correlation is not required for video summarization strategies. Hence, the principle commitment of this paper is to apply phase correlation in video summarization technique. The computational time of phase correlation is low and rich movement data is acquired by phase correlation system [4].
The structure of the remaining paper is as per the following. Area 2 audits related examination. The proposed technique is depicted in section 3. Experimental results are given in section 4. At last, a conclusion is given in section 5.

II. RELATED WORK

In the literature writing, diverse methodologies have been proposed for summarizing different sorts of videos. For egocentric video summarization, locale saliency is anticipated in [5] utilizing a regression model and storyboard are created taking into account on region significance score. In the technique proposed in [6], story driven egocentric video is summarized by finding the most influential objects inside of a video. Gaze tracking data is utilized in [7] for summarization. If there should arise an occurrence of client produced video summarization, adaptive sub modular maximization function is applied in [8]. A collaborative sparse coding model is used in [9] for creating summary of the same sort of videos. Web pictures are utilized as a part of [10] to upgrade the procedure of summarizing the user created video. To summarize movie, aural, visual and textual are converged in [11]. Role community network is employed in [12]. Film comic is created utilizing eye tracking information in [13]. Notwithstanding these, systems proposed in [14][15][16][17] for wireless capsule endoscopic video summarization.
Be that as it may, the significance of surveillance video for industrial application is exceptionally higher than different sorts of videos (e.g., egocentric, user created, motion picture, and so forth). To compress surveillance video, object focused system is utilized in [18]. Dynamic Video Book is proposed in [19] for presenting to the surveillance video in a hierarchical order. Learned separation metric is presented in [20] for summarizing nursery school surveillance video. In [21], salient motion data is connected. Maximum a posteriori probability (MAP) is utilized as a part of [22] for synopsis generation. Now-a-days, a technique is proposed in [1] for a multi-view surveillance video summarization. Firstly, a single view summarization is created in this methodology for every sensor autonomously.
For this reason, MPEG-7 color format descriptor is required to every video frame and an online-Gaussian mixture model (GMM) is utilized for clustering. The key frames are chosen based on the parameters of cluster. As the choice of selecting or dismissing a frame is performed based on the consistently updates of these clustering parameters, a video segment is extricated rather than key frames. In conclusion, multi-view summarization is delivered by applying distributed view selection technique utilizing the video segments removed for every sensor in the past step.
To the best of our knowledge, phase correlation technique has not employed for video summarization. In this proposed technique, phase correlation approach is needed to incorporate movement data in frequency domain and fused with moving foreground object and spatial movement data.

III. PROPOSED WORK

The proposed technique depends on region of moving foreground objects and their movement data in spatial and frequency domain. The primary steps of the proposed technique are (1) moving foreground object extraction (2) movement data count in spatial domain, (3) movement approximation in frequency domain, (4) combination of foreground object range and spatial and frequency movement data, and (5) video summary generation. The flow chart of the proposed technique is presented in Fig. 1. The detail of every step is clarified in the subsequent sub-segment.

3.1. Foreground Object Extraction

In the proposed technique, Gaussian mixture-based parametric BGM [3] is employed. In this parametric BGM, every pixel is displayed by the K Gaussian distributions (K=3) and each Gaussian model addresses either static background or dynamic foreground object on time frame. For instance, assume a pixel intensity x_t at time t is demonstrated by k^{th} Gaussian with recent measure γ_k^t, mean μ_k^t, standard deviation σ_k^t and weight ω_k^t such that $\sum \omega_k^t = 1$. The learning parameter α is utilized to redesign parameter measures, for example, mean, standard deviation, and so on. Toward the starting, the framework contains unfilled arrangement of Gaussian models. In the wake of discovering the

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

primary pixel (t=1), another Gaussian model (K=1) is created with $\gamma_k^t = \mu_k^t = x_t$, standard deviation $\sigma_k^t = 30$ and weight $\omega_k^t = 0.001$. At that point for each new notification of pixel intensity X_t of the same area at t, it attempts to locate a coordinated model from the current models such that $| x_t - \mu_k | \leq 2.5\sigma_k$.

In order to get gray scale background frame, background modeling [3] is employed in the wake of converting every color frame into gray scale picture. At that point, a color video frame at time t is converted into gray picture I(t) and subtracted from the relating gray background frame B(t) acquired by the background modeling. A pixel is took as a foreground pixel and fix the measure to one, if the pixel intensity difference between I(t) and B(t) is more than or equivalent to a specific threshold (Thr1). In the event that the pixel intensity does not fulfill this condition, it is viewed as a background pixel and set to zero. In such manner, a foreground area pixel $G_{i,j}(t)$ is gotten as follows

$$G_{i,j}(t) = \begin{cases} 1 & if\ | I_{i,j}(t) - B_{i,j}(t) | \geq Thr1 \\ 0 & otherwise \end{cases} \quad (1)$$

where (i, j) is the pixel position. The estimation of Thr1 is set to 20 to keep away from inconspicuous changes between background and foreground. This is a typical practice to set the threshold limit to 20 to distinguish object from the background as said in [24]. Then, the aggregate number of non-zero pixels in $G_{i,j}(t)$ is utilized as region of foreground object feature F(t) that is gotten by the accompanying equation where r and c demonstrate the row and column of F respectively.

$$G(t) = \sum_{i=1}^{r} \sum_{j=1}^{c} G_{i,j}(t) \quad (2)$$

As indicated by the psychological theories of human consideration, movement data is huger than the static consideration hints [2]. Consequently, movement data is incorporated into the proposed technique notwithstanding the foreground object.

3.2. Motion Information Calculation in Spatial Domain

Generally, a human focus more on the motion of articles in a video [2]. With a specific end goal to acquire object movement data in spatial area, edge to-edge distinction is connected. Assume two consecutive frames such as I(t-1) and I(t) at time t-1 and t in video. In order to discover spatial movement data, the color contrast in red, green, and blue channel among these frames is evaluated. In the event that the distinctions at every pixel in three unique channels are more or equivalent to a threshold, this type of pixel is thought as movement pixel and set to esteem one. Else, it is certain that this pixel does not contain any movement data. Consequently, the spatial movement data $S_{i,j}(t)$ in pixel (i,j) at time t can be acquired by the following equation

$$S_{i,j}(t) = \begin{cases} & if\ | I_{i,j}^r(t) - I_{i,j}^r(t-1) \geq T2 \\ 1, & and\ | I_{i,j}^g(t) - I_{i,j}^g(t-1) \geq T2 \\ & and\ | I_{i,j}^b(t) - I_{i,j}^b(t-1) \geq T2 \\ 0, & otherwise \end{cases} \quad (3)$$

where, $I_{i,j}^r$, $I_{i,j}^g$, and $I_{i,j}^b$ show red, green and blue colors at (i,j)channels in a respective manner. This is a typical practice to set 20 as a threshold limit to get data between two consecutive frames [24]. In this way, the estimation of T2 is set to 20. The spatial movement data S(t) is acquired at time t by summing all qualities in $S_{i,j}(t)$ as follows

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

$$S(t) = \sum_{i=1}^{r}\sum_{j=1}^{c} S_{i,j}(t) \qquad (4)$$

where r and c demonstrate row and column of S respectively.

In any case, movement extricated in spatial domain is not delicate to diffuse event [25]. For instance, in the event of worldwide light changes, it doesn't function admirably. Furthermore, spatial movement approximation is inclined to local inaccuracies and little movement discontinuities [26].

3.3. Motion Estimation in Frequency Domain

In order to defeat the issue of spatial movement figuring, movement data is ascertained in frequency domain. Movement assessed in frequency domain has a few advantages over spatial domain [25]. It is proficient for worldwide changes of illumination and powerful to movement estimation close object limits. To acquire movement data in frequency domain, every frame is isolated into various blocks of 16×16 pixels estimate. At that point, phase correlation technique [4] is employed between the current block and reference block. The phase correlation peak (β) is extracted from phase correlation technique is utilized as movement indicator for that block. The phase difference (ϕ) is figured between the current block and its co- located reference block after employing Fast Fourier Transformation (FFT) on every block. The inverse FFT is performed on the figured phase difference and lastly two dimensional (2-D) movement vector (dx, dy) is gotten [27]. This 2-D movement vector contains peaks (□) where there are move between the current and reference blocks.

$$\phi = fftshift\,|\,ifft(e^{j(\angle F_r - \angle F_c)}\,|\qquad (5)$$

where Fr and Fc demonstrate FFF of current and reference block respectively

$$(dx, dy) = \max(\phi) - b/2 - 1 \qquad (6)$$

$$\beta = \phi(dx + b/2 + 1, dy + b/2 + 1) \qquad (7)$$

where b is block size. For instance, b will be 16 if 16×16 is applied.

If the value of □ of a block greater than a threshold (T3), it is considered that this block contains sufficient motion information. In this method, the value of T3 is set to 0.6. All the values greater than T3 are summed to obtain motion information F(t) in frequency domain.

In the event that the estimation of β of a block more prominent than a threshold (T3), it is viewed as that this block comprises adequate movement data. In this technique, the estimation of T3 is set to 0.6. All the measures are more prominent than T3 are summed to acquire movement data F(t) in frequency domain.

$$F(t) = \sum_{n=1}^{N}\sum_{m=1}^{M} F_{n,m}(t) \qquad (8)$$

where N and M show row/b and column/b of F respectively.

The movement data got in frequency domain utilizing phase correlation technique at various blocks of frame no 3869 of bl-14 video. No movement is demonstrated to in block (4,4) with just a single highest pick.

Conversely, frequency based movement estimation techniques absences of limitation issue [25]. In this way, movements acquired in both spatial and frequency domains are consolidated with moving foreground objects for creating video synopsis.

3.4. Fusion of Foreground and Motion Information

With a specific end goal to choose more exact frame sequences, both regions of foreground object and movement data are joined. In this approach, a weighted linear fusion is employed to join the features for positioning every frame as indicated by their example in a video. Before utilizing fusion technique, every feature is converted into z-score standardization applying the accompanying mathematical equation

$$Z(t) = (X(t) - \mu)/\sigma \qquad (9)$$

where is a feature esteem at time t,μ is the mean and σ is standard deviation of the feature measures. Z-score, Z(t) is a standardized type of X(t). In this technique, z-score standardization is the preferable technique since it produces

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

significant data about every information point, and gives better results in the vicinity of exceptions than min-max based standardization [28]. The weighted linear fusion is acquired as follows

$$A(t) = \phi_1 * Z_G(t) + \phi_2 * Z_S(t) + \phi_3 * Z_F(t) \quad (10)$$

where A(t) is fusion measure; $Z_G(t)$, $Z_S(t)$ and $Z_F(t)$ are z-score standardization of foreground feature (G(t)), spatial movement feature (S(t)), and movement data in frequency domain (F(t)) respectively at time t. Experimentally, it is assessed that if the estimations of weights , ϕ_1, ϕ_2 and ϕ_3 are set to 15, 60, and 25 respectively, it gives better results to all videos in BL-7F dataset. The discernment to give higher weight to movement feature contrasted with the foreground zone is that as indicated by the psychological theories of human consideration, movement data is more noteworthy than the static consideration intimations [2]. After that, A(1, ... T) is sorted base on descending order where T is absolute number of frames in a video.

IV. EXPERIMENTAL RESULTS

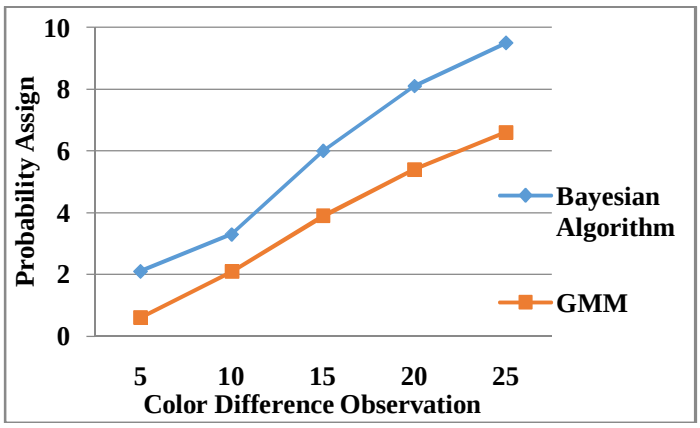

Figure 1: Color Difference Observation vs. Probability Assign

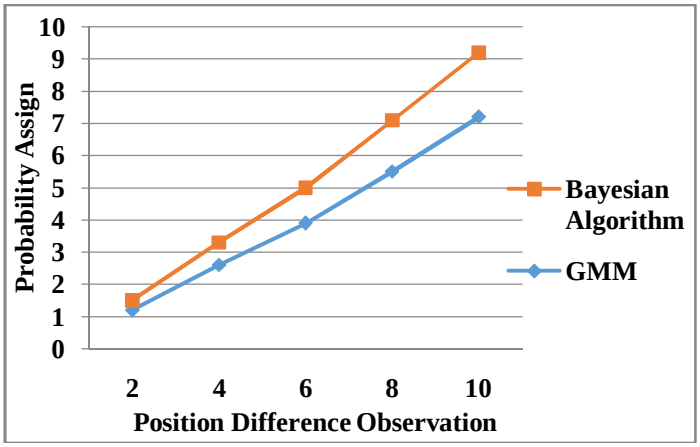

Figure 2: Position Difference Observation vs. Probability Assign

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

V. CONCLUSION

Thus, we proposed a novel system to summarize surveillance video combining foreground object with movement data in spatial and frequency domain. As indicated by [1], foreground object actually contain detailed data of the video contents. Also, a person actually gives more thoughtfulness regarding object movement in a video [2]. Accordingly, there two vital properties of a video are incorporated into this methodology. In order to incorporate movement data in frequency domain, phase correlation technique [4] is employed. To the best of our knowledge, phase correlation method is employed for the first time for video summarization. The experimental results show that the proposed technique beats the state-of-the-art strategy.

REFERENCES

1. Ou, S., LEE, C., Somayazulu, V., Chen, Y., Chien, S.: On-line Multi-view Video Summarization for Wireless Video Sensor Network. IEEE J. Sel. Top. Signal Process. 9, 165–179 (2015).
2. Gao, D., Mahadevan, V., Vasconcelos, N.: On the plausibility of the discriminant center-surround hypothesis for visual saliency. J. Vis. 8, 1–18 (2008).
3. Paul, M., Lin, W., Lau, C., Lee, B.: Explore and model better I-frames for video coding. IEEE Trans. Circuits Syst. Video Technol. 21, 1242–1254 (2011).
4. Paul, M., Lin, W., Lau, C.T., Lee, B.-S.: Direct intermode selection for H.264 video coding using phase correlation. IEEE Trans. image Process. 20, 461–73 (2011).
5. Lee, Y.J., Ghosh, J., Grauman, K.: Discovering Important People and Objects for Egocentric Video Summarization. IEEE Conf. Comput. Vis. Pattern Recognit. 1346–1353 (2012).
6. Lu, Z., Grauman, K.: Story-Driven Summarization for Egocentric Video. IEEE Conf. Comput. Vis. Pattern Recognit. 2714–2721 (2013).
7. Xu, J., Mukherjee, L., Li, Y., Warner, J., Rehg, J.M., Singh, V.: Gaze-enabled Egocentric Video Summarization via Constrained Submodular Maximization. IEEE Conf. Comput. Vis. Pattern Recognit. 2235–2244 (2015).
8. Gygli, M., Grabner, H., Gool, L. Van: Video Summarization by Learning Submodular Mixtures of Objectives. IEEE Conf. Comput. Vis. Pattern Recognit. 3090–3098 (2015).
9. Liu, Y., Liu, H., Sun, F.: Outlier-attenuating summarization for user-generated-video. IEEE Int. Conf. Multimed. Expo. 1 – 6 (2014).
10. Khosla, A., Hamid, R.: Large-scale video summarization using web-image priors. IEEE Conf. Comput. Vis. Pattern Recognit. 2698 – 2705 (2013).
11. Evangelopoulos, G.: Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. IEEE Trans. Multimed. 15, 1553–1568 (2013).
12. Tsai, C., Kang, L.: Scene-Based Movie Summarization Via Role-Community Networks. IIEEE Trans. Circuits Syst. Video Technol. 23, 1927–1940 (2013).
13. Sawada, T., Toyoura, M., Mao, X.: Film Comic Generation with Eye Tracking. Adv. Multimed. Model. 467–478 (2013).
14. Schoeffmann, K., Del Fabro, M., Szkaliczki, T., Böszörmenyi, L., Keckstein, J.: Keyframe extraction in endoscopic video. Multimed. Tools Appl. (2014).
15. Spyrou, E., Diamantis, D., Iakovidis, D.K.: Panoramic Visual Summaries for Efficient Reading of Capsule Endoscopy Videos. 2013 8th Int. Work. Semant. Soc. Media Adapt. Pers. 41–46 (2013).
16. Mehmood, I., Sajjad, M., Baik, S.W.: Video summarization based tele-endoscopy: a service to efficiently manage visual data generated during wireless capsule endoscopy procedure. J. Med. Syst. 38, 109 (2014).
17. Ismail, M. Ben: Endoscopy video summarization based on unsupervised learning and feature discrimination. Vis. Commun. Image Process. 1–6 (2013).
18. Fu, W., Wang, J., Zhao, C., Lu, H., Ma, S.: Object-centered narratives for video surveillance. IEEE Int. Conf. Image Process. 29–32 (2012).

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

19. Sun, L., Ai, H., Lao, S.: The dynamic VideoBook: A hierarchical summarization for surveillance video. IEEE Int. Conf. Image Process. 3963–3966 (2013).
20. Wang, Y., Kato, J.: A distance metric learning based summarization system for nursery school surveillance video. IEEE Int. Conf. Image Process. 37–40 (2012).
21. Mehmood, I., Sajjad, M., Ejaz, W., Wook, S.: Saliency-directed prioritization of visual data in wireless surveillance networks. Inf. Fusion. 24, 16–30 (2015).
22. Huang, C., Chung, P.J.: Maximum a Posteriori Probability Estimation for Online Surveillance Video Synopsis. IEEE Trans. Circuits Syst. Video Technol. 24, 1417–1429 (2014).
23. Chakraborty, S., Paul, M.: A novel video coding scheme using a scene adaptive non-parametric background model. IEEE 16th Int. Work. Multimed. Signal Process. 1–6 (2014).
24. Haque, M., Murshed, M., Paul, M.: A hybrid object detection technique from dynamic background using Gaussian mixture models. IEEE 10th Work. Multimed. Signal Process. 915 – 920 (2008).
25. Ahuja, N., Briassouli, A.: Joint Spatial and Frequency Domain Motion Analysis. Int. Conf. Autom. Face Gesture Recognit. 203–208 (2006).
26. Briassouli, A., Ahuja, N.: Integration of Frequency and Space for Multiple Motion Estimation and Shape-Independent Object Segmentation. 657–669 (2008).
27. Paul, M., Frater, M.R., Arnold, J.F.: An Efficient Mode Selection Prior to the Actual Encoding for H . 264 / AVC Encoder. IEEE Trans. Multimed. 11, 581–588 (2009).
28. Han, J., Kamber, M., Pei, J.: Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan Kaufmann (2006).