Advanced Research in Electrical and Electronic Engineering
p-ISSN: 2349-5804; e-ISSN: 2349-5812 Volume 2, Issue 12 July-September (2015) pp. 18-22
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/Publication.html

A Review of Dynamic Texture Clustering Technique for Video Modelling

Anagha Vijaysing Raghuwanshi[1] and Sudesh Gupta[2]

[1,2]Technocrats Institute of Technology & Science, Bhopal, India
E mail: [1]anagharaghuwanshi@gmail.com, [2]sudesh.gupta75@gmail.com

Abstract—*Video modelling is vital research area in the field of video tracking and video motion detection. The process of motion detection and video tracking is very challenging task. The texture feature of video is major part of analysis in segmentation and tracking. The dynamic nature of texture, the process of clustering and segmentation are very difficult. In this paper presents the dynamic texture clustering technique and problem of current system and video texture modeling. One significant limitation of the original dynamic texture is, however, its inability to provide a perceptual decomposition into multiple regions, each of which belongs to a semantically different visual process: for example, a flock of birds flying in front of a water fountain, highway traffic moving in opposite directions, video containing both smoke and fire, and so forth. One possibility to address this problem is to apply the dynamic texture model locally, by splitting the video into a collection of localized spatiotemporal patches, fitting the dynamic texture to each patch, and clustering the resulting models. However, this method, along with other recent technique, lacks some of the attractive properties of the original dynamic texture model.*

Keywords*: Video Modeling, Dynamic Feature, Clustering Technique, Segmentation.*

1. INTRODUCTION

Multimedia information systems are becoming increasingly important with the development of broadband networks, high-powered workstations, and compression standards. Since visual media requires large amounts of memory and computing power for storage and processing, there is a need to efficiently index, store, and retrieve the visual information from multimedia databases [1]This work focuses on video data, video representation and segmentation. Advanced video representation schemes are needed to enable compact video storage as well as a concise model for indexing and retrieval applications. Segmenting an input video stream into interesting "events" is becoming an important research objective. Motion is an important cue in object perception. While many static cues such as color or shape can be used to generate object hypotheses, common motion is a further fundamental grouping cue that is especially useful for active perception by robots. Many recommendation systems rely on semantic tags, which are words or short phrases that describe musically meaningful concepts such as genres, instrumentation, mood and usage. Exploiting the tremendous amount of multimedia data, and specifically video data, requires developing methods able to extract high level information related to semantic aspects. Video summarization, video retrieval, and video surveillance are examples of applications [4,5]. These objectives may involve different video processing issues. Most previously proposed auto-taggers rely either on discriminative algorithms, or on generative probabilistic models, including Gaussian mixture models (GMMs), hidden Markov models (HMMs) ,hierarchical Dirichlet processes (HDPs), code word Bernoulli average models (CBA), and dynamic texture mixture models (DTMs. The bag of features representation extracts audio features from the song at a regular time interval, but then treats these features independently, ignoring the temporal order or dynamics between them. The dynamic texture is a probabilistic generative model, defined over space and time, that represents a video (i.e., spatiotemporal volume) as the output of a linear dynamical system (LDS) [6].The model includes a hidden-state process which encodes the motion of the video over time and an observation variable that determines the appearance of each video frame, conditioned on the current hidden state. Both the hidden-state vector and the observation vector are representative of the entire image, enabling a holistic characterization of the motion for the entire sequence [7]. One major current limitation of the dynamic texture framework is, however, its inability to account for visual processes consisting of multiple, co-occurring, and dynamic textures. For example, a flock of birds flies in front of a water fountain, highway traffic moving in opposite directions, video containing both smoke and fire, and so forth [8]. The dynamic texture (DT) model, a generative time series model that captures longer-term time dependencies, for automatic tagging of musical content. The DT model represents a time series of audio features as a sample from a linear dynamical system (LDS), which is similar to the hidden Markov model (HMM) that has proven robust in music identification [10].A video object is then defined as a collection of video regions that have been grouped together under some criteria across several frames. Namely, a video object is a collection of regions exhibiting consistency across several frames in at least one feature. A video shot is decomposed into a set of sub-objects. The sub objects are obtained by tracking. Each sub-object consists of a sequence of tracked regions. The regions are obtained by segmentation. An HEM algorithm for clustering dynamic textures through their probability distributions [15].The resulting algorithm is capable of both clustering DTs and learning novel DT cluster centers that are representative of the cluster members, in a manner that is consistent with the underlying generative probabilistic model of the DT. We then demonstrate the efficacy of the clustering algorithm on several computer vision problems, hierarchical motion clustering; semantic motion annotation using weakly-supervised learning, codebook generation for the bag- of-systems motion representation. The above section discuss introduction of video modeling. In section 2 we describe related work of dynamic texture clustering. In section 3 discuss dynamic texture. In section 4 discuss problem formulation of dynamic video molding and finally conclude in section 5.

2. RELATED WORK

In this section describe the related work of video modelling based on dynamic texture data for clustering and segmentation process. Segmentation and clustering is important technique used in video motion detection and surveillance.

1. In this paper authors proposed clustering algorithm is capable of both clustering DTs and learning novel DT cluster centers that are representative of the cluster members in a manner that is consistent with the underlying generative probabilistic model of the DT. We also derive an efficient recursive algorithm for sensitivity analysis of the discrete-time Kalman smoothing filter, which is used as the basis for computing expectations in the E-step of the HEM algorithm. Finally, we demonstrate the efficacy of the clustering algorithm on several applications in motion analysis, including hierarchical motion clustering, semantic motion annotation, and learning bag-of-systems (BoS) codebooks for dynamic texture recognition. we address the problem of clustering dynamic texture models, i.e., clustering linear dynamical systems. Given a set of DTs (e.g., each learned from a small video cube extracted from a large set of videos), the goal is to group similar DTs into K clusters, while also learning a representative DT “center” that can sufficiently summarize each group. This is analogous to standard K-means clustering, except that the data points are dynamic textures instead of real vectors.

2. In this paper authors proposed a novel method for object discovery and dense modelling in RGB-D image sequences using motion cues. We develop our method as a building block for active object perception, such that robots can learn about the environment through perceiving the effects of actions. Our approach simultaneously segments rigid-body motion within key views, and discovers objects and hierarchical relations between object parts. The poses of the key views are optimized in a graph of spatial relations to recover the rigid-body motion trajectories of the camera with respect to the objects. In experiments, we demonstrate that our approach finds moving objects, aligns partial views on the objects, and retrieves hierarchical relations between the objects.

3. In this paper authors proposed a content-based automatic tagging system for music that relies on a high-level, concise “Bag of Systems”(BoS) representation of the characteristics of a musical piece. The BoS representation leverages a rich dictionary of musical code words, where each code word is a generative model that captures tumbrel and temporal characteristics of music. Songs are represented as a BoS histogram over code words, which allows for the use of traditional algorithms for text document retrieval to perform auto-tagging. Compared to estimating a single generative model to directly capture the musical characteristics of songs associated with a tag, the BoS approach offers the flexibility to combine different generative models at various time resolutions through the selection of the BoS code words. Additionally, decoupling the modelling of audio characteristics from the modelling of tag-specific patterns makes BoS a more robust and rich representation of music.

4. In this paper authors present a content-based auto-tagger that leverages a rich dictionary of musical code words, where each code word is a generative model that captures timbrale and temporal characteristics of music. This leads to a higher-level, concise “Bag of Systems” (BoS) representation of the characteristics of a musical piece. Once songs are represented as a BoS histogram over code words, traditional algorithms for text document retrieval can be leveraged for music auto tagging. Compared to estimating a single generative model to directly capture the musical characteristics of songs associated with a tag, the BoS approach offers the flexibility to combine different classes of generative models at various time resolutions through the selection of the BoS code words. Experiments show that this enriches the audio representation and leads to superior auto-tagging performance.

5. In this paper authors gives statistical model for an ensemble of video sequences that is sampled from a finite collection of visual processes, each of which is a dynamic texture. An expectation maximization (EM) algorithm is derived for learning the parameters of the model, and the model is related to previous works in linear systems, machine learning, time-series clustering, control theory, and computer vision. Through experimentation, it is shown that the mixture of dynamic textures is a suitable representation for both the appearance and dynamics of a

variety of visual processes that have traditionally been challenging for computer vision (e.g. fire, steam, water, vehicle and pedestrian traffic, etc.). When compared with state-of-the-art methods in motion segmentation, including both temporal texture methods and traditional representations (e.g. optical flow or other localized motion representations), the mixture of dynamic textures achieves superior performance in the problems of clustering and segmenting video of such processes

6. In this paper authors describe a novel approach to automatic music annotation and retrieval that captures temporal (e.g., rhythmical) aspects as well as timbral content. The proposed approach leverages a recently proposed song model that is based on a generative time series model of the musical content the dynamic texture mixture (DTM) model that treats fragments of audio as the output of a linear dynamical system. To model characteristic temporal dynamics and timbales content at the tag level, a novel, efficient hierarchical EM algorithm for DTM (HEM-DTM) is used to summarize the common information shared by DTMs modelling individual songs associated with a tag. Experiments show learning the semantics of music benefits from modelling temporal dynamics.

7. In this paper author handle the challenging problem of recognizing dynamic video contents from low-level motion features. We adopt a statistical approach involving modeling, (supervised) learning, and classification issues. Because of the diversity of video content (even for a given class of events), we have to design appropriate models of visual motion and learn them from videos. We have defined original parsimonious global probabilistic motion models, both for the dominant image motion (assumed to be due to the camera motion) and the residual image motion (related to scene motion). Motion measurements include affine motion models to capture the camera motion and low-level local motion features to account for scene motion. Motion learning and recognition are solved using maximum likelihood criteria. To validate the interest of the proposed motion modelling and recognition framework, we report dynamic content recognition results on sports videos.

8. Authors proposed a mixture of dynamic textures, which models a collection of videos consisting of different visual processes as samples from a set of dynamic textures. We derive the EM algorithm for learning a mixture of dynamic textures, and relate the learning algorithm and the dynamic texture mixture model to previous works. Finally, we demonstrate the applicability of the proposed model to problems that have traditionally been challenging for computer vision. the expectation maximization (EM) algorithm is derived for maximum likelihood estimation of the parameters of a mixture of dynamic textures. Second, the relationships between this mixture model and various other models previously proposed in the machine learning and computer vision literatures, including mixtures of factor analyzers, linear dynamical systems, and switched linear dynamic models, are analyzed. Finally, we demonstrate the applicability of the new model to the solution of traditionally difficult vision problems that range from clustering traffic video sequences to segmentation of sequences containing multiple dynamic textures.

9. The proposed solution approximates the nonlinear manifold and dynamics using piecewise linear models. The interactions among the linear models are captured in a graphical model. By exploiting the model structure, efficient inference and learning algorithms are obtained without oversimplifying the model of the underlying dynamical process. Evaluation of the proposed framework with competing approaches is conducted in three sets of experiments: dimensionality reduction and reconstruction using synthetic time series, video synthesis using a dynamic texture database, and human motion synthesis, classification and tracking on a benchmark data set.

10. Authors describe a novel approach to automatic music annotation and retrieval that captures temporal (e.g., rhythmical) aspects as well as tumbrel content. The proposed approach leverages a recently proposed song model that is based on a generative time series model of the musical content the dynamic texture mixture (DTM) model that treats fragments of audio as the output of a linear dynamical system. To model characteristic temporal dynamics and timbale content at the tag level, a novel, efficient, and hierarchical expectation–maximization (EM) algorithm for DTM (HEM-DTM) is used to summarize the common information shared by DTMs modelling individual songs associated with a tag. Experiments show learning the semantics of music benefits from modelling temporal dynamics.

3. VIDEO BACKGROUND

The background of video plays an important role in video segmentation and object tracking. The automatic background updating of video increase the efficiency of video object tracking and reduces the frame loss of video. Various researchers proposed a background updating algorithm for video tracking some are perform better performance for video tracking. A general background subtraction algorithm applies a Kalmar filter (or α-blending) to the pixel intensities to find the background.

$$B_{t+1} = B_t + (\alpha_1(1 - M_t) + \alpha_2 M_t)D_t, \ldots \qquad (1)$$

Frame and$M_t$is a binary moving object hypothesis mask.Where represents the background model at time t, is the difference between the present such an approach works well when foreground objects appear infrequently, but when the background is occluded by an object for a significant time, the algorithm begins to fail. Another problem is that Mt is usually generated from by there holding and applying morphological operators. Such self-feedback can make the filtering unstable. For ex-ample, a single detection failure or a sudden illumination change can result in a permanent failure (or a ghost) which may even grow in size until it covers up the entire image. Sudden illumination changes commonly occur in many field video images because most video cameras have an auto-iris feature. Various augmentations have been applied to the back-ground subtraction, for example, to use temporal median instead of the α-blending [4]. More recently, Batista et al. introduced various augmentations including the use of multi-layer background models and dynamic thresholding [2]. Such augmentations significantly improve the robust-features) to generate more robustM_t. In addition, we also made the following modifications to Equation 1: less, but the problem of self-feedback is still there. Therefore, we incorporate an external cue (corner)

- The temporal median approach is combined with the α-blending;
- An illumination correction procedure is added to deal with sudden/temporary illumination changes; and

  We use an update equation

$$B_{t+1} = \begin{cases} I_c(B_t) \quad M_t = 1 \\ I_c((1-\alpha)B_t + \alpha N_t \; M_t = 0, \end{cases} \qquad (2)$$

Where I_c ()is an illumination-correction function and *Nt* is the temporal median of the recent, say 15, frames. Note that our background update rate is about 2 frames per second and the 15 frames spans about 7 to 8 seconds. The illumination-correction is applied to each of the RGB value since the auto-iris can also change the color distribution (hue):

$$I_c = (k_R R, k_G G, k_b B), \qquad (3)$$

Where,

K_R ,K_G&K_B are determined by voting on R_C/R, G_C/G, andB_C/B over all the pixels in the images. ($R_{C,}G_{C,}B_C$) are the pixel values of the current frame. For M_t we start with the standard procedure which is to 1) threshold the difference, 2) apply morphological operators (or threshold after over-smoothing), and 3) perform connected component analysis to fill holes, remove small regions, and find object blobs. After the object blobs are found, we apply an additional validation step to remove the ghosts. We assume that within all the non-ghost foreground region there exists at least one valid corner, i.e., a corner feature which is not found from the background image. For more details on the valid corner, The illumination challenge caused by an auto-iris camera. The two white vehicles in the bottom changes the entire scene darker and it causes significant false alarms. However, the error is minimized by applying the illumination correction. Here also discuss another background updating model. Each pixel in the scene is modeled by a mixture of k Gaussian distributions. The probability that a certain pixel has a value of X_N at time N can be written as,

$$p(X_N) = \sum_{j=1}^{k} w_j \eta(X_N; \theta_j) \ldots\ldots\ldots\ldots\ldots\ldots \qquad 4)$$

Where W_K is the weight parameter of the k^{th} Gaussian component $\eta(X;\theta_K)$is the Normal distribution ofK^{th} component represented by

$$\eta(X;\theta_k) = \eta(X;\theta_k,\Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_k|^{\frac{1}{2}}} e^{\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k)} \ldots \qquad (5)$$

Where μ_k is the mean and $\Sigma_k = \sigma^2{}_k I$ is the covariance of the k^{th} component [8].

The K distributions are ordered based on the fitness value $^{w_k}/_{\sigma_k}$and the first B distributions are used as a model of the background of the scene where B is estimated as

$$B = arg_b min\left(\sum_{j=1}^{b} w_j \, T\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \, (6)$$

The threshold T is the minimum fraction of the background model. In other words, it is the minimum prior probability that the background is in the scene. Background subtraction is performed by marking a foreground pixel any pixel that is more than 2.5 standard deviations away from any of the B distributions. The first Gaussian component that matches the test value will be updated by the following update equations,

$$\hat{w}_k^{N+1} = (1-\alpha)\hat{w}_k^{N} + \alpha\hat{w}\text{ÊÞ}\varphi\text{Ṕ} \ldots\ldots\ldots\ldots\ldots \qquad (7)$$

Where ω_k is the k^{th} Gaussian component,1/α defines the time constant which determines change. If none of the K distributions match that pixel value, the least probable component is replaced by a distribution with the current value as its mean, an initially high variance, and a low weight parameter. According to their papers [1, 2, 3], only two parameters, α&T, needed to be set for the system. The details of its robustness were explained in their papers [1, 2, 3]; however, with a simple discussion, we can see its incapability. Firstly, if the first value of a given pixel is a foreground object, there is only one Gaussian where its weight equals unity. With only one-color subsequent background values, it will take frames until the genuine background can be considered as a background and $\log_{(1-\alpha)}$(T) frames until it will be the dominant background component. For example, if we assume that at least 60% of the time the background is present and α is 0.002 (500 recent frames), it would take 255 frames and 346 frames for the component to be included as part of the background and the dominant background component, respectively. The situation can be worse in busy environments where a clean ck ground is rare. This paper presents a solution to the problem in the next section. Secondly, ρ is too small due to the likelihood factor. This leads to too slow adaptations in the means and the covariance matrices, therefore the tracker can fail within a few seconds after initialization. One solution to this is to simply cut out the likelihood term from ρ.

4. VIDEO SEGMENTATION

Filter is important tools in video tracking for estimation of frame loss and reduction of AWGN noise. Various filters are used for video processing such as Gaussian filter, kamala filter and partial least square filter. The working mode of filter in video is spatial and temporal. In this section we discuss we discuss some filter used in video tracking. Instead of extracting complex features from a connected component, the raw shape of a connected component itself is an important distinguishable feature for classifying structured video and random or irregular components [4]. Together with the shape of connected component, the surrounding area of a connected component can also play an important role for video and background classification, similarly because of the structured video and non-structured non-video surrounding areas. Neighborhood surrounding areas for video and non-video regions. We refer connected component with its neighborhood surrounding as confider. Based on the above mentioned hypothesis, our feature vector of connected component is composed of shape and con video information [8]. Detail description of the feature vector is presented below. In order to improve the segmentation results, a nearest neighbor analysis by using class probabilities is performed for rending the class label of each connected component. For this purpose, a region of 70 _ 70 (empirically chosen) is selected from document image by keeping targeted connected component at centre. The probabilities of connected components within the selected regions are already computed during classification [14].

5. CONCLUSION AND FUTURE WORK

In this paper we study of video modelling based on clustering and segmentation process. Background segmentation play important role in object tracking system. The correct background updating function improved the performance of video tracking algorithm. Video background segmentation implies by different approach such as background subtraction, particle least technique and many more. Such method creates a difference between actual video and background of motion video. In the process of study we also found that noise filter process for video object tracking system, now a day various filter are used such as particle filter, Gaussian filter and kamala filter. The filtration mechanism of video improves the performance of video tracking. In future we modified the partial least square filter for processing of video object tracking.

REFERENCES

[1] Adeel Mumtaz, Emanuele Coviello, Gert R.G. Lanckriet, Antoni B. Chan “Clustering Dynamic Textures with the Hierarchical EM Algorithm for Modelling Video” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35. , 2013. Pp 1606-1621.

[2] Jorg Stuckler,Sven Behnke “Hierarchical Object Discovery and Dense Modelling From Motion Cues in RGB-D Video” In Processing of 23rd IJCAI , China 2013. Pp 1-8.

[3] Katherine Ellis, Emanuele Coviello, Antoni B. Chan, Gert Lanckriet “A Bag of Systems Representation for Music Auto-Tagging” IEEE Trans.On Audio, Speech, and Language Processing, vol. 21, 2013. Pp 2454-2569.

[4] Katherine Ellis, Emanuele Coviello, Gert R.G. Lanckriet “Semantic Annotation And Retrieval Of Music Using A Bag Of Systems Representation” 2011. Pp 1-6.

[5] Antoni and Segmenting Video with Mixtures of Dynamic Textures” IEEE Trans. Pattern Analysis and Machine Intelligence, 2007B. Chan, Nuno Vasconcelos “Modeling, Clustering,. Pp 1-18.

[6] Emanuele Coviello, Luke Barrington, Antoni B. Chan, Gert. R. G. Lanckriet ” Automatic Music Tagging Wih Time Series Model” 2010. Pp 1-6.

[7] Gwenaelle Piriou, Patrick Bouthemy, Jian-Feng Yao “Recognition of Dynamic Video Contents With Global Probabilistic Models of Visual Motion” IEEE Trans. On Image Processing, vol.15, 2006. Pp 3418-3431.

[8] Antoni B. Chan, Nuno Vasconcelos “Mixtures of Dynamic Textures”IEEE International Conference on Computer Vision, Beijing, 2005. Pp 1-7.

[9] Rui Li, Tai-Peng Tian, Stan Sclaroff “Simultaneous Learning of Nonlinear Manifold and Dynamical Models for High-dimensional Time Series” in Proc. International Conference on Computer Vision, Oct. 2007. Pp 1-9.

[10] Emanuele Coviello, Antoni B. Chan, Gert Lanckriet “Time Series Models for Semantic Music Annotation” IEEE Trans.On Audio, Speech, and Language Processing, vol.. 19, 2011. Pp 1343-1360.

[11] Antoni B. Chan, Emanuele Coviello, Gert. R. G. Lanckriet “Clustering Dynamic Textures with the Hierarchical EM Algorithm” 2010. Pp 1-8.

[12] A.Ravichandran,R.Chaudhry, and R. Vidal, “Categorizing Dynamic Textures Using a Bag of Dynamical Systems,” IEEE Trans.Pattern Analysis and Machine Intelligence, vol. 35, no. 2, 2013, Pp 342-353.

[13] R. Pe´teri, S. Fazekas, and M.J. Huiskes, “DynTex: A Comprehensive Database of Dynamic Textures” Pattern Recognition Letters, vol. 31, no. 12, 2010, Pp 1627-1632,

[14] K.G.Derpanis and R.P.Wildes,“Dynamic Texture Recognition Based on Distributions of Spacetime Oriented Structure” Proc.IEEE Conf.Computer Vision and Pattern Recognition, 2010. Pp 1-6.

[15] H.Cetingul and R.Vidal,“Intrinsic Mean Shift for Clustering on Stiefel and Grassmann Manifolds” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009.Pp