Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 9, Pages 40-45, October 2017

A Review on Content Based Video Retrieval, Classification and Summarization

Mr.S.Gnana Saravanan[1], Mr.T.Sivaprakasam[2] and Dr.D.Somasundaram[3]

[1]*Assistant Professor, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India. Email: gnanasaravanan@siet.ac.in*
[2]*Assistant Professor, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India. Email: tsivaprakasam@siet.ac.in*
[3]*Associate Professor, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamilnadu, India Email: somgce@gmail.com*

Article Received: 03 August 2017 *Article Accepted: 28 September 2017* *Article Published: 10 October 2017*

ABSTRACT

This paper provides an overview of different video content modeling techniques, retrieval techniques and also the classification techniques employed in existing content-based video indexing and retrieval (CBVR) systems. Based on the requirements for modeling of a CBVIR system, this paper analyzes and categorizes the existing modeling approaches. Starting with a review of video content modeling and representation techniques, this paper presents view-invariant representation approaches and the corresponding performance analysis too. Based on the current status of research in CBVR systems, we have identified the video retrieval approaches from spatial and temporal analysis. Finally, a summary of future trends and open problems of content-based video modeling retrieval and classification is also provided.

Keywords: Video content modeling technique, Classification technique and Video retrieval approach.

1. CONTENT MODELING AND REPRESENTATION TECHNIQUES IN VIDEO SEARCHING

In this section, this paper presents a generalization of video content modelling and representation. The general problems in video content modelling and representation is first investigated. With regard to that, the state-of-the-art approaches: curvature scale space (CSS) and centroid distance functions (CDF)-based representations is presented. Subsequently the null space invariant (NSI) representations for video classification and retrieval due to camera motions is also presented. Finally, brief overview of the other approaches in video content modelling and representation and future trends is also presented.

1.1 Need for CBVR Systems

Content-based video collections are growing rapidly in both the professional and consumer environment, these are characterized by a steadily increasing capacity and content variety. Since searching them manually through the available collections is tedious and time-consuming, the classification and retrieval tasks to the automated systems become crucial for stored video volumes. The development of such systems is based on the algorithms for video content analysis. These algorithms are built around the models, bridging the gap between unifying query-by-example based indexing and retrieval systems and high-level semantic query-based activity recognition.

1.2 General Problems in modelling and Representation techniques

Within the last few years, object motion trajectory-based recognition has gained significant interest in diverse application areas including, Global Positioning System (GPS), sign language gesture recognition, animal mobility experiments, Car Navigation System (CNS), automatic video surveillance and sports video trajectory analysis. Psychological studies show that human beings can easily discriminate and recognize an object's motion pattern even from large viewing distances or poor visibility conditions where other features of the object vanish.

The development of such accurate activity classification and recognition algorithms in multiple view situations is still an extremely challenging task which is needed to be addressed. Object trajectories captured from different view-points lead to completely different representations, which can be modeled by affine transformation approximately. To get a view of independent representation, the trajectory data is represented in an affine invariant feature space. With regard to that, a compact, robust view invariant representation is highly desirable.

1.3 View Invariant Representation method

View-Invariant representation is a very important and sometimes difficult aspect of an intelligent system. The representation is an abstraction of the sensory data, which should reflect a real world situation. The view invariant representation includes scale view invariant , affine view invariant, and projective view invariant, etc. Once the representation has been defined, various recognition or classification algorithms can be performed. The methods usually involve some kind of distance calculation between a model and an unknown input. The model with smallest distance is taken to be the class of motion to which the input belongs to. The problem with this is that the system can only recognize a predefined set of behaviors [1] [2] [3].

1.3.1 Centroid Distance Functions (CDF) and Curvature Scale Space (CSS) - based Representation Technique

The scale-space is a multi-resolution technique used to represent data of arbitrary dimension without any knowledge of noise level and preferred scale of smoothness. The notion of scale in the measured data is handled by representing a measured signal at multiple levels of detail for example,from the finest to the coarsest (i.e. from original signal to the most-smoothed version). The CSS representation takes

curvature data of a parametric curve and represents it by its different evolved versions at increasing levels of smoothness.

1.3.2 Null Space View Invariance Representation Technique (NSI)

This subsection, introduces a simple but highly efficient view invariant representation based on Null Space Invariant (NSI) matrix. This is the first use of Null space in motion-based classification/retrieval applications. Indexing and classification of the NSI operator is obtained by extracting the features of the null space representation using PCNSA (Principal Components Null Space Analysis), which provides an efficient analysis tool when different classes may have different non-white noise covariance matrices. Dimensionality reduction for indexing of the NSI is achieved by first performing Principal Components Analysis (PCA) as part of PCNSA. Classification is performed in PCNSA by determining the *ith* class M_i-dimensional subspace by choosing the M_i eigenvectors that give the smallest intra-class variance. The M_i-dimensional space is referred to as the Approximate Null Space (ANS). A query is classified into the class if its distance to the class mean in ANS space is lowest among all the other classes. A fundamental set of 2-D affine invariants for an ordered set of n points in R^2 (not all collinear) is expressed as an n-3 dimensional subspace, H^{n-3}, of R^{n-1}, which yields a point in the 2n-6 dimensional Grassmannian Gr_R(n-3,n-1), which also shows number of invariants is 2n-6 in 2-D.

In Null Space Invariant (NSI) of a trajectories matrix (each row in the matrix corresponds to the positions of a single object over time) and this is introduced as a new and powerful affine invariant space to be used for trajectory representation. This invariant, which is a linear subspace of a particular vector space, is the most natural invariant and is definitely more general and more robust than the familiar numerical invariants. It does not need any assumptions and after invariant calculations it conserves all the information of original raw data.

1.3.3 Tensor Null Space Invariance Representation Technique (TNSI)

Among affine view-invariance systems, majority of them represent affine view invariance in a single dimension, limiting the system to only single dimension affine view-invariance and single object motion based queries. In many of the applications, it is not only the individual movement of an object that is of interest, but also the motion patterns that emerge while considering synchronized or overlapped movements of multiple objects. For example, in sports video analysis, one is often interested in a group activity involving activity of multiple players, rather than the activity of an individual player. Also, due to camera movement, same motion trajectory has completely a unique representation from different viewing angles. Hence, a highly efficient classification and retrieval system which is invariant to multidimensional affine transformations is desirable.

1.4 Other Representation techniques

In other literature [10], the author Zhang et al uses a multiple different criteria like zoom-in type of effects in a shot content and color content changes. In [9] the author presents a technique for shot content representation and similarity measure using sub-shot extraction and representation. The author uses two content descriptors, Spatial Structure Histogram (SSH) and Dominant Color Histogram (DCH) to measure content variation to represent sub-shots. The author C. Faloutsos [9] represent a shot using a tree structure called shot tree, formed by clustering frames in a shot. This approach addresses the problem of scene content representation for both similarity matching and browsing where for browsing only the root node of the tree (key frame) is used, while for similarity matching two or three levels of tree can be used employing the standard tree matching algorithms. In addition the authors in [11] represent other views of invariant representations such as geometry invariants (GI).

1.5 Open Problems and Future Trends

From the above reviews, it can be seen that there are remarkable advances in the field of video content modelling and representation techniques. However, to make the full use out of visual information retrieval systems, there are many open research issues that are still needed to be addressed. In the sections below, several open problems and a brief summary of future trends are listed.

1) Optimal Sampling Strategy for High Dimensional Representation [7]: This paper proposes the optimal sampling scheme for 2D null space representation. In future, the work can be considered the optimal sampling strategy for 3D Null space representations.

2) Segmentation of Tensor Null Space Invariants: The authors in [12] represents that The dimensionality of feature vectors employed in most systems for representing visual media can be solved by tensor null space invariants (TNSI), but, the computational burden for TNSI will be very heavy. To segment the space properly and reduce the dimension of the representation, there is a need in practical view invariant systems.

2. VIDEO INDEXING AND RETRIEVAL TECHNIQUES

This section, presents the concepts, problems and state-of-the-art approaches for content-based video retrieval (CBVR). The general problems of content-based retrieval is overviewed. Following to that, one of the key problems in video retrieval-video summarization is presented, which summarizes the state of-the-art approaches for both key frame extractions and shot-boundary detection. Spatial and Temporal Analysis is very important to access video content, hence the focus is shown on spatial-temporal motion trajectory analysis, presenting both single and multiple motion trajectory-based CBVR techniques, specifically, (i) unfolded multiple-trajectory indexing and retrieval (UMIR) (ii) geometrical multiple trajectory indexing and retrieval (GMIR) algorithm, and (iii) concentrated multiple-trajectory indexing and retrieval (CMIR) algorithm. The above algorithms doesn't only reduces the dimensionality of the indexing space but also enables the realization of fast retrieval systems. Finally, a brief overview of other

approaches in video retrieval and future trends in this research area are also suggested.

2.1 Definitions of CBVR systems

By definition, a Content-Based Video Retrieval (CBVR) system aims at assisting a human operator to retrieve sequence (target) within a potentially large database [11]. The authors in [11] has just presented a natural extension of the well-known Content-Based Image Indexing and Retrieval (CBIR) systems. Both systems are aiming at accessing image and video by its content, namely, the spatial (image) and spatial-temporal (video) information. A Typical spatial information includes texture, color, edge, etc., while a typical temporal information includes change of scenes and motions. Moving from images to video adds several orders of complexity to the retrieval problem due to indexing, analysis and browsing over the inherently temporal aspect of video [15].

2.2 General Problems in Video Indexing and Retrieval

One of the key issues in CBVR is, as pointed out by authors in [15], is bridging the ”semantic gap”, which refers to the gap between low level features (such as texture, color, shape, layout, structure and motion) and high level semantic meanings of content (such as people, or car-rasing scenes , indoor and outdoor). Low level features such as textures and colors are easy to measure and compute, however, it is a great challenge to connect the low level features to a semantic meaning, especially involving intellectual and emotional aspects of the human operator. Another issue is how to efficiently access the rich content of video information, these involves video content summarization, and spatial and temporal analysis of videos, which is discussed in detail in the following sections.

2.3 Video Summarization technique

Video summarization is the process of creating a presentation of visual information about the structure of video, which should be shorter than the original video [14]. Typically, a subset of video data such as key frames or highlights such as entries for shots, scenes, or stories is extracted for compact representation and fast browsing of video content. A *shot* is a set of contiguous frames acquired through a continuous camera recording, and is considered as the fundamental building block of a video. Several shots consist a *scene*, which is a set of contiguous shots that have a common semantic significance. Altogether, a video usually consists of several scenes. The significance of shots in videos, shot boundary detection or extraction of shots is the key towards video summarization.

2.3.1 Shot boundary Detection

Shot boundary detection, or video segmentation, is to identify the frame or frames of videos where a transition takes place from one shot to another [17]. The locations of these changes are referred to as a Break or a Cut. Examples of transitions include cuts, dissolves, fades, wipes and other special effect edits. Some more efficient solutions have been proposed for shot boundary detection in [17] [18] [19].

2.3.2 Key Frame Extraction

After video segmentation, key frames are extracted from each shot as features for compact summarization of video. The Key frame extraction must be automatic and content-based, such that important video shot content are retained while redundancies are removed after extraction. There are several key frame extraction criteria have been proposed, such as heuristic decisions [20], to use intra frame or first frame of video shot as key frame; or visual features [21], such as brightness, color or dominant colors; or motion patterns [22][23], such as dominant motion components of camera or large moving objects. These key frames are further utilized to summarize video content for efficient indexing and retrieval.

2.4. Open Problems and Future Trends Video Indexing and Retrieval

From the above review, one can notice the potential and promising future of video indexing and retrieval systems, however, there are still many open research issues that need to be addressed to make the full use of visual information retrieval systems. In the following sections, several open problems are identified and also a brief summary of future trends is provided.

1) Low to High Level Semantic Gap: Current research efforts are more inclined towards high-level description and retrieval of visual content. Most of the techniques are at high level of abstraction which assumes the availability of high-level representation and processes the information for indexing. The techniques that bridge this semantic gap between pixels and predicates are a field of growing interest. Intelligent systems are needed to take low-level feature representation of the visual media to provide a model for the high-level object representation of the content.

2) Dynamic matching, updating of query and databases: The practical utility of a robust CBVR system must address the problem of dynamic updating of video databases and feature spaces, as well as dynamic matching of queries and databases [39].

3. VIDEO CLASSIFICATION AND RECOGNITION TECHNIQUES

Video classification differs from video indexing and retrieval, since in video classification, all videos are put into various categories, and each video is assigned a meaningful label. Recently many automatic video classification algorithms have been proposed, most of them can be categorized into any one of following four groups: audio-based approaches, text-based approaches, visual-based approaches, and combination of text, audio and visual features. Many standard classifiers, such as Bayesian, support vector machines (SVM), Gaussian Mixture Models (GMM), hidden Markov Models (HMMs) and neural networks have been applied in video classification and recognition. For more details one can refer [40] which proposes a novel distributed multi-dimensional hidden Markov Model (DHMM) for modelling of interacting trajectories involving multiple objects. This model is capable

of conveying not only dynamics of each trajectory, but also interactions information between multiple trajectories.

3.1 Model-based Classification Technique

3.1.1 Multi-dimensional Distributed Hidden Markov Models

This paper introduces some basic idea for a novel distributed multi-dimensional hidden Markov Model (DHMM). In this model, each object-trajectory is modelled as a separate Hidden Markov process; while "interactions" between objects are modelled as dependencies of state variables of one process on states of the others. The interesting fact is that, HMM is very powerful tool to model temporal dynamics of each process (or trajectory); each process (or Trajectory) has its own dynamics, while it may be influenced or influence others. Figure 1 demonstrates examples of multiple interactive motion trajectories.

(a) (b)

Fig. 1 Examples of multiple interactive trajectories: (a) "Two people walk and meet". (b) "Two air planes fly towards each other and pass by".

The training and classification algorithms presented in [45] to a general causal model. Also a new conditional independent subset-state sequence structure decomposition of state sequences is proposed for the 2D Viterbi algorithm. The new model can be applied to many problems in pattern analysis and classification.

3.1.2 Hidden Markov Model

Hidden Markov model (HMM) is very powerful tool to model temporal dynamics of processes, and has been successfully applied to many applications such as gesture recognition [41], speech recognition [42], musical score following [43]. The authors of [44] has presented a novel classification algorithm of object motion trajectory based on 1D HMM. They have segmented single trajectory into atomic segments called sub-trajectories, based on curvature of trajectory, then the sub-trajectories are represented by their PCA (principal component analysis) coefficients. Temporal relationships of sub-trajectories are represented by fitting a 1D HMM. However, all the above applications rely only on a one-dimensional HMM structure. Simple combinations of 1D HMMs cannot be used to characterize multiple trajectories, since 1D models fail to convey interaction information of multiple interacting objects. The major challenge is to develop a new model that will semantically reserve and convey the "interaction" information.

3.2 Open Problems and Future Trends in Model-based Classification Techniques

Many works have been done in video classification and recognition, most of them use audio, Text or visual features or combination of features. Despite many methods are proposed, there are still many open problems that need to be addressed in video classification and recognition.

1) Efficient Fusion of Various Features: It is a method which combines more features such as text, audio and visual feature would improve the performance of video classification systems. However, very few works has been done in efficient fusion of different features.

2) Although many HMM models have been applied in the field of video classification, A rich spatial-temporal structure of video has not yet been fully explored. A robust model that captures spatial-temporal structure of video and utilizes various features of videos is highly desired.

4. CONCLUSION

This paper gives a review of all methodologies being used in Video retrieval methods based on contents and also about video classification techniques. Along with it this paper also adds up the challenges available in every technique or method which the reader can take it as their research area trying to figure out the solution for the problems or challenges.

REFERENCES

[1] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Object trajectory-based activity classification and recognition using hidden Markov models", IEEE Transactions on Image Processing, vol. 16, pp. 1912-1919, 2007.

[2] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Real-time motion trajectory-based indexing and retrieval of video sequences", IEEE Transactions on Multimedia, vol. 9, pp. 58-65, 2007.

[3] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "A Hybrid System for Affine-Invariant Trajectory Retrieval", Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, New York, New York, 2004.

[4] F. I. Bashir, A. A. Khokhar, and D. Schonfeld, "Real-time affine-invariant motion trajectory based retrieval and classification of video sequences from arbitrary camera view," ACM Multimedia Systems Journal, Special Issue on Machine Learning Approaches to Multimedia Information Retrieval, vol. 12, pp. 45-54, 2006.

[5] N. Vaswani, R. Chellappa, "Principal Components Null Space Analysis for Image and Video Classification," IEEE Trans. Image Processing, July 2006.

[6] The University of California at Irvine Knowledge Discovery in Databases (KDD) archive, [Online]. Available: http://kdd.ics.uci.edu, URL.

[7] X. Chen, D. Schonfeld and A. Khokhar, "Robust null space representation and sampling for view invariant motion trajectory analysis", IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[8] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and Effective Querying by Image Content", Journal of Intelligent Information Systems, 3, 3/4, July 1994, pp. 231-262.

[9] T. Lin, H. J. Zhang, and Q. Y. Shi, "Video Content Representation for Shot Retrieval and Scene Extraction", International Journal of Image and Graphics, Vol. 1, No. 3, July 2001.

[10] H. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing", Pattern Recognition, Vol. 30, no.4, pp. 643-658, 1997.

[11] [Online] Stephane Marchand-Maillet, "Content-based Video Retrieval: An overview", http://viper.unige.ch/ marchand/CBVR/, URL

[12] X. Chen, D. Schonfeld and A. Khokhar, "View-Invariant Tensor Null-Space Representation for Multiple Motion Trajectory Retrieval and Classification", IEEE International Conference on Acoustics, Speeach, and Signal Processing, 2009.

[13] T. Ng, S. Chang and M. Tsui, "Using Geometry Invariants for Camera Response Function Estimation", IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[14] N. Sebe, M. S. Lew and A. W. M. Smeulders, "Video retrieval and summarization", editorial introduction, Computer Vision and Image Understanding (CVIU),

[15] N. Sebe, M. S. Lew, X. Zhou, T. S. Huang and E. M. Bakker, "The State of the Art in Image and Video Retrieval", in proceedings of the International Conference on Image and Video Retrieval (CIVR), pages 1-8, 2003.

[16] J. Yuan, H. Wang, W. Zheng, J. Li, F. Lin and B. Zhang, "A Formal Study of Shot Boundary Detection", IEEE Transactions on Circuit and Systems for Video Technology, pp. 168-186, 2007..

[17] Y. Yang and L. Ming, "A Survey on Content based Video Retrieval", Hong Kong University of Science and Technology.

[18] A. Hanjalic, "Shot-boundary detection: unraveled or resolved?", IEEE Transactions on Circuit and Systems for Video Technoligy, vol. 12, issue 2, pp. 90-105, 2002.

[19] R. Lienhart, "Reliable Transition Detection in Videos: A Survey and Practitioner's Guide", International Journal of Image and Graphics, vol. 1, pp. 469-486, 2001.

[20] M. M. Yeung and B. Liu, "Efficient Matching and Clustering of Video Shots", technical report, Princeton University, 1995.

[21] H. J. Zhang, S. W. Smoliar and J. H. Wu, "Content-based Video Browsing Tools", SPIE Conference on Multimedia Computing and Networking, San Jose, CA, 1995.

[22] W. Wolf, "Key Frame Selection by Motion Analysis", IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1228-1231, 1996.

[23] X. Liu, C. B. Owen and F. Makedon, "Automatic Video Pause Detection Filter", Technical Report PCS-TR97-307, Dartmouth College, Computer Science, Hanover, NH, 1997.

[24] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram and D. Zhong, "A Fully Automated ContentBased Video Search Engine Supporting Spatiotemporal Queries", IEEE Trans. Circuits Syst. Video Technol., vol.8, no.5, pp. 602-615, 1998.

[25] N. AbouGhazaleh, Y. E. Gamal, "Compressed Video Indexing Based on Object's Motion", Int. Conf. Visual Communication and Image Processing, VCIP'00, Perth, Australia, pp. 986993, 2000.

[26] B. Katz, J. Lin, C. Stauffer and E. Grimson, "Answering questions about moving objects in surveillance videos", in proceedings of 2003 AAAI Spring Symp. New Directions in Question Answering, Palo Alto, CA, 2003.

[27] N. Rea, R. Dahyot and A. Kokaram, "Semantic Event Detection in Sports Through Motion Understanding", in proc. 3rd Int. Conf. on Image and Video Retrieval (CIVR), pp. 21-23, 2004.

[28] E. Sahouria, A. Zakhor, "A Trajectory Based Video Indexing System For Street Surveillance", IEEE Int. Conf. on Image Processing (ICIP), pp. 24-28, 1999.

[29] W. Chen and S. F. Chang, "Motion Trajectory Matching of Video Objects", IS&T/ SPIE, pp. 544-553, 2000.

[30] F. I. Bashir, A. A. Khokhar and D. Schonfeld, "Segmented trajectory based indexing and retrieval of video data", in proc. IEEE Int. Conf. Image Processing, pp.623-626, 2003.

[31] T. Zhao and R. Nevatia, "Tracking multiple humans in crowded environment", in proc. IEEE Int. Conf. Compt. Vision and Pattern Recognit., vol. 2 (2004), pp. 406-413, 2004.

[32] C. Chang, R. Ansari and A. Khokhar, "Multiple Object Tracking with Kernal Particle Filter", in proc. IEEE Int. Conf. Compt. Vision and Pattern Recognit., vol. 1 (2005), pp. 566-573, 2005.

[33] X. Ma, F. Bashir, A. Knokhar and D. Schonfeld, "Event Analysis Based on Multiple Interactive Motion Trajectories", IEEE Trans. on Circuits and Syst. for Video Technology, vol 19, no 3, 2009.

[34] X. Ma, F. Bashir, A. Knokhar and D. Schonfeld, "Tensor-based Multiple Object Trajectory Indexing and Retrieval", in proc. IEEE Int. Conf. on Multimedia and Expo.(ICME), toronto, Canada, pp. 341-344, 2006.

[35] L. Lathauwer, B. D. Moor and J. Vandewalle, "A multilinear singular value decomposition", SIAM Journal on Matrix Analysis and Applicat. (SIMAX), vol. 21, issue 4, pp. 1253-1278, 2000.

[36] L. D. Lathauwer and B. D. Moor, "From Matrix to Tensor: Multilinear Algebra and Signal Processing", in proc. 4th IMA Int. Conf. Mathmatics in Signal Process., pp. 1-15, 1996.

[37] R. A. Harshman, "Foundations of the PARAFAC procedure: Model and Conditions for an "explanatory" multi-mode factor analysis", UCLA Working Papers in Phonetics, pp.1-84, 1970.

[38] The Context Aware Vision using Image-based Active Recognition (CAVIAR) dataset, [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/, URL.

[39] X. Ma, D. Schonfeld, and A. Khokhar, "Dynamic updating and down dating matrix SVD and tensor HOSVD for adaptive indexing and retrieval of motion trajectories," IEEE International conference on acostics, speech and signal Processing, Taipei, Taiwan, 2009

[40] D. Brezeale and D. J. Cook, "Automatic Video Classification: A Survey of the Literature", IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 38, no. 3, 2008.

[41] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition From Video Using Hidden Markov Models", Technical Report, MIT Media Lab, Perceptual Computing Group, vol. 375, 1995.

[42] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition", Proceedings of the IEEE, vol. 77, pp. 257-286, 1989.

[43] C. Raphael, "Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, issue 4, pp. 360-370, 1999.

[44] F. I. Bashir and A. A. Khokhar and D. Schonfeld, "HMM based motion recognition system using segmented pca", IEEE International Conference on Image Processing (ICIP'05), vol. 3, pp. 1288-1291, 2005.

[45] J. Li and A. Najmi and R. M. Gray, "Image classification by a two-dimensional hidden markov model", IEEE Trans. on Signal Processing, vol. 48, pp. 517-533, 2000.

[46] X. Ma, D. Schonfeld and A. Khokhar, "Distributed multidimensional hidden Markov model: theory and application in multiple-object trajectory classification and recognition", SPIE International Conference on Multimedia Content Access: Algorithms and Systems, San Jose, California, 2008.

[47] X. Ma, D. Schonfeld and A. Khokhar, "Image segmentation and classification based on a 2D distributed hidden Markov model", SPIE International Conference on Visual Communications and Image Processing (VCIP 08'), San Jose, California, 2008

[48] X. Ma, D. Schonfeld and A. Khokhar, "Distributed Multi-dimensional Hidden Markov Models for Image and Trajectory-Based Video Classification", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 08'), Las Vegas, Nevada, 2008.

[49] X. Ma, D. Schonfeld and A. Khokhar, "Video Event Classification and Image Segmentation Based on Non-Causal Multi-Dimensional Hidden Markov Models", IEEE Transactions on Image Processing (T-IP), to appear, May 2009.

[50] L. E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains", Ann. Math. Stat., vol. 1, pp. 164-171, 1970.

[51] D. Schonfeld and N. Bouaynaya, "A new method for multidimensional optimization and its application in image and video processing", IEEE Signal Processing Letters, vol. 13, pp. 485488, 2006.