MPEG Video Traffic Modeling and Classification Using Fuzzy C-Means Algorithm with Divergence-Based Kernel

Chung Nguyen Tran and Dong-Chul Park

ICRL, Dept. of Information Engineering, Myong Ji University, Korea
{tnchung, parkd}mju.ac.kr

Abstract. A modeling and classification model for MPEG video traffic data using a Fuzzy C-Means algorithm with a Divergence-based Kernel (FCMDK) for clustering GPDF data is proposed in this paper. The FCMDK is based on the Fuzzy C-Means clustering algorithm and thus exploits advantageous features of fuzzy clustering techniques. To further improve classification accuracies and deal with nonlinear data, the input data is projected into a feature space of a higher dimensionality. Consequently, nonlinear problems existing in the input space can be solved linearly in the feature space. The divergence-based kernel method adopted in the FCMDK employs a divergence measure between two probability distributions for its similarity measure. By adopting the divergence-based kernel method for probability data, the FCMDK can not only utilize advantageous features of the kernel method but also exploit the statistical nature of the input data. Experiments and results on several MPEG video traffic data sets demonstrate that the classification model employing the FCMDK for clustering GPDF data can archive improvements of 28.19% and 34.60% in terms of False Alarm Rate (FAR) over the models using the conventional k-means and SOM algorithms, respectively.

1 Introduction

Content-based retrieval of video data has attracted a great attention in recent years. Many video applications such as video on demand, video databases, and video teleconferencing can benefit from retrieval of the video data based on their content. However, with the rapid increase in various multimedia services, numerous video databases are available through the internet. Organizing these huge video databases into libraries and providing effective indexing require an efficient modeling and classification model.

Recently, various video data classification models have been proposed [1,2,3,4]. Most of classification models are based on pattern recognition approaches which often use a Gaussian Mixture Model (GMM) for modeling video traffic data and a Bayesian classifier [4]. In order to obtain mixture components , also called Gaussian Probability Density Function (GPDF) data, in GMMs, clustering algorithms are often employed. For clustering GPDF data, conventional Self Organizing Map (SOM) [5] and k-means [6] algorithms have been most widely used in

I. King et al. (Eds.): ICONIP 2006, Part II, LNCS 4233, pp. 439–447, 2006.
© Springer-Verlag Berlin Heidelberg 2006

practice because of their simplicity. Later, the Fuzzy C-Means (FCM) clustering algorithm is proposed as an improvement of the k-means and the SOM [7,8]. The FCM has been successfully applied in clustering the probabilistic distribution of the log-value of the frame size in the MPEG video classification model proposed by Liang and Mendel [4]. However, these algorithms were designed with the Euclidean distance. This implies that most of video classification models using these clustering algorithms used only mean values of GPDF data for clustering while leaving out covariance information of GPDF data. To exploit entire information in data including the mean value and covariance information, Park and Kwon proposed a divergence-based centroid neural network (DCNN) algorithm for clustering GPDF data [9]. The DCNN has been successfully applied to the clustering GPDF data for Hidden Markov Model (HMM) in speech applications.

In this paper, a MPEG video traffic classification model using a Fuzzy C-Means Algorithm with a Divergence-based Kernel (FCMDK) is proposed. The proposed classification model is designed for the classification of compressed video data without going through the decompressing procedure. The FCMDK adopted in the proposed classification model is used for clustering the GPDF data. The FCMDK is based on the FCM algorithm and thus utilizes advantageous features of fuzzy clustering techniques. Before clustering, the input data is projected to a feature space using a kernel method. The kernel method adopted in the FCMDK is used to transform the input data from a low dimensional space to a feature space of a higher dimensionality [10,11]. Consequently, nonlinear problems associated with the input space can be solved linearly in the feature space according to the well-known Mercer theorem [12]. Furthermore, the statistical nature of the data is utilized by using both the mean value and covariance information in GPDF data. For clustering of probability data, a divergence-based kernel using a divergence measure as its measure distance between two probability distributions is employed.

The remainder of this paper is organized as follows. Section 2 summarizes the Fuzzy C-Means and the Kernel-based Fuzzy C-Means algorithms. Section 3 introduces the Fuzzy C-Means algorithm with Divergence-based Kernel. Section 4 presents experiments and results on several MPEG video data sets including comparisons with other conventional algorithms. Conclusions are presented in Section 5.

2 Kernel-Based Fuzzy C-Means Algorithm

2.1 Fuzzy C-Means Algorithm

The FCM algorithm has successfully been applied to a wide variety of clustering problems. The FCM algorithm attempts to partition a finite collection of elements $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N\}$ into a collection of C fuzzy clusters. Bezdek first generalized the *fuzzy ISODATA* by defining a family of objective functions $J_m, 1 < m < \infty$, and established a convergence theorem for that family of objective functions [7,8]. For the FCM, the objective function is defined as :

$$J_m(U, \boldsymbol{v}) = \sum_{i=1}^{C} \sum_{k=1}^{N} \mu_{ik}^{m} \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2 \tag{1}$$

where $\|.\|^2$ denotes Euclidean distance measure, $\boldsymbol{x}_k$ and $\boldsymbol{v}_i$ is the input data, k, and cluster prototype, i, respectively. μ_{ki} is the membership grade of the input data $\boldsymbol{x}_k$ to the cluster $\boldsymbol{v}_i$, and m is the weighting exponent, $m \in \{1, \cdots, \infty\}$, while N and C are the number of input data and clusters, respectively.

The FCM objective function is minimized when high membership grades are assigned to objects which are close to their centroid and low membership grades are assigned when objects are far from their centroid [8].

By using the Lagrange multiplier to minimize the objective function, the center prototypes and membership grades can be updated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{C} \frac{\|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2}{\|\boldsymbol{x}_k - \boldsymbol{v}_j\|^2}} \tag{2}$$

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} \mu_{ik}^{m} \boldsymbol{x}_k}{\sum_{k=1}^{N} \mu_{ik}^{m}} \tag{3}$$

The FCM finds the optimal values of group centers iteratively by applying Eq. (2) and Eq. (3) in an alternating fashion.

2.2 Kernel-Based Fuzzy C-Means Algorithm

Though the FCM has been applied to numerous clustering problems [13], it still suffers from poor performance when boundaries among clusters in the input data are nonlinear. One alternative approach is to transform the input data into a feature space of a higher dimensionality using a nonlinear mapping function so that nonlinear problems in the input space can be linearly treated in the feature space according to the well-known Mercer theorem [12,11]. One of the most popular data transformation methods adopted in recent studies is the kernel method [10]. One of the advantageous features of the kernel method is that input data can be implicitly transformed into the feature space without knowledge of the mapping function. Further, the dot product in the feature space can be calculated using a kernel function.

With the incorporation of the kernel method, the objective function in the feature space using the mapping function Φ can be rewritten as follow:

$$F_m = \sum_{i=1}^{C} \sum_{k=1}^{N} \mu_{ik}^{m} \|\Phi(\boldsymbol{x}_k) - \Phi(\boldsymbol{v}_i)\| \tag{4}$$

Through kernel substitution, the objective function can be rewritten as:

$$F_m = 2 \sum_{i=1}^{C} \sum_{k=1}^{N} \mu_{ik}^{m} (1 - K(\boldsymbol{x}_i, \boldsymbol{v}_k)) \tag{5}$$

where $K(\boldsymbol{x}, \boldsymbol{y})$ is a kernel function used for calculating the dot product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in the feature space. To calculate the kernel between two vectors, the Gaussian kernel function is widely used:

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{\sigma^2}\right) \tag{6}$$

By using the Lagrange multiplier to minimize the objective function, the cluster prototypes can be updated as follow:

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m K(\boldsymbol{x}_k, \boldsymbol{v}_i)\boldsymbol{x}_k}{\sum_{k=1}^{N} \mu_{ik}^m K(\boldsymbol{x}_k, \boldsymbol{v}_i)} \tag{7}$$

And the membership grades can be updated as follow:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{C} \left(\frac{1-K(\boldsymbol{x}_k, \boldsymbol{v}_i)}{1-K(\boldsymbol{x}_k, \boldsymbol{v}_j)}\right)^{\frac{1}{m-1}}} \tag{8}$$

3 Fuzzy C-Means Algorithm with Divergence-Based Kernel

Since conventional kernel-based clustering algorithms were designed for deterministic data, they cannot be used for clustering probability data. In this paper, we propose a Fuzzy C-Means algorithm with a Divergence-based Kernel (FCMDK) in which a divergence distance is employed to measure the distance between two probability distributions. The proposed FCMDK incorporates the FCM for clustering data and the divergence-based kernel method for data transformation.

For GPDF data, each cluster prototype is not represented by a deterministic vector in the input space but is represented by a GPDF with a mean vector and covariance matrix. In order to calculate the kernel between two GPDF data, a divergence-based kernel is employed. The divergence-based kernel is an extension of the standard Gaussian kernel. While the Gaussian kernel is the negative exponent of the weighted Euclidean distance between two deterministic vectors as shown in Eq. 6, the divergence-based kernel is the negative exponent of the weighted divergence measure between two GPDF data. The divergence-based kernel function between two GPDF data is defined as follows:

$$DK(g_{\boldsymbol{x}}, g_{\boldsymbol{y}}) = \exp\left(-\alpha D(g_{\boldsymbol{x}}, g_{\boldsymbol{y}}) + \beta\right) \tag{9}$$

where $DK(g_{\boldsymbol{x}}, g_{\boldsymbol{y}})$ is the divergence distance between two Gaussian distributions, $g_{\boldsymbol{x}}$ and $g_{\boldsymbol{y}}$. α and β are the constants which depend on the data. After evaluating several divergence distance measures, the popular Bhattacharyya distance measure is employed. The similarity measure between two distributions using the Bhattacharyya distance measure is defined as follows:

$$D(G_i, G_j) = \frac{1}{8}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \left[\frac{\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j}{2}\right]^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \frac{1}{2}\ln\frac{\left|\frac{\boldsymbol{\Sigma}_i+\boldsymbol{\Sigma}_j}{2}\right|}{\sqrt{|\boldsymbol{\Sigma}_i|\,|\boldsymbol{\Sigma}_j|}} \tag{10}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ denote the mean vector and covariance matrix of a Gaussian distribution G_i, respectively. T denotes the transpose matrix.

Similar to the cluster prototypes and membership grades in the kernel-based FCM, the cluster prototypes and membership grades in the FCMDK can be updated using a Lagrange multiplier to minimize its objective function. However, each cluster prototype representing a cluster in the FCMDK is a probability distribution with a mean vector and a covariance matrix. Therefore, cluster prototypes in each iteration are updated by modifying their mean vector and covariance matrix as follows:

$$m_{\boldsymbol{v}_i} = \frac{\sum_{k=1}^{N} \mu_{ik}^m DK(\boldsymbol{x}_k, \boldsymbol{v}_i) m_{\boldsymbol{x}_k}}{\sum_{k=1}^{N} \mu_{ik}^m DK(\boldsymbol{x}_k, \boldsymbol{v}_i)} \tag{11}$$

$$\Sigma_{\boldsymbol{v}_i} = \frac{\sum_{k=1}^{N} \mu_{ik}^m DK(\boldsymbol{x}_k, \boldsymbol{v}_i) \Sigma_{\boldsymbol{x}_k}}{\sum_{k=1}^{N} \mu_{ik}^m DK(\boldsymbol{x}_k, \boldsymbol{v}_i)} \tag{12}$$

where $m_{\boldsymbol{v}_i}$ and $m_{\boldsymbol{x}_k}$ are the mean of the cluster prototype $\boldsymbol{v}_i$ and the vector in input $\boldsymbol{x}_k$, respectively. $\Sigma_{\boldsymbol{v}_i}$ and $\Sigma_{\boldsymbol{x}_k}$ are the covariance of the cluster prototype $\boldsymbol{v}_i$ and the vector in input $\boldsymbol{x}_k$, respectively. $DK(\boldsymbol{x}_k, \boldsymbol{v}_j)$ is the divergence-based kernel function between two Gaussian distributions $\boldsymbol{x}_k$ and $\boldsymbol{v}_j$.

The membership grades are similar to those in the KFCM and can be updated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{1-DK(\boldsymbol{x}_k,\boldsymbol{v}_i)}{1-DK(\boldsymbol{x}_k,\boldsymbol{v}_j)}\right)^{\frac{1}{m-1}}} \tag{13}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{v}_i$ are the probability distribution input vector and probability distribution cluster prototype, respectively. $DK(\boldsymbol{x}_k, \boldsymbol{v}_j)$ is the divergence-based kernel function between two Gaussian distributions, $\boldsymbol{x}_k$ and $\boldsymbol{v}_j$.

With the incorporation of the divergence-based kernel method and the FCM, the proposed FCMDK can be used for clustering GPDF data while utilizing the advantageous features of the fuzzy clustering techniques and the kernel method. Thus, it provides an efficient clustering algorithm for GPDF data.

4 Experiments and Results

To demonstrate the performance of MPEG video traffic classification model using the FCMDK, several MPEG video traces were used for experiments. These MPEG video traces were coded with the MPEG-1 standard according to the Moving Picture Expert Group. Table 1 shows the list of video traces used in our experiments. These data are provided by the University of Wuerzburg, Wuerzburg, Germany and are available at the following website:

http://www3.informatik.uni-wuerzburg.de/MPEG/

Table 1 consists of 5 *"movie"* traces and 5 *"sports"* traces. Each trace consists of 40,000 frames which result in 3,333 GOPs. Each GOP can be represented by the sequence $IBBPBBPBBPBB$ with 12 frames for each GOP. More details on these video traces can be found in [14].

Table 1. MPEG-1 Video used for experiments

MOVIE	SPORTS
"Jurassic Park"	*"ATP Tennis Final"*
"The Silence of the Lambs"	*"Formula 1 Race: GP Hockenheim 1994"*
"Star Wars"	*"Super Bowl Final 1995: SanDiego-San Francisco"*
"Terminator 2"	*"Two 1993 Soccer World Cup Matches"*
"A 1994 Movie Preview"	*"Two 1993 Soccer World Cup Matches"*

From video traces in Table 1, we used the first 24,000 frames, resulting in 2,000 GOPs from each trace, for training and the remaining frames from each trace for testing. Fig. 1 shows an example of MPEG-1 data with I-, P-, and B-frame from the video data *"Two 1993 soccer World Cup matches"*.

The proposed classification model using the FCMDK is based on a Gaussian Mixture Model (GMM) and a Bayesian classifier. In order to model and classify the MPEG video data, we consider the MPEG data as Gaussian Probability Density Function (GPDF) data [14]. The classification process of proposed classification model can be divided into two steps: the modeling step and the classification step. In the modeling step, mixture components of GMMs are obtained using the FCMDK algorithm. Then, in the classification step, a Bayesian classifier is employed to decide the genre, *"movie"* or *"sports"*, to which a video sequence belongs. The genre decision procedure can be summarized by the following equations:

$$Genre(x) = \arg\max_i P(x|v_i) \tag{14}$$

$$P(x|v_i) = \sum_{i=1}^{M} c_i \aleph(x, \mu_i, \Sigma_i) \tag{15}$$

$$\aleph(x, \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-0.5(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)} \tag{16}$$

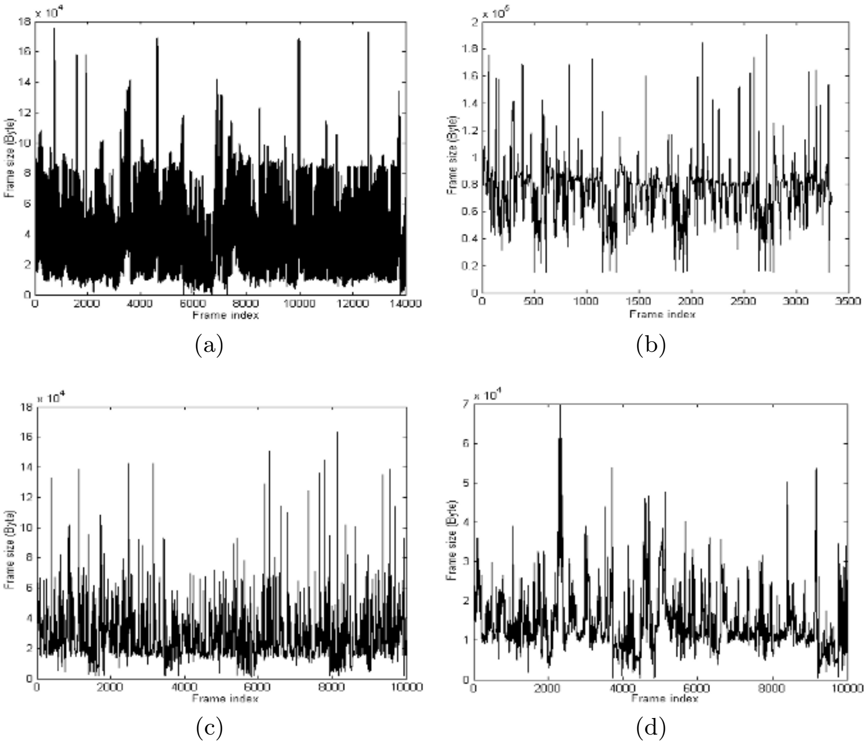

Fig. 1. Example of MPEG data: (a) whole data (b) I-frame (c) P-frame (d) B-frame

where M is the number of code vectors, c_i is weight of the code vectors, d is the number of dimensions of feature vectors (d = 12), and m_i and Σ_i are the mean and covariance matrix of the i-th group of the genre's distribution, respectively.

In order to evaluate the performance of the proposed classification model, the classification performance is measured by the False Alarm Rate (FAR) which is calculated by the following equation:

$$\text{FAR}(\%) = \frac{\text{Number of misclassification GOPs}}{\text{Total number of GOPs}} \times 100 \qquad (17)$$

One of the most important parameters that has to be selected in most clustering algorithms is the number of clusters in the data. Most clustering algorithms partition data into a specified number of clusters, regardless of whether the clusters are meaningful. In our experiments, the number of code vectors is varied from 3 to 8 in order to determine a sufficient number of code vectors to represent the number of mixture components in the GMMs. Fig. 2 shows the classification performance in terms of FAR of classification models using the SOM, the k-means, and the FCMDK. As can be seen from Fig. 2, the FARs of all classification models are decreased significantly when the number of code vectors is increased from 3 to 5 while they tend to saturate when the number of code

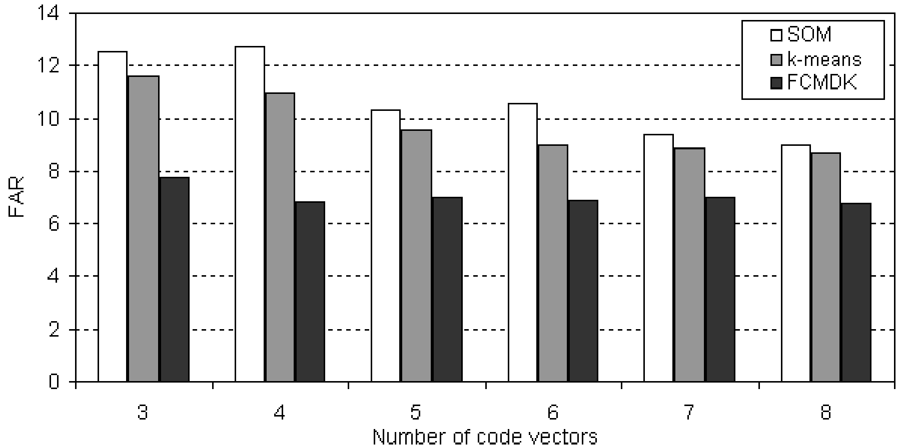

Fig. 2. Overall classification accuracies using different algorithms

Table 2. Average FAR (%) of different classification models

	Overall FAR(%)
SOM	**10.748**
k-means	**9.789**
FCMDK	**7.029**

vectors is greater than 5. This implies that using 5 code vectors for representing the number of mixture components is sufficient.

Table 2 summarizes the classification performance in terms of FAR for different models using the SOM, the k-means, and the proposed FCMDK. As can be seen from Table 2, the classification model using the proposed FCMDK outperforms the models using the SOM and k-means. Improvements in terms of FAR of 28.19% and 34.60% are achieved over the k-means and the SOM algorithms, respectively. These results imply that the covariance information plays an important role in modeling and classification of MPEG video traffic data. By using divergence-based kernel, the FCMDK can utilize the covariance information of the GPDF data for clustering. Thus, it can be used as an efficient tool for clustering GPDF data in GMMs.

5 Conclusion

A new approach for modeling and classification of MPEG video traffic data using a Fuzzy C-Means (FCMDK) algorithm with a Divergence-based Kernel is proposed in this paper. The proposed classification model is based on a Gaussian Mixture Model (GMM) and a Bayesian classifier. The FCMDK adopted in the proposed classification model is employed for clustering of the GPDF data in GMMs. The proposed classification model using the FCMDK for clustering of

GPDFs is applied to a modeling and classification problem of MPEG video traffic data. Our experiments and results for several MPEG video traffic data sets show that respective improvements of 28.19% and 34.60% in terms of FAR are archived over the conventional k-means and the SOM algorithms, respectively. The proposed MPEG video traffic classification model provide an efficient tool for organizing and retrieval of video databases.

References

1. Dawood, A.M., Ghanbari,M.: MPEG Video Modeling Based on Scene Description. IEEE Int. Conf. Image Processing, Chicago, IL., Vol. 2 (1998) 351-355.
2. Manzoni,P.,Cremonesi,P.,Serazzi,G.: Workload models of VBR video traffic and their use in resource allocation policies. IEEE Trans. on Networking 7 (1999) 387-397
3. Krunz, M., Sass, R., Hughes, H.: Statistical Characteristics and Multiplexing of MPEG streams. Proc. IEEE Int. Conf. Comput. Commun., INFOCOM'95, Boston, MA, Vol. 2 (1995) 445-462
4. Liang, Q., Mendel, J.M.: MPEG VBR Video Traffic Modeling and Classification Using Fuzzy Technique. IEEE Trans. on Fuzzy Systems 9 (2001) 183-193
5. Kohonen,T.: The Self-Organizing Map. Proc. IEEE, Vol. 78 (1990) 1464-1480
6. Hartigan, J.: Clustering Algorithms. New York, Wiley (1975)
7. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. New York(Plenum), (1981)
8. Bezdek, J.C.: A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms. IEEE Trans. on Pattern Analysis and Machine Intelligence PAMI-2(1) (1980) 1-8
9. Park, D.C., Kwon, O.H.: Centroid Neural Network with the Divergence Measure for GPDF Data Clustering. IEEE Trans. on Neural Networks (in review).
10. Muller, K.R., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An Introduction to Kernel-Based Learning Algorithms. IEEE Transactions on Neural Networks 12(2) (2001) 181-201
11. Girolami,M.: Mercer Kernel-Based Clustering in Feature Space. IEEE Trans. on Neual Networks 13(3) (2002) 780-784.
12. Cover, T.M.: Geomeasureal and Statistical Properties of Systems of Linear Inequalities in Pattern Recognition. Electron. Computing, Vol. EC-14 (1965) 326-334.
13. Chen, S., Zhang, D.: Robust Image Segmentation using FCM with Spatial Constraints Based on New Kernel-Induced Distance Measure. IEEE Trans. on Systems, Man and Cybernetics 34(4) 2004 1907-1916
14. Rose, O.: Satistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM systems. Univ. Wurzburg, Inst. Comput. Sci., Rep., 101 (1995)