Principal Component Analysis of Video Streams for Video-on-Demand Services

Raul V. Ramirez-Velarde, Raul Perez-Cazares and Carlos F. Pfeiffer Celaya [1]

[1] Instituto Tecnológico y de Estudios Superiores de Monterrey- Campus Monterrey, Eugenio Garza Sada 2501 Sur, Monterrey, N.L., 64849. Mexico
{rramirez, raul.perez, pfeiffer}@itesm.mx

Abstract. The success of today's video-on-demand services, such as the ones provided by cable companies, depends on large availability of video titles. But this availability makes video server desgn very difficult because each video sequence will have different statistical characteristics such as mean, variance, symmetry, kurtosis and fractal dimension. Thus, a server needs to be designed as if the most demanding of all video traces will always be selected by users. This is a dimensionality problem. We can say that if a video server offers thousands of titles to subscribers, then its design has thousands of dimensions, since we don't know what would be the particular arrangement of subscriber selections at any given time, nor can we predict future releases statistical characteristics. These characteristics can require demand of resources that the hardware infrastructure might be able to meet. In this paper, we show a very well known technique known as principal component analysis that can help to dramatically reduce dimensionality by using only a few principal components when creating a characteristic video trace, that will be used as a representative of all video traces for video server design.

Keywords: video-on-demand, principal component analysis, statistical analysis, probability distribution.

1 Introduction

Now day's digital multimedia entertainment systems are becoming ever more common. These entertainment systems place heavy demands over the current infrastructure. They require large storage systems, high bandwidth, strong computing power, very precise algorithms, etc. And the most demanding of all is video. Current video-on-demand (VOD) systems deliver either low quality video, such as YouTube and MetaCafe, or high quality video at relatively high prices, such as cable systems services. The main disadvantage of all these services though, is the availability of interesting content. For example, Internet VOD sites have available tens of thousands of videos, but most of them are community contributed or short clips of commercial productions. At the same time, Cable VOD systems store between 100 and 300 different programs. If we take into consideration that about 64,000 movies have been produced since the industry began and take into consideration that many more TV programs, sport films, documentaries, newsreels, shopping catalogs and other types of video could be included in VOD services, we can see that current high quality VOD availability is not enough [1].

This is shown by the fact that although initial expectation for VOD based revenue by cable companies was very high, and still is, the average revenue per unit (ARPU) has been calculated a low as $13 [2], and at best is about $1.99 per video selection [3], which is very low compared to the $87 ARPU that cable companies are used to.

The main problem, like the sister music on demand and music download industry, is the availability of high quality materials. If such availability of titles satisfies demand, Spencer Wang of JPMorgan Chase predicts that revenue form VOD and download related services will out perform revenue based on advertisement spot broadcasting.

But large availability of high quality content in current VOD technology is not as simple as just storing the videos and servicing them. Each video possesses different statistical qualities such as GOP structure, mean, variance, symmetry and Kurtosis of frame size, fractal dimension and Hurst parameter, playback frame rate, entropy energy, time correlations, etc. These variables affect VOD system parameters such as disk scheduling algorithm, real-time jobs processing and programming, memory buffer size, CPU core allocation, etc.

Furthermore, although all statistical characteristics can be computed off-line, before a video request is serviced, the particular combination of user tastes cannot. That is to say, that in a particular moment in time, most videos requested by users could be cartoons, at other times could be motion pictures, at other times there will not be a particular gender that is dominant, etc. Even if a Zipf based VOD server design can be carried out [4, section 7.4.8], that is to say that the effect each video has on VOD server will be weighted using estimated using preferences [5], this preferences curve will most likely change month after month, as new movies are produced, and old ones are included into video server storage.

And even after that, a model that correctly predicts VOD server behavior based on estimated stored video statistics is still a work in progress. The main problem here is the chaotic behavior of video streams. Such behavior makes it very difficult to design VOD servers, as it is very likely that real-time behavior will deviate significantly from sample, and even population statistics, a phenomenon called in probability Heavy-Tails. There are several models, such as [6] and [7] based on self-similarity, [8] and [9] based on heavy-tailed probability distributions, and [10], [11] and [12] based on the Gamma probability distribution; but some of them are asymptotic models that can only be used on very specific situations, or use probability distributions that cannot approximate the heavy-tail observed in video data (and thus are not useful for very high degrees of service quality) and do not in general, solve the problem of high diversity of available video streams. Furthermore, although there are currently commercial VOD services available, mainly delivered by cable companies, those services will, in the near future, need be able to service HDTV programming. This quality jump on video services will make the price of video delivery hardware skyrocket. Our research shows that a HDTV video stream can require, if the server load is high and the quality of service requirements are high, up to 1.5 Gigabytes of RAM. So if we take into consideration that high-end VOD server RAM can cost up to $1000 a Gigabyte, in order to service 100 different simultaneous streams we would need an investment of up to $150,000 only in server memory.

To address these problems we use the following technique. We treat the availability of large numbers of video traces as a problem of muti-dimensionality. We consider each video file to be a dimension for video frame size data. Then we use principal components analysis (PCA) to determine a Characteristic Video Trace (CVT), that is, a video trace that captures as much of variability of all stored video files as possible. We use PCA to reduce the dimensionality of video file availability from possible hundreds and thousands to one, two or at most three. The CVT is created by combining the Component Loadings (CL) generated by just a few principal components into Component Scores (CS).

In regards to VOD server modeling and design this paper continues previous work on VOD design in which we have created models that can determine the maximum amount of simultaneous subscribers that a video server can attend based on disk throughput, the size of each video stream memory buffer (arguably the most expensive part of a VOD server) and service rate regulation, thus the models provide scalability rules that will make HDTV VOD economically successful [13][14].

The rest of the articles is organized a follows. In section 2 we present the results of the principal component analysis. In section three we present the characteristics of the CVT and in section 4 we present our conclusions.

2 Statistical Analysis of Video Traces

2.1 Statistics of the Video Traces

The video traces that were used in this study are the following (Table 1) [15]:

Mr. Bean	Silence of the Lambs
Star Trek – First Contact	Alpine Ski
Die Hard III	Soccer European Championship 1996
From Dusk Till Dawn	Star Wars IV
The Firm	Starship Troopers
Jurassic Park	

Table 1. Analyzed video streams

We used the first 65534 elements form each video trace and got the following results (Table 2):

Video Trace	Minimum Frame	Maximum Frame	Mean Frame	Standard Deviation
bean	93.000	14274.000	3079.286	1815.192
contact	47.000	11090.000	1510.531	1125.738
Die ha	71.000	16840.000	3402.324	2132.224
dusk	74.000	15745.000	3108.756	1811.266
firm	32.000	10204.000	1539.672	1196.799
jurasic	72.000	16745.000	3917.219	2296.278
silence	158.000	22239.000	3008.143	2457.317
skii	307.000	15780.000	3815.135	2224.231
soccer	130.000	17657.000	5554.905	2253.875
star wa	26.000	9370.000	1364.831	917.981
tropers	224.000	14592.000	2856.820	1545.513

Table 2. Statistical characteristics of video streams

2.2 Principal Components Analysis

Principal component Analysis is a statistical technique used to reduce multidimensional data sets to lower dimensions for analysis. PCA is mathematically defines as an orthogonal lineal transformation that rotates the data to a new coordinate system such that the variance of the projection of the data to first coordinate is greatest, the variance of the second coordinate is the second largest and so on.

PCA projections (called Component Scores or Component Observations) conserve those characteristics of the data ser that contribute the most to the variance. But since the variance of the projections λ_k decreases (l1 being the largest) as k increases, PCA can be used to select a few derived projections that preserve the most information about variance, covariance and correlation.

Suppose that **x** is a vector of p random variables, and that the variances of the p random variables and the structure of the covariances or correlations between the p variables are of interest. We wish to look for a few (<<p) derived variables that preserve most of the information given by these variances and correlations or covariances. Let $\mathbf{\Sigma}$ be the covariance matrix of **x** [16]. This is the matrix whose (i,j)th elements the covariance between the ith and the jth elements if $\boldsymbol{x}$ when $i \neq j$, and the variance of the jth element of **x** when $i=j$. For k=1, 2, … , p, the Component Observation or Component Score (CS) on the kth Principal Component (PC) is given by $z_k = \boldsymbol{\alpha}_k^t \mathbf{x}$ where $\boldsymbol{\alpha}_k$ is an eigenvector of $\mathbf{\Sigma}$ corresponding to it s kth largest eigenvalue λ_k. Furthermore $\boldsymbol{\alpha}_k^t \boldsymbol{\alpha}_k = 1$, then var($z_k$)=$\lambda_k$. We call $\boldsymbol{\alpha}_k$ the Component of Loadings (CL) for PC k.

Let $\mathbf{z} = \mathbf{A}^t \mathbf{x}$, where **A** is the orthogonal matrix whose kth column, $\boldsymbol{\alpha}_k$, is the kth eigenvector of $\mathbf{\Sigma}$ (**A** is orthogonal because $\boldsymbol{\alpha}_k^t \boldsymbol{\alpha}_k = 1$). Then

$$\mathbf{\Sigma A} = \mathbf{A \Lambda}, \tag{1}$$

Where Λ is the diagonal matrix whose kth diagonal element is λ_k, the kth eigenvalue of Σ and λ_k=var(z_k). Thus

$$\mathbf{A}^t \boldsymbol{\Sigma} \mathbf{A} = \boldsymbol{\Lambda} , \qquad (2)$$

and,

$$\boldsymbol{\Sigma} = \mathbf{A} \boldsymbol{\Lambda} \mathbf{A}^t . \qquad (3)$$

The PCs are found by a singular value decomposition of the covariance matrix Σ. A similar technique which uses the correlation matrix can be found in [16].

2.3 Video Trace PCA

To find the PCs, a correlation matrix was built (shown in Table 3):

Video	bean	contact	die ha	dusk	firm	jurasic	silence	skii	soccer	star w	tropers
bean	1	0.421	0.370	0.444	0.384	0.357	0.116	0.269	0.355	0.466	0.352
contact	0.421	1	0.267	0.319	0.322	0.298	0.102	0.180	0.245	0.370	0.304
die ha	0.370	0.267	1	0.271	0.302	0.244	0.118	0.148	0.261	0.315	0.227
dusk	0.444	0.319	0.271	1	0.341	0.314	0.206	0.286	0.324	0.391	0.380
firm	0.384	0.322	0.302	0.341	1	0.303	0.282	0.326	0.281	0.443	0.367
jurasic	0.357	0.298	0.244	0.314	0.303	1	0.064	0.181	0.244	0.394	0.296
silence	0.116	0.102	0.118	0.206	0.282	0.064	1	0.230	0.077	0.275	0.190
Ski	0.269	0.180	0.148	0.286	0.326	0.181	0.230	1	0.181	0.287	0.199
soccer	0.355	0.245	0.261	0.324	0.281	0.244	0.077	0.181	1	0.334	0.298
star w	0.466	0.370	0.315	0.391	0.443	0.394	0.275	0.287	0.334	1	0.375
tropers	0.352	0.304	0.227	0.380	0.367	0.296	0.190	0.199	0.298	0.375	1

Table 3. Video stream correlation matrix

The resultant PCs and their effect are shown in Table 4, with the Scree and Cumulative Variance plots shown in Fig. 1.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Eigenvalue	4,359	1,400	1,028	0,798	0,781	0,754	0,655	0,646	0,567	0,537	0,476
Variability (%)	36,322	11,665	8,563	6,652	6,505	6,287	5,457	5,381	4,723	4,478	3,967
Cumulative %	36,322	47,987	56,550	63,203	69,708	75,994	81,451	86,833	91,555	96,033	100,000

Table 4. PCA results

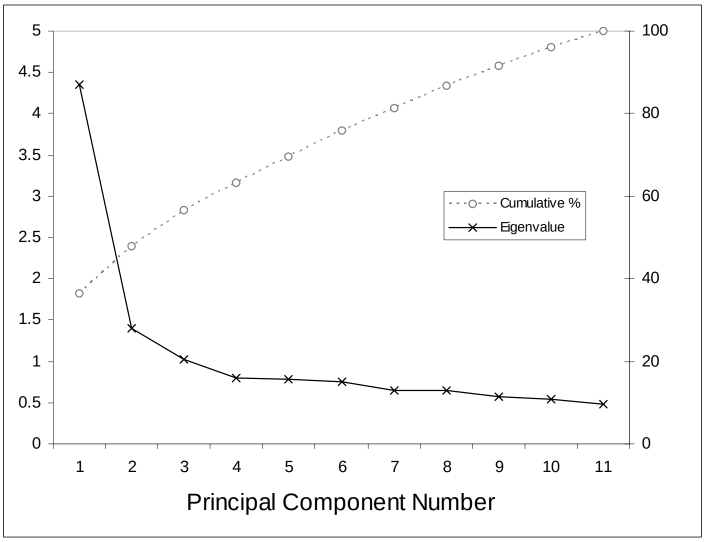

Figure 1. PCA Scree plot and cumulative variance plots

In Table 5 we show each video's contribution to each PC. The first PC, almost three times as large as the second largest, is not dominated by any particular stream, which means that it truly represents the average or characteristic behavior of the data. The video stream that contributes the most to this PC contributes 11.81% whereas the 4.64% marks the smallest contribution. The rest of the PCs have varying degrees of dominance from different videos, with usually one video dominating one of the smaller PCs.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
bean	11.318	0.540	7.878	0.054	2.245	2.473	0.946	2.998	13.035	5.890	**52.623**
contact	7.763	0.156	7.318	1.874	0.008	**48.237**	9.655	3.170	18.101	2.253	1.467
die ha	6.082	0.744	10.734	**36.274**	15.499	11.578	8.471	2.667	6.241	0.128	1.583
dusk	9.570	2.112	0.504	6.538	1.395	0.044	**37.028**	15.554	2.599	15.852	8.805
firm	9.635	4.618	2.083	1.281	0.139	0.041	7.895	24.890	15.311	**33.958**	0.149
jurasic	11.094	**32.644**	5.067	0.026	0.010	0.386	0.000	0.057	0.410	0.139	0.167
silence	2.298	16.009	**34.327**	13.368	6.132	2.292	0.386	14.195	5.067	0.002	5.924
Ski	4.641	7.146	12.863	24.047	**37.878**	0.424	0.198	2.299	7.428	2.942	0.134
soccer	6.452	0.853	13.572	13.268	1.714	33.788	21.437	3.331	4.579	0.923	0.083
star w	11.810	1.062	0.332	1.219	0.324	0.037	5.946	1.046	**19.836**	29.983	28.405
tropers	8.245	1.471	0.254	2.026	34.646	0.315	8.038	**29.736**	6.984	7.793	0.492

Table 5. Video contributions to PCs

To investigate if there are any correlations between videos, apart form the correlation matrix we can use the biplot. In Fig. 2 we show the biplot between PC1 and PC3. The red triangles indicate PC positions. The gray cloud maps the PC observations from each PC (see below). The fan-like spread of the PCs indicates that there are no statistical similarities between videos. Sadly, this fan-like structure and the almost even biplot-cloud also indicate that there are no clear statistical classes to use for video classification [16].

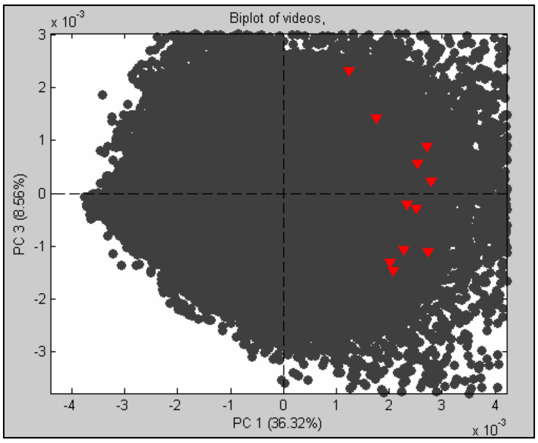

Figure 2. PC1 and PC3 biplot

2.4 Choosing Principal Component

There are several rules or criteria that can be used to select the number of m components that will keep most of the variability present in p variables. Naturally, we wish $m << p$, and this is called dimension reduction. In our case, we wish to derive a single video constructed by using only a very small number of PCs that will represent a collection of possible thousands of videos stored for VOD services. This single video, called the characteristic video will be used to design the video server. Table 6 shows how many principal components should be chosen according to different criteria. More complicated selection criteria can be found on [16]

Rule	Description	Minimum PCs to Keep
Cumulative percentage	Accumulated variability is at least 50% (80% is better)	3
Size of variance	PC variance (eigenvalue) exceeds 1	3
Jollife size of variance	PC variance (eigenvalue) exceeds 0.7	6
Scree graph	One PC after graph changes slope	2
Log-eigenvalue	The last PC is the first point which starts a straight line	4

Table 6. PC selection criteria

After choosing *m* principal components, the data dimension is reduced by carrying out the orthonormal lineal transformation:

$$y_i=\mathbf{B}^t x_i, \qquad (4)$$

where **B** is a (p x m) matrix (with orthonormal columns), $\mathbf{x}_i$, i=1, 2, … , n is the vector of n observations on the p-element random vector **x**, and note that the observations form the matrix **X** (n x p), and $\mathbf{y}_i$ is the p-element vector of scores produced by using only the first m PCs. Then $\mathbf{B}=\mathbf{A}_m$, where $\mathbf{A}_m$ consists of the first m columns of **A**.

Additionally, let **l** be the vector containing the first m components of λ, the vector of PC variances, and let $\mathbf{S}_m$ be the orthonormal-column matrix containing the first m columns of **Z**, the (n x p) matrix of component scores (**Z**=**XA**). Then

$$\mathbf{w}=\mathbf{S}_m\mathbf{l}, \qquad (5)$$

where **w** is the n-element weighted linear combination of m PCs. That is, our ***characteristic video*** vector.

Since the purpose of this investigation is to choose a very small number of PCs compared to the amount of available video titles, we took only one PC, the first one and largest.

3 Characteristics of the New Data Set

The new data set generated by PCA does indeed preserve the characteristics behavior of the generating video streams, such as GOP structure, frame size histogram distribution and autocorrelation. This is shown in Figs. 3, 4 and 5.

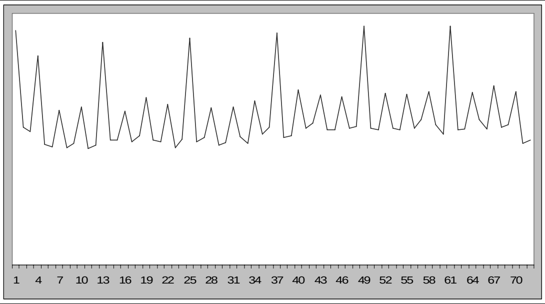

Figure 3a. Frame size plot for First Contac movie

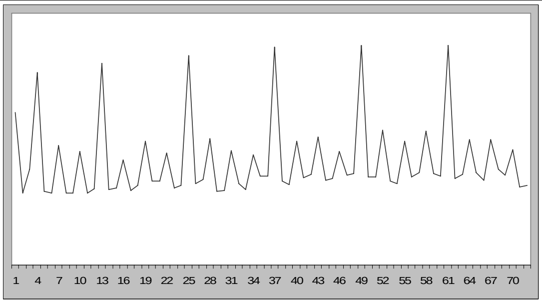

Figure 3b. Frame size plot for first PC data set

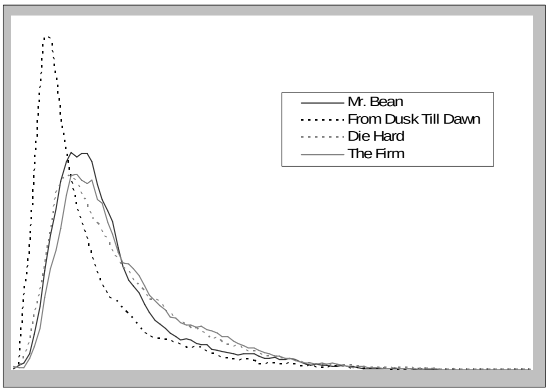

Figure 4a. Frame size histogram for several videos

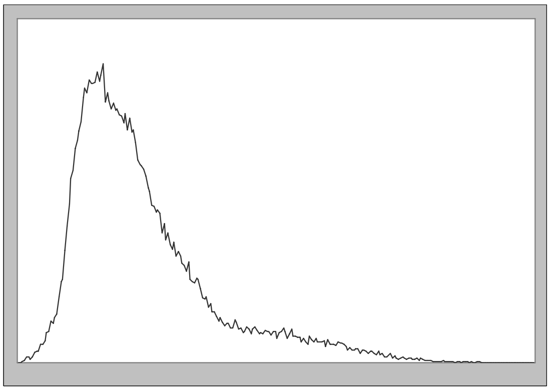

Figure 4b. Frame size histogram first PC data set

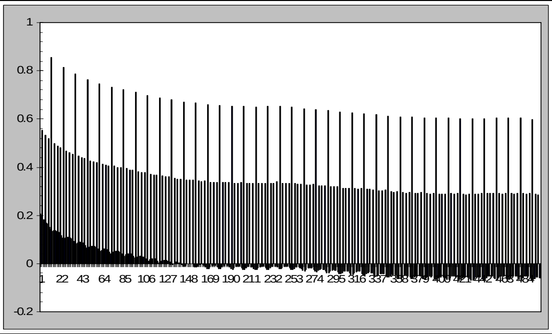

Figure 5a. Autocorrelation for Mr. Bean video

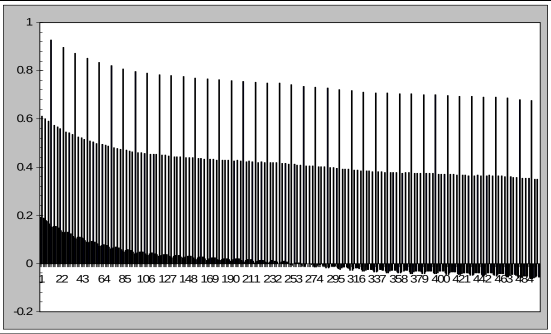

Figure 5b. Autocorrelation for first PC data set

It is important to note, that as reported in [17], it can be seen in Fig. 1 that the eigenvalues of the PCs decay following power-law rule. That is that the Spree graph presents a few large value PCs and then a long-tail of very slowly decaying eigenvalues with a rule similar to $k^{-\beta}$.

We also found that the CVT has fractal dimension as the Hurst parameter $0.5<H<1$.

This has important implications for the design of VOD services since it means that in order to capture a large percentage of total variation, many PCs must be taken. But it also means that since the very first few PCs are much larger than the rest, that it might be possible to use only one or two PCs. In our case the first

PC captures 36.32% of total variability which is very good, since a single video should capture only 9%, whereas two PCs capture 48%. We would expect two videos to represent only 18% of total variability.

4 Conclusions

PCA over video stream data for high-end commercial video services is a powerful tool. First, it shows that it is entire possible that there might not clear statistical classes of video sequences with similar characteristics that one could use for multi-class design of video servers. Secondly, although nominally the very few first PCs do in fact preserve no more than 50% of variability, this percentage is much high compared to each video expected contributed variability, simplifying video server design by permitting to generate a very representative new video data set that can be used as a characteristic video. This characteristic video acts as a proxy for all video streams which means that subjective procedures such as popularity weighting, worst case scenario and mean case scenario are not longer necessary. A set of multiple copies of the characteristic video represents all combinations of video streams, even an all worst-video scenario.

References

[1] D. Minolli. “Video Dialtone Technology”. McGraw-Hill, New York, 1995.

[2] D. Mermigas. “Reshaping of revenue at forefront of NATPE”. The Hollywood Reporter, Jan 24th, 2006.

[3] D. Sabbagh. “Cable TV Companies to Launch Video on Demand”. Times On Line, Jan 19th, 2005.

[4] Tanenabum, A. S. “Computer Networks. 4th Edition”. Prentice Hall, 2003.

[5] D. Sitaram and A. Dan. “Multimedia Servers”. Morgan Kauffman Publishers, San Francisco, U.S.A, 2000.

[6] J. Beran, R. Sherman, M.S. Taqqu, W. Willinger. “Long-range dependence in variable-bit-rate video traffic”. IEEE Transaction on Communications 43 (2–4) (1995) 1566–1579.

[7] I. Norros, “A Storage Model with Self-Similar Input”. Queueing Systems: Theory and Applications 16 (1994) 387–396.

[8] M. Garret, W. Willinger. “Analysis, modeling and generation of self-similar VBR”. Video Traffic, ACM SigComm, London, pp. 269–280, September, 1994.

[9] M.S. Taqqu, A representation for self-similar processes, Stochastic Processes and their Applications 7 (1978) 55–64.

[10] A. Chodorek, R.D. Chodorek. “Characterization of MPEG-2 video traffic generated by DVD applications”. 1st European conference on universal multi-service networks, pp. 62–70. ECOMN, 2000.

[11] O. Rose. “Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems”. IEEE Conference on Local Computer Networks, pp. 397–406, October, 1995.

[12] U. K. Sarkar, S. Ramakrishnan, D. Sarkar. ”Segmenting full-length VBR video into shots for modeling with markov-modulated Gamma-Based Framework”. Internet Multimedia Management Systems II, SPIE vol. 4519, pp. 191–202, 2001.

[13] R. V. Ramirez-Velarde and R. M. Rodriguez-Dagnino. "A gamma fractal noise source model for variable bit rate video servers". Computer Communications, Volume 27, Issue 18, 1 December 2004, Pages 1786-1798.

[14] R. V. Ramirez-Velarde, and R. M. Rodriguez-Dagnino. "Performance Analysis of a VBR Video Server with Gamma Distributed MPEG Data". SPIE, ITCOM 2003, Performance and Control of Next Generation Communication, Orlando, USA, October 2003.

[15] H. P. Fitzek and M. Reisslein. “MPEG-4 and H.263 Video Traces for Network Performance Evaluation”. Technical University Berlin, Technical Report TKN-00-06, October 2000.

[16] I. T. Jolliffe. “Principal Component Analysis”. Springer, U. S. A, 2002.

[17] J. B. Gao, Y. Cao, J.-M. Lee. “Principal component analysis of 1/fα noise”. Physics Letters A 314 (2003) 392–400, Elsevier.