A Study on the Distribution of Shot Lengths for Video Analysis

Cuneyt M. Taskiran and Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907-1285

ABSTRACT

In this paper we investigate the distribution of shot lengths for video sequences containing diverse content. Accurate models for shot lengths are important to model video both for content-based retrieval applications and for performing queuing analysis for the design of video buffers in multimedia networks. Using a large dataset collected from CSPAN programs we have analyze the Pareto, Weibull, and gamma distributions as possible models for the shot length distribution. We have compare the goodness of fit of these possible distribution models using the Kolmogorov-Smirnov statistic.

Keywords: Shot length modeling, heavy-tailed models, statistical video models, Pareto distribution, subexponential distributions

1. INTRODUCTION

In this paper we investigate the problem of choosing and fitting models for the distribution of shot lengths for broadcast video programs. Modeling the distribution of shot lengths for video is important for the following applications, among others:

Shot boundary detection. A shot, which is the basic component of a video sequence, may be defined as a group of frames that has continuity in some general conceptual or visual sense. Virtually all systems for indexing, characterization, and understanding of video rely on the identification of individual shots. Hence, usually the first task in processing video data is segmenting shots by detecting shot boundaries, thereby breaking the video sequence into distinct "semantic chunks." The general approach is to derive one or more low-level features from video frames, derive a dissimilarity value between frames using these features, and flag a shot transition whenever this value shows some nontypical behavior. A better approach would be to use a model for the probability distribution of shot lengths as *a priori* knowledge in a Bayesian framework to derive an adaptive threshold for shot boundary detection.[1] This method is superior to using a fixed threshold to detect shot boundaries and is more robust to changes caused by camera and object motion.

Video genre classification. It is evident that, for an information source as rich as video, efficient indexing and retrieval require some form of automatic analysis of semantic content which may be achieved by assigning semantic labels to individual shots. Currently, however, automatic extraction of truly semantic features such as "young girl running," or "park scene" is not possible. One way to circumvent this problem is to define features that correlate well with high level semantic labels and hence are useful in bridging the gap between low-level image features and semantic labels.[2] We call such features *pseudo-semantic features*.[3] Shot length is one such feature. For example, action sequences contain a large number of consecutive short shots whereas for other content, like news stories and talk shows, longer shot lengths are more common. It has been shown that shot length distribution is a simple yet effective feature which correlates well with the content and genre of a video sequence.[3]

Video Traffic Modeling. Accurate traffic models of variable bit rate (VBR) coded video is necessary for prediction of performance of any multimedia network. Various models have been proposed for modeling variations in the bit rate of VBR video.[4,5] In the last decade researchers have found that video conference data and packet counts per unit time in Ethernet traffic exhibit long range dependence and self-similarity, while quantities like World Wide Web file transmission times, UNIX file sizes, and cpu time to complete a job appear to be generated by distributions

This work was supported by CSPAN and a grant from the Indiana 21st Century Research and Technology Fund. Address all correspondence to E. J. Delp, `ace@ecn.purdue.edu`, `http://www.ece.purdue.edu/~ace` or telephone: +1 765 494 1740.

with heavy tails.[6,7] This fact is disturbing since it makes invalid the classical assumptions widely used in queuing analysis such as exponential waiting times and iid-ness of events. VBR coded video bit rate also exhibits multiple-scale variations which is most probably caused by the fact that the length of shots are generated from heavy-tailed distributions. Based on these observations, models that incorporate long range dependence were recently proposed as more faithful models for VBR video traffic.[8] The accurate determination of the distribution function of shot lengths is central in these and most other traffic models.

1.1. Previous Work

Sarkar *et al.*[5] use a geometrical distribution to model the length of shots which is expressed in number of GOPs in the shot. Frater *et al.*[4] propose the following probability mass function for the length of shots

$$f(m) = \frac{a}{m^k + b^2}$$

where $n \approx 2$ is used to model two movie sequences. Krunz and Ramasamy[8] propose a formula in which the probability mass function of shot lengths is expressed in terms of the empirical autocorrelation function. They also derive expressions which relate the performance at a video buffer to the distribution of shot lengths of video. Pareto and Weibull distributions were considered for shot lengths. Heyman and Lakshman[9] examine the gamma, Weibull, and Pareto densities for modeling shot lengths. They use a plot of $\gamma_1 = \sigma/\mu$ versus $\gamma_3 = \mu_3/\sigma^3$ where μ, σ, and μ_3 are the mean, variance, and the third central moment estimated from data, respectively, to identify the underlying distribution function for a variety of broadcast video programs. Unfortunately, such high moments may not exist if the underlying distribution has heavy tails. Their analysis revealed that different video programs followed different distributions. For some sequences no good fit was found. Dawood and Ghanbari[10] use a second order gamma distribution to model shot lengths. Jelenkovic *et al.*[11] develop a rigorous video model for queuing analysis. Based on their experiments they conclude that MPEG traffic exhibits subexponential behavior. They use a Pareto distribution to model shot lengths.

Unfortunately, previous studies are limited by the following facts: first, the number of video sequences for which the proposed shot length models were fit was generally too small to decide which distribution to use and to obtain accurate estimates for distribution parameters. Second, the video content used was not diverse enough, which makes it impossible to compare distributions from different program genres. Generally researchers have just used sequences containing movies or video conferences. Third, statistical analysis for the goodness of fit is lacking in some studies.[4,1]

In previous work on modeling shot lengths often one of the following distributions were used

Pareto distribution

$$\begin{aligned} f(x) &= \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \\ F(x) &= 1 - \left(\frac{\beta}{x}\right)^\alpha, \qquad \alpha, \beta > 0, x > \beta \end{aligned} \tag{1}$$

Gamma distribution

$$\begin{aligned} f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \\ F(x) &= \Gamma(\alpha, \beta x), \qquad \alpha, \beta > 0, x > 0 \end{aligned} \tag{2}$$

Weibull distribution

$$\begin{aligned} f(x) &= \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp(-(x/\beta)^\alpha) \\ F(x) &= 1 - \exp(-(x/\beta)^\alpha), \qquad \alpha, \beta > 0, x > 0 \end{aligned} \tag{3}$$

In this paper we will consider only these three distributions as possible distribution models to model shot lengths for broadcast video.

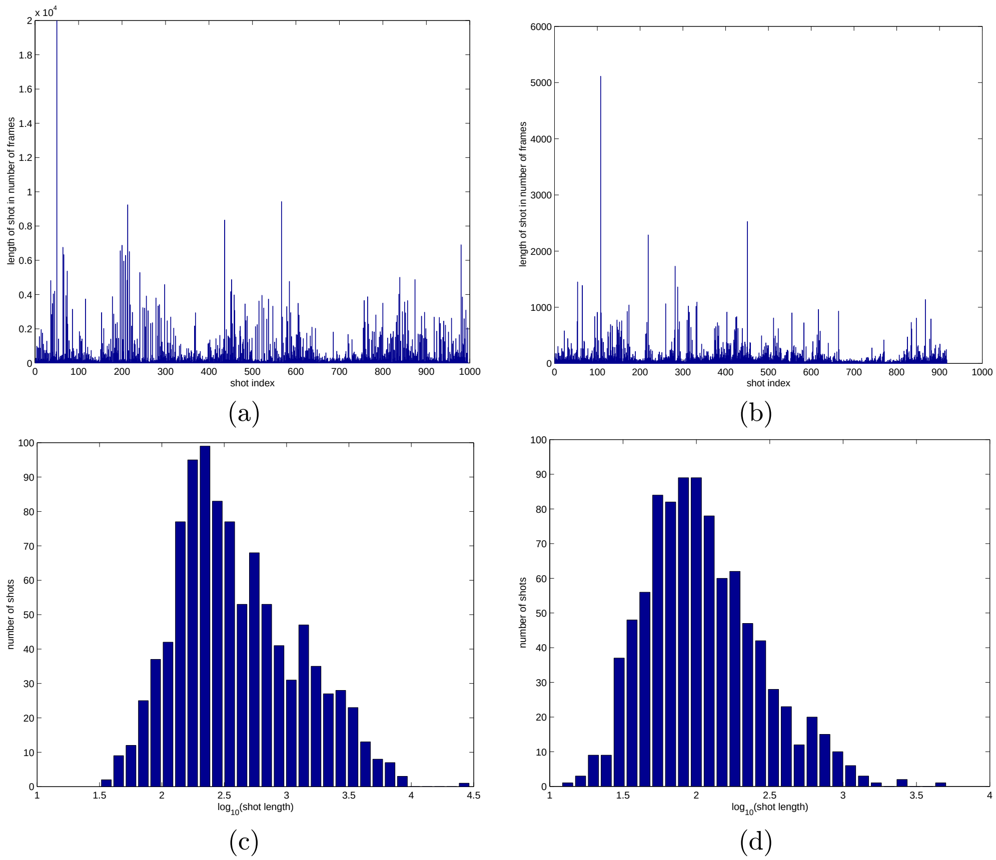

Figure 1. Plots (a) and (b) show the lengths of shots, for *cspan* and *movie*, respectively, in number of frames. Note different scales on the Y-axis for these plots. Plots (c) and (d) are the log histograms of shot lengths for *cspan* and *movie*, respectively.

1.2. The Experimental Data Set

We have used two data sets in our experiments which were obtained as follows: Incoming broadcast video was digitized at a rate of 1.5 Mb/sec in CIF format, i.e., 352×240, using MPEG-1 compression. Commercials in the programs, if they exist, were edited out. The locations of all the shot transitions in the sequences were recorded by a human operator. We grouped the shot length data from sequences into two sets:

- *cspan* (25 sequences, approximately 7.5 hours, 996 shots). Obtained from CSPAN and CSPAN-2 networks. Consists of House meetings, *Washington Journal*, *Booknotes*, and various other programs. CSPAN programs generally contain minimum editing. For example, our dataset includes a shot of approximately 16 minutes which contains only a static shot of a speaker.

- *movie* (7 sequences, approximately 1.5 hours, 917 shots). Obtained from various networks. The movies we used were *Lawrence of Arabia*, *Before He Wakes*, *Pretty Woman*, *Blues Brothers II*, *Witches Of Eastwick*, *Monday After The Miracle*, and *Groundhog Day*. The shot lengths distribution may differ from movie to movie and from director to director. In fact it has been shown that the shot length distribution is a good feature to identify the director.[12] However, different editing patterns for movies generally fall within a narrow range widely used norms.

We do not include more than one sequence from a given airing of a particular program, in order to achieve maximum content variation. Basic statistical properties of the two data sets are given in Figure 1 and Table 1.

Data	*cspan*	*movie*
n	996	917
min	32	12
median	332	98
mean	818.67	178.55
max	29703	5113
variance	2.16×10^6	7.78×10^4

Table 1. Basic statistics for the two data sets used in our experiments.

2. HEAVY-TAILED DISTRIBUTION FUNCTIONS

From the log histogram plots in Figure 1 we observe that the data is heavily skewed to the right. This implies that there is a non-negligable number of very long shots. This observation is verified by the data plots in Figure 1(a) and (b). Throughout this paper we will assume that the shot lengths are independent and identically distributed (iid).

Consider the Gaussian distribution, for which, as $x \to \infty$

$$P[X > x] \sim \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x} \to 0. \tag{4}$$

Such a distribution is said to possess light tails. In contrast, a random variable X is called *heavy-tailed* if, as $x \to \infty$, we have

$$P[X > x] \sim x^{-\alpha} L(x) \tag{5}$$

where $L(x)$ is a slowly varying function for which $\lim_{t\to\infty} L(tx)/L(x) = 1$. The variable α is called the *tail index* of the distribution. Heavy-tailed distributions belong to the family of *subexponential distributions*. An important property[13] of these distributions is that

$$\lim_{x\to\infty} \frac{P(X_1 + \cdots + X_n > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1$$

for every $n \geq 2$. This property implies that the largest value in $\{X_n\}$ exerts a strong influence on the total distribution. If X is a heavy-tailed random variable, for $X \geq 0$, the l^{th} moment of X, $E(X^l)$, exists only if $l < \alpha$. This property implies that if $\alpha < 2$, a heavy-tailed random variable has infinite variance; if $\alpha < 1$, then it also has infinite mean.

Heavy-tailed distributions include the Pareto, lognormal, and loggamma distributions and the Weibull distribution for $0 < \alpha < 1$. Light-tailed distributions include the normal, exponential, and gamma distributions and the Weibull distribution for $\alpha \geq 1$.

3. ESTIMATING DISTRIBUTION PARAMETERS

3.1. The Hill Plot

One way to estimate the tail index, α, is based on the following observation: Suppose $X_1, \ldots, X_n$ are iid random variables with distribution function F. Let $X_{1,n} \geq \cdots \geq X_{n,n}$ be the order statistics for this data. If X has a Pareto distribution with $\beta = 1$

$$P(X > x) = 1 - F(x) = x^{-\alpha}, \qquad x \geq 1 \tag{6}$$

then $Y = \log X$ will have an exponential distribution with distribution function

$$P(Y > y) = y^{-\alpha y}, \qquad y \geq 0$$

Since for this distribution the mean is α^{-1}, the maximum likelihood estimate (MLE) of α is given by

$$\hat{\alpha}_n^{-1} = H_n = \frac{1}{n} \sum_{i=1}^{n} \log X_{i,n}$$

We can generalize this estimation method for the case where X has a general heavy-tailed distribution

$$1 - F(x) = x^{-\alpha} L(x)$$

by assuming that above a certain threshold $1 - F(x)$ behaves like the Pareto distribution function in Equation 6. This means we should base our estimate of α on the part of the distribution that looks most Pareto-like. Based on these arguments we can define the Hill estimator to be

$$\hat{\alpha}_{k,n}^{-1} = H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{i,n}}{X_{k+1,n}} \tag{7}$$

Note that only k upper statistics are used in the estimation. The Hill estimator is asymptotically normal and it can be shown that as $n \to \infty$ and $k \to \infty$ but $k/n \to 0$, we have $H_{k,n} \to \alpha^{-1}$ in probability. In practice the Hill estimator is used by plotting

$$\text{Hill Plot:} \left\{ (k, H_{k,n}^{-1}),\ 1 \leq k < n \right\} \tag{8}$$

One should bear in mind that the Hill estimator has optimality properties only when the underlying distribution is close to a Pareto distribution. If the distribution is far from Pareto, there may be large bias even for large sample sizes. Hence it is best to use the Hill plot with other analysis techniques and compare results. The Hill plots for the datasets are shown in Figure 2(a) and (b). These plots suggest the tail index values $\alpha_{\text{cspan}} \approx 0.8.$ and $\alpha_{\text{movie}} \approx 1.2$

3.2. The Moment Estimator

Another estimate for α may be obtained using the moment estimator proposed in.[14] This method is designed to estimate the *extreme value index*, γ, where we have $\gamma = 1/\alpha$. For this method we first define the quantities

$$H_{k,n}^{(r)} = \frac{1}{k} \sum_{i=1}^{k} \left(\log \frac{X_{i,n}}{X_{k+1,n}} \right)^{r}$$

which for $r = 1$ is the Hill estimator described in Section 3.1. Then the estimate for γ is given by

$$\hat{\gamma}_{k,n} = H_{k,n}^{(1)} + 1 - \frac{1/2}{1 - (H_{k,n}^{(1)})^2 / H_{k,n}^{(2)}} \tag{9}$$

The quantity $\hat{\gamma}_{k,n}$ is also very useful in deciding if a given data sample is drawn from a heavy-tailed distribution or not. If $\hat{\gamma}_{k,n}$ is negative or close to zero, then most probably the underlying distribution is not heavy-tailed. The use of the moment estimator to estimate the value of γ and decide if the underlying distribution is heavy-tailed is similar to the Hill plot. We plot the graph

$$\text{Moment Plot:} \{ (k, \hat{\gamma}_{k,n}),\ 1 \leq k < n \} \tag{10}$$

to obtain an estimate of γ. The moment estimator plots are shown in Figure 2(c) and (d). These plots suggest the tail index values $\gamma_{\text{cspan}} \approx 0.7$ and $\gamma_{\text{movie}} \approx 0.6$ which correspond to the values $\alpha_{\text{cspan}} \approx 1.4$ and $\alpha_{\text{movie}} \approx 1.7$. From these plots we can also conclude that heavy-tailed distribution is an appropriate model for these shot distributions.

3.3. The Mean Excess Function

Let X be a random variable with a right endpoint x_F. The function defined by

$$e(u) = E(X - u | X > u),\ 0 \leq u < x_F) \tag{11}$$

is called the *mean excess function* of X. The mean excess function provides another useful graphical tool, especially for discrimination in the distribution tails. The mean excess function of a heavy-tailed distribution function, for large values of u, typically will lie between a constant function, corresponding to the exponential distribution, and a straight line with positive slope, corresponding to the Pareto distribution. Any continuous distribution function F is uniquely determined by its mean excess function.[13]

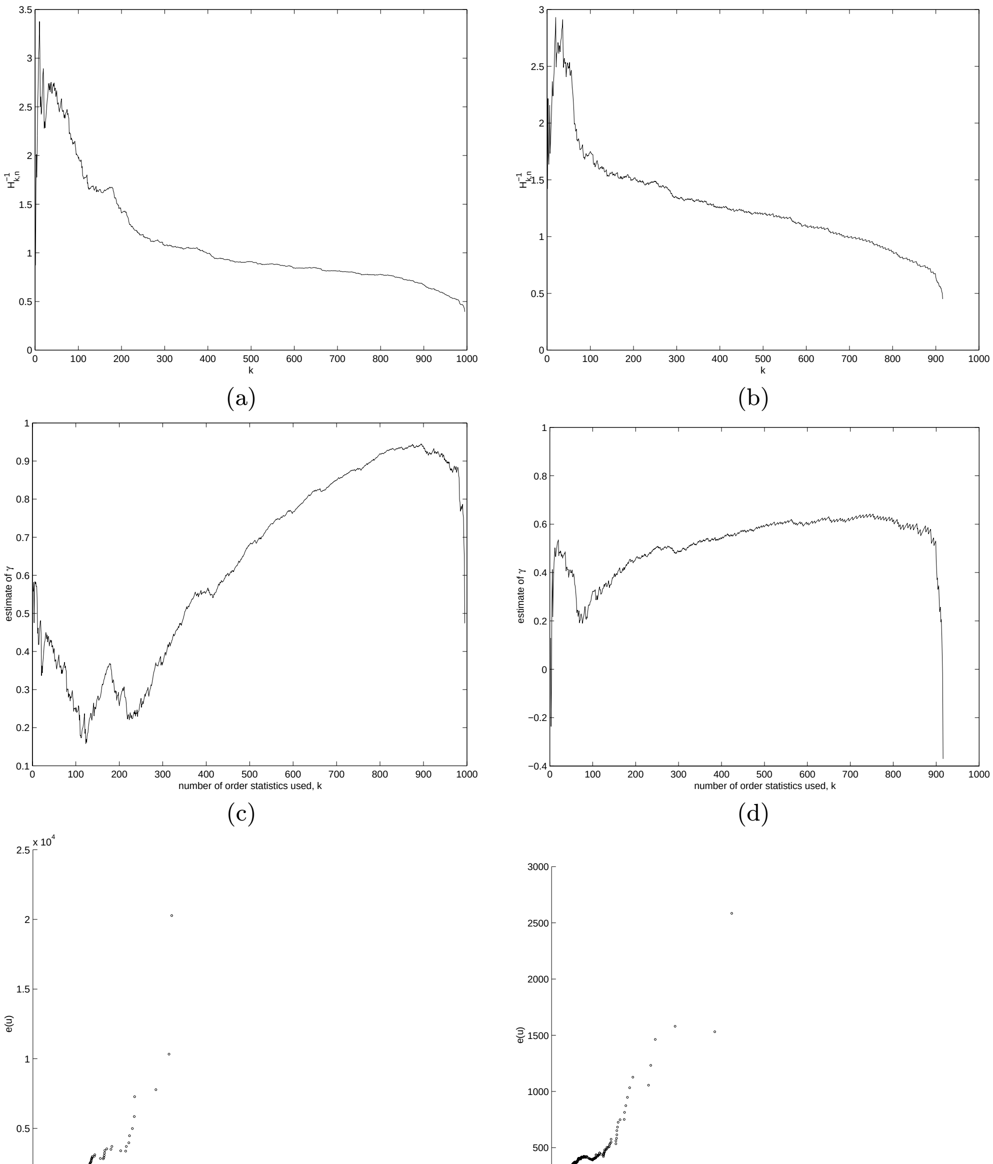

Figure 2. Plots (a), (c), and (e) are the Hill, moment estimator, and mean excess function plots for *cspan*. Plots (b), (d), and (f) are the same plots for *movie*.

Model	mean	variance
Pareto	$\alpha\beta/(\alpha-1)$	$\alpha\beta^2/(\alpha-1)(\alpha-2)$
Gamma	α/β	α/β^2
Weibull	$\beta\Gamma(1+1/\alpha)$	$\beta^2[\Gamma(1+2/\alpha)-\Gamma^2(1+2/\alpha)]$

Table 2. Formulas for the mean and variance for the three distributions considered as possible models. Note that for the Pareto distribution the mean exists only if $\alpha > 1$ and the variance exists only if $\alpha > 2$.

Suppose that $X_1, \ldots, X_n$ are iid and define $\Delta_n(u) = \{i : X_i > u\}$, that is, $\Delta_n(u)$ is the set of samples in $\{X_n\}$ with values larger than the threshold u. Then we define the empirical mean excess function, $e_n(u)$, to be

$$e_n(u) = \frac{1}{\mathrm{card}\Delta_n(u)} \sum_{i \in \Delta_n(u)} (X_i - u), \qquad u \geq 0 \tag{12}$$

where we use the convention $0/0 = 0$. The mean-excess plot is then obtained from the graph

$$\text{ME Plot:} \{(X_{k,n}, e_n(X_{k,n})),\ 1 \leq k < n\} \tag{13}$$

where $X_{1,n} \geq \cdots \geq X_{n,n}$ are the order statistics. The mean excess plots for the two datasets are shown in Figure 2(e) and (f). Since the graphs are linear we have further proof that the underlying distributions are heavy-tailed. Since linear mean excess functions correspond to the Pareto distribution, we also conclude that the Pareto distribution is a good candidate to model the given data.

4. COMPARING DISTRIBUTIONS

In this section we examine the problem of choosing the best distribution for our shot length data. Formulas for the means and variances for the three distributions are given in Table 2. We shall use some of these formulas to estimate distribution parameters using the method of moments.

For the gamma distribution equating the first moments of the data, we obtain

$$\hat{\alpha}_g = \frac{\hat{\mu}^2}{\hat{\sigma}^2}, \qquad \hat{\beta}_g = \frac{\hat{\alpha}_g}{\hat{\mu}}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimates for the mean and variance, respectively.

For the Pareto distribution the maximum likelihood estimate of β is $\min\{X_n\}$. We then either obtain an estimate of α from the Hill and moment estimator plots or check to see if $\alpha > 1$ from these plots and use the mean formula to obtain

$$\hat{\alpha}_p = \frac{\hat{\mu}}{\hat{\mu} - \hat{\beta}_p}, \qquad \hat{\beta}_p = \min\{X_n\}, \qquad \text{if } \alpha > 1$$

For the Weibull distribution there is no closed form expression for the method of moments estimators for the parameters. Therefore, for this distribution parameters have to be estimated by plotting the equation

$$\ln\left(\ln\left(\frac{1}{1-F(x)}\right)\right) = \alpha \ln x - \alpha \ln \beta$$

which is obtained by manipulating the expression given in Equation 3 and fitting a line through the data points to estimate $\hat{\alpha}_w$ and $\hat{\beta}_w$.

Once the parameters of the distributions are estimated, we measure the goodness of fit of the proposed distribution to the empirical distribution function using the Kolmogorov-Smirnov statistic. The K-S statistic is defined to be[15]

$$D = \max_{-\infty < x < \infty} |F_n(x) - F(x)| \tag{14}$$

Data	*cspan* α	β	*movie* α	β
Pareto	0.3	32	0.40	12
	$D = 0.2148$		$D = 0.2732$	
Gamma	0.31	3.79×10^{-4}	0.41	23.0×10^{-4}
	$D = 0.3282$		$D = 0.3319$	
Weibull	1.05	697.00	1.32	171.02
	$D = 0.1440$		$D = 0.1287$	

Table 3. Estimated parameter values for the three distribution models. For each model the K-S statistic, D, is also given.

where $F_n(x)$ is the empirical cumulative distribution function estimated from data using the formula

$$F_n(x) = \frac{n - k + 1}{n + 1}.$$

The values of the estimated parameters and the corresponding values of the K-S statistic are tabulated in Table 3. From this table we see that for our data set the Weibull distribution gives the best fit followed closely by the Pareto distribution. However, among the three models the Pareto distribution provides the best fit for the right tail of the shot length distribution. Therefore, it may be preferred when it is important to characterize the distribution of very long shots.

5. CONCLUSIONS

In this paper we have addressed the problem of choosing a good probabilistic model for the duration of shots for broadcast video. Accurate models for shot lengths are important to model video both for content-based retrieval applications and for performing queuing analysis for the design of video buffers in multimedia networks. Using a large dataset collected from CSPAN programs and movies we have illustrated various exploratory data analysis techniques. We have shown that the shot length distributions possess heavy tails which is a very important fact for queuing analysis. Using various graphical tools we have shown how the parameters of the distributions may be estimated. Finally we have compared the three possible distribution models using the Kolmogorov-Smirnov statistic.

REFERENCES

1. N. Vasconcelos and A. Lippman, “A bayesian video modeling framework for shot segmentation and content characterization,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 97)*, October 4-7 1997, San Juan, Puerto Rico.
2. D. Stan and I. K. Sethi, “Mapping low-level image features to semantic concepts,” *Proceedings of SPIE Conference on Storage and Retrieval for Media Databases 2001*, January 2001, San Jose, CA, pp. 172–179.
3. C. M. Taskiran, C. A. Bouman, and E. J. Delp, “Discovering video structure using the pseudo-semantic trace,” *Proceedings of SPIE Conference on Storage and Retrieval for Media Databases 2001*, January 2001, San Jose, CA, pp. 571–578.
4. M. R. Frater, J. F. Arnold, and P. Tan, “A new statistical model for traffic generated by VBR coders for television on the broadband ISDN,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 6, pp. 521–526, December 1994.
5. U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, “Segmenting full-length VBR video into shots for modeling with markov-modulated gamma-based framework,” *SPIE Symposium on the Convergence of Information Technologies and Communications (ITCom 2001)*, 19-24 August 2001, Denver, CO.

6. S. I. Resnick, "Heavy tail modeling and teletraffic data," *The Annals of Statistics*, vol. 25, no. 5, pp. 1805–1869, 1997.
7. M. E. Crovella, M. S. Taqqu, and A. Bestavros, "Heavy-tailed probability distributions in the World Wide Web," in *A Practical Guide to Heavy Tails* (R. J. Adler, R. E. Feldman, and M. S. Taqqu, eds.), pp. 3–25, Birkhauser, 1998.
8. M. Krunz and A. M. Ramasamy, "The correlation structure for a class of scene-based video models and its impact on the dimensioning of video buffers," *IEEE Transactions on Multimedia*, vol. 2, no. 1, pp. 27–36, March 2000.
9. D. P. Heyman and T. V. Lakshman, "Long-range dependence and queueing effects for VBR video," in *Self-similar Network Traffic and Performance Evaluation* (K. Park and W. Willinger, eds.), pp. 285–318, John Wiley and Sons, Inc., 2000.
10. A. M. Dawood and M. Ghanbari, "Content-based MPEG video traffic modeling," *IEEE Transaction on Multimedia*, vol. 1, no. 1, pp. 77–87, March 1999.
11. P. R. Jelenkovic, A. A. Lazar, and N. Semret, "The effect of multiple time scales and subexponentiality in MPEG video streams on queing behavior," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1052–1071, August 1997.
12. S.-F. Chang, J. R. Smith, M. Beigi, and A. Benitez, "Visual information retrieval from large distributed online repositories," *Communications of the ACM*, vol. 40, no. 12, pp. 63–71, 1997.
13. P. Embrechts, C. Kluppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance.* Berlin, Germany: Springer, 1997.
14. A. Dekkers and L. D. Haan, "On the estimation of the extreme value index and large quantile estimation," *Annals of Statistics*, vol. 17, pp. 1883–1855, December 1989.
15. W. Press, B. Flannery, S. Teukolsky, and Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge University Press, 1988.