A Robust Estimator of Image Thumbnail and Video Histogram Representation

Cheng Cai
[1] Northwest A&F University
No.3 Taicheng Road, Yangling, Shaanxi, China, 712100
cheney.chengcai@gmail.com

ABSTRACT

For the browsing and retrieval system, images are represented by thumbnails and video shots are represented by histogram representations. In order to achieve better visual quality and retrieval performance, the representation estimator is expected to be accurate and robust. From the statistical perspective, representation extraction can be treated as central value estimation. In this paper, we propose an adaptive alpha-trimmed average estimator based on the Gaussian distribution hypothesis test. For a set of values, this estimator extracts the representation by trimming extreme values and then averaging the rest. The criterion adopted to distinguish between extreme values and useful data is derived from the Gaussian distribution hypothesis test on the basis of global statics. Experimental results from standard images and videos show that our proposed scheme outperforms traditional methods.

General Terms

Algorithms, Experimentation, Performance

Keywords

Adaptive alpha-trimmed average, image thumbnail, video shot histogram representation, Gaussian distribution hypothesis test.

1. INTRODUCTION

For a browsing and retrieval system, the extraction of effective representation of a set of samples is one of the key issues because of limited storage of index structure [1].

1.1 Image Thumbnail

In order to display more images in the limited screen space, original images are usually represented by their thumbnails. As shown in Figure 1, an original image is divided into $D\times D$ non-overlap blocks, and each block is then represented by a value, which means the size of original image is represented by its $1/(D\times D)$ size thumbnail.

Contextual and Social Media Understanding and Usage, Seminar 08251, Schloss Dagstuhl, Leibniz-Zentrum für Informatik, Germany, June 15-20, 2008.

1.2 Video Histogram Representation

Because the size of video databases increases, it becomes necessary to represent groups of frames or shots in a video with effective features. A reliable histogram-based content representation should consider the content of all the frames with a shot for computation [2]. As shown in Figure 2, a video shot is a set of frames. The single representation for the entire collection is obtained by combining individual frame histograms [3].

2. REPRESENTATION ESTIMATOR

For thumbnail extraction and video histogram presentation, average is a commonly used as an estimator [3, 4]. The reason is that average is optimal according to the least squared error. And it is also an unbiased estimator for Gaussian distributed samples. However, average might be inadequate in high noise situations [5]. To suppress noise, median value is employed instead as the representation. Median value has the least probability to be corrupted by impulse noise because impulses typically occur near the end of the sorted samples. However, because median value may differ significantly from the noise-free value, it may not be the optimal value to represent samples [6]. To compromise between average and median, the alpha-trimmed average (ATA) estimator is widely used [2]. The key design issue of ATA estimator is to select its only parameter α. In this paper, we propose a new adaptive alpha-trimmed average estimator based on a Gaussian distribution hypothesis test (AATA-GDHT) for content representation. This estimator intends to keep as many useful samples as possible as well as removing extreme values.

2.1 Alpha-trimmed Average

Let $\left\{x_n, 0 \le n \le N-1\right\}$ denotes a set of N samples, and $\left\{s_n, 0 \le n \le N-1\right\}$ the sorted samples in descending order where $s_k \in \{x_n\}$, $s_k \le s_{k+1}$ and $0 \le k \le N-1$. The first step of alpha-trimmed average is to trim the most extreme values, both low and high with an equal percentage of samples $\{t_k\} = T\left(\alpha, \{s_n\}\right) = \{s_n, \lfloor \alpha N \rfloor \le n \le N-1-\lfloor \alpha N \rfloor\}$, $0 \le k \le N-2\lfloor \alpha N \rfloor$, where $0\le\alpha<0.5$ controls the number of samples excluded from the average computation, $\lfloor \cdot \rfloor$ denotes the greatest integer part and the trimming percentage parameter α, and $T(\alpha,\{s_n\})$ is the sample set left after being trimmed..

The alpha-trimmed average then averages the remainders to estimate the central value:

$$ATA(\alpha,\{s_n\}) = \frac{1}{N-2\lfloor \alpha N \rfloor}\sum_{n=\lfloor \alpha N \rfloor}^{N-1-\lfloor \alpha N \rfloor} s_n = \frac{1}{N-2\lfloor \alpha N \rfloor}\sum_{k=0}^{N-1-2\lfloor \alpha N \rfloor} t_k \quad (1)$$

where $ATA(\alpha,\cdot)$ denotes the alpha-trimmed average operator. As trimming percentage parameter $0 \le \alpha < 0.5$, the ATA performs like an average when α is close to 0, and like a median operator when α is close to 0.5. How to select a proper parameter α is the key design issue of ATA.

2.2 Adaptive Alpha-trimmed Average

The alpha-trimmed average intends to be used for central value estimation when the Gaussian distributed data contains some outliers. Let $\{s_n, 0 \le n \le N-1\}$ denote a set of sorted Gaussian distributed samples with some outliers. With a given trimming percentage parameter α, the remaining sample set left after being trimmed is $T(\alpha,\{s_n\})$. The optimal α should guarantee to trim out the outliers as well as to keep as many useful samples as possible.

$$\alpha_{opt} = \min\left\{\alpha : T(\alpha,\{s_n\}) \text{ are Gaussian distributed}\right\} \quad (2)$$

Jarque-Bera criterion is selected to test Gaussian distribution hypothesis because it is sensitive to outliers, which means that a set of Gaussian distributed samples with outliers will fail in this test. The Jarque-Bera criterion is defined as follows:

$$J(\alpha,\{s_n\}) = \frac{N(\alpha)}{6}\left\{\frac{S^2(T(\alpha,\{s_n\}))}{\sigma^6(T(\alpha,\{s_n\}))} + \frac{1}{4}\left[\frac{K(T(\alpha,\{s_n\}))}{\sigma^4(T(\alpha,\{s_n\}))} - 3\right]^2\right\} \quad (3)$$

where $N(\alpha) = N - 2\lfloor \alpha N \rfloor$, σ, S and K are standard deviation, skewness, and kurtosis respectively [1]. $J(\alpha,\{s_n\})$ satisfies Chi-square distribution. Suppose the significance level is 0.05. The optimal trimming percentage then is:

$$\alpha_{opt} = \min\left\{\alpha \mid \chi_2^2(0.025) \le J(\alpha,\{s_n\}) \le \chi_2^2(0.975)\right\} \quad (4)$$

where $\chi_v^2(\gamma)$ is:

$$\gamma = \int_0^{\chi_v^2(\gamma)} \frac{x^{(v-2)/2} e^{-t/2}}{2^{v/2}\Gamma(v/2)} dt \quad (5)$$

3. EXPERIMENTS

3.1 Image Thumbnail

The standard 8-bit, gray-scale, "Lena" with resolution 512 × 512 is used as an example to test the representation estimation performance of our adaptive alpha-trimmed average. The original image is degraded with Gaussian noise (the noise density is 0.005) and impulse noise (the noise density is 0.05). As illustrated in Fig. 1, the 512 × 512 image is divided into 8 × 8 non-overlap blocks, and each block is then represented by a value. Compared with the adaptive alpha-trimmed average based on nonsymmetrical detection [7], the peak signal-to-noise ratios (PSNR) is improved with 0.2 dB [1]. This indicates that our proposed estimator can achieve a more accurate and more robust representation.

3.2 Video Shot Histogram Representation

For a video shot totally contains N frames, let $h_n(b)$, $1 \le b \le B$ denote the histogram of the n-th frame, where b denotes the bin number and B denotes the total number of quantization bin. Let bin values of each frame histogram form a set of samples denoted as $\{h_n(b)\}$, $0 \le n \le N-1$. The corresponding optimal trimming percentage parameter of each bin is denoted as $\alpha_{opt}(b)$ according to the Jarque-Bera criterion:

$$\alpha_{opt}(b) = \min\left\{\alpha \mid \chi_2^2(0.025) \le J(\alpha,\{h_n(b)\}) \le \chi_2^2(0.975)\right\} \quad (6)$$

The representation of each bin is extracted as follows:

$$h_{opt}(b) = ATA\left(\alpha_{opt}(b),\{h_n(b)\}\right),\ 1 \le b \le B. \quad (7)$$

Our experiments are conducted on the basis of 427 shots extracted from the three video sequences of the MPEG-7 content set. The shot boundaries have been identified manually, and query samples are selected as the first, middle and last frames of each shot. To evaluate the representation estimator performance, the average recall (AR) [4] and the average normalized modified retrieval rank (ANMRR) [4] are adopted. The 8-bit gray-level alpha-trimmed average histograms with fixed parameters $\alpha = \{0, 0.1, 0.15, 0.20, 0.25, 0.50\}$, and adaptive parameters according to the Gaussian distribution hypothesis test are shown in Table I. From Table I, both AR and ANMRR are improved, which shows that our proposed algorithm can estimate a more accurate and robust representation.

Table 1. The performances of the alpha-trimmed average histograms with α = {0, 0.1, 0.15, 0.2, 0.25, 0.5}, and the adaptive optimal alpha-trimmed average histogram.

Algorithm	Performance	
	ANMRR	***AR***
0 alpha-trimmed average histogram	0.4520	0.5885
0.10 alpha-trimmed average histogram	0.4464	0.5942
0.15 alpha-trimmed average histogram	0.4451	0.5950
0.20 alpha-trimmed average histogram	0.4462	0.5944
0.25 alpha-trimmed average histogram	0.4468	0.5957
0.5 alpha-trimmed average histogram	0.4453	0.5968
adaptive optimal alpha-trimmed average histogram	0.3021	0.7372

4. CONCLUSIONS

We have proposed a new adaptive alpha-trimmed average for image thumbnail and video shot histogram representation by analyzing the global statistics derived from the Gaussian distribution hypothesis test. Experimental results from standard image and videos show that our algorithm outperforms traditional methods. The drawback of our proposed algorithm is its heavy computational complexity. The future research direction is to find a faster algorithm to reduce this weakness.

5. ACKNOWLEDGMENTS

This is joint work with Dr. Kin-Man Lam from the Multimedia Center, Electronics and Information Engineering Department in the Hong Kong Polytechnic University.

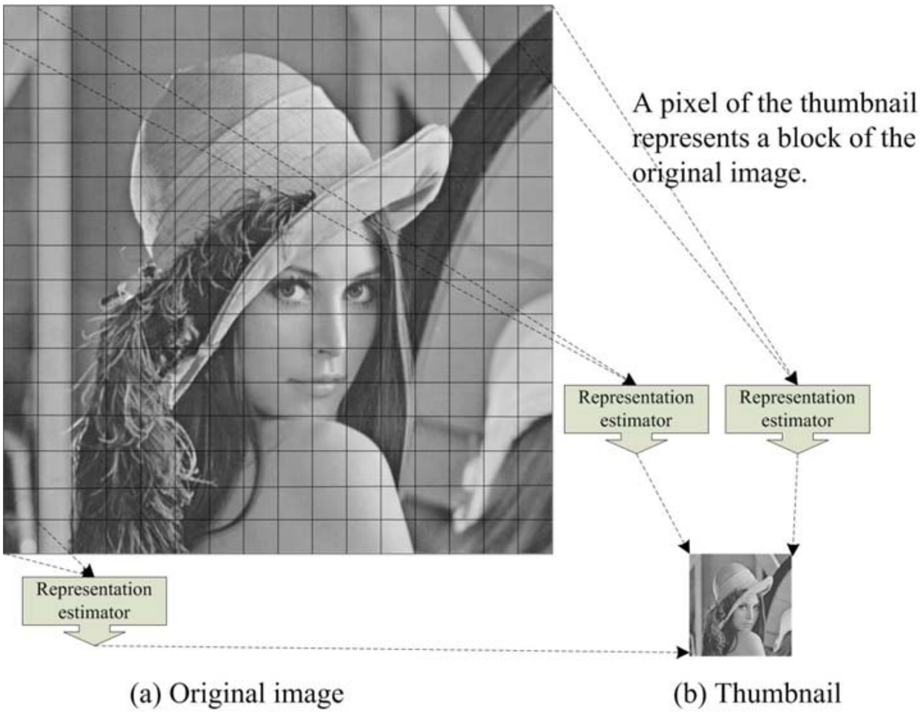

Figure 1. Image thumbnail representation.

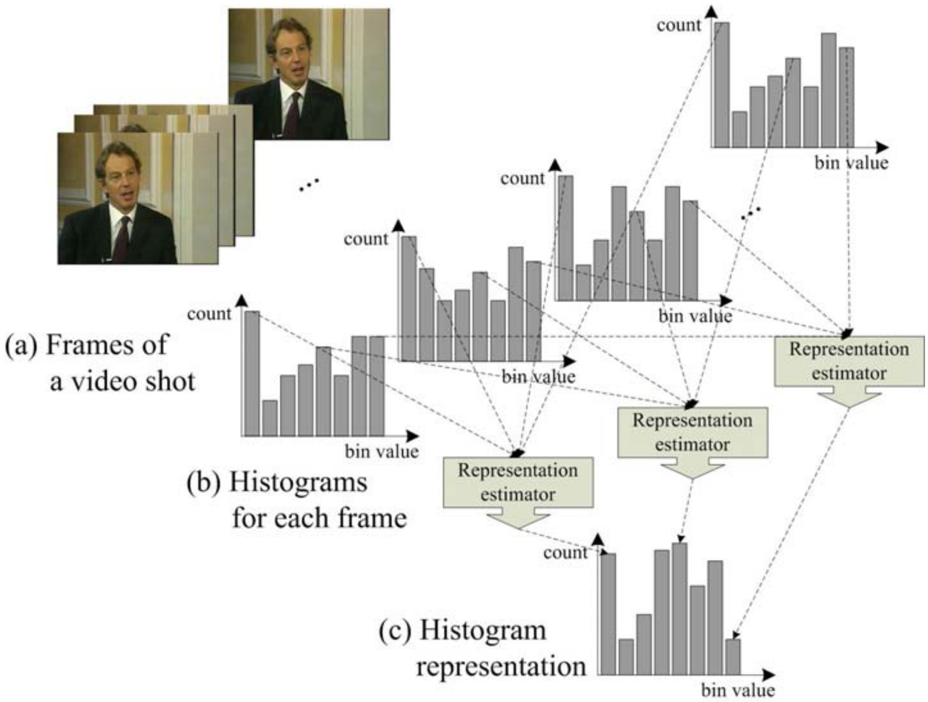

Figure 2. Video shot histogram representation.

6. REFERENCES

[1] Cai, C., Lam K.-M., and Tan, Z. 2007. Adaptive alpha-trimmed average operator based on Gaussian distribution hypothesis test for image representation. In Proceedings of the 4th International Conference on Image and Graphics (Chengdu, China), IEEE Computer Society

[2] Ferman, A. M., Tekalp A. M., and Mehrotra, R. 2002. Robust color histogram descriptors for video segment retrieval and identification. IEEE Trans. Image Processing. 11, 5 (May. 2002), 497-508.

[3] Ferman, A. M., Krishnamachari, S., Tekalp, A. M., Abdel-Mottaleb, M., and Mehrotra, R. 2000. Group-of-frames/pictures color histogram descriptors for multimedia applications. In Proceedings of the International Conference on Image Processing.

[4] Sze, K.-W., Lam, K.-M., and Qiu, G. 2005. A new key frame representation for video segment retrieval. IEEE Trans. Circuits Syst. Video Technol. 15, 9 (Sept. 2005), 1148-1155.

[5] Bednar, J. and Watt, T. 1984. Alpha-trimmed means and their relationship to median filters. IEEE Trans. Acoust., Speech, Signal Processing. 32, 1 (Feb. 1984) 145-153.

[6] Zhang,D., Zhuoer,S., Haixiang, W., Kouri, D. J., and Hoffman, D. K. 2000. Nonlinear filtering impulse noise removal from corrupted images. In Proceedings of the International Conference on Image Processing. 285-287.

[7] Oten, R. and Figueiredo, R. J. P. de "Adaptive alpha-trimmed mean filters under deviations from assumed noise model," *IEEE Trans. Image Process.*, vol. 13, pp. 627-639, 2004.