BOOSTING QUANTIZATION FOR L_P NORM DISTORTION MEASURE

Liming Wang[1] *Nicola Piotto*[2] *Dan Schonfeld*[3]

[1]Dept. of Electrical Engineering, Columbia University
500 W. 120th Street, New York, NY 10027
[2]Dept. of Engineering & Computer Science, University of Trento
via Sommarive 14, 38100, Povo (Trento), Italy
[3]Dept. of Electrical & Computer Engineering, University of Illinois at Chicago
851 S. Morgan Street, Chicago, IL 60607

ABSTRACT

Quantization is a widely used technique in signal processing. The purpose of many quantization schemes is to faithfully reproduce the input signals. However, in many situations, one is more interested in comparison of different classes of signals in order to classify them into different categories. The classification criterion is based on comparing distances under certain metric. For classical quantizers, although the individual quantized signal may show high fidelity to its original signals, the distance features characterizing different categories may not be well reserved, which results in poor performance of classification in spite of relatively good reproduction of individual signals. In this paper, we propose a special optimal quantization under the L_p norm distortion measure called Boosting Quantization. The quantization is guaranteed to preserve the distances of different classes. We provide a quantization algorithm to generate the quantizer. We also derive several theoretical properties for the proposed quantization. Finally, we provide numerical examples to illustrate our proposed boosting quantization.

***Index Terms*—** Quantization, classification.

1. INTRODUCTION

Quantization is a widely used technique and an important intermediate step in many signal processing systems [1]. It can be seen as a function which maps a continuous input signal to a discrete signal with finite number of representation levels. The primary goal in classical quantization is to faithfully reproduce the input signal, i.e. the quantized signal should preserve the fidelity to the original signal. Lloyd and Max proposed the optimal quantizer [2, 3] under mean square error measurement. Linde, Buzo and Grey proposed the vector quantizer [4]. A wide variety of vector quantizer algorithms have been developed for speech, images, video, and other signal sources. For example, in video trajectory analysis, the quantization technique is adopted in order to bring the observable information from a continuos to a symbolic level: after an initial stage of feature extraction, some signal quantization and symbolic coding is typically computed. Once the quantized discrete data is available, the similarity between sequences is evaluated adopting, among others, simple metrics (e.g., L_p norms, Euler or Hausdorff distances, or string alignment algorithms) [5]. The authors of [6] present a complete framework for motion-based video retrieval in which they adopt the minimum cumulative square distance as a metric.

A parallel problem is the quantization for a noisy source [7]. The problem can be viewed as trying to compress a dirty source into a clean reproduction, or as an estimation of the original source based on a quantized version of the noisy signal. The process is equivalent to a de-noising filter. Many of these quantizers can be categorized as the Rate-distortion quantizers [1].

In many situations, one is often more interested in comparison and classification of different classes of signals than processing one individual signal class. Meanwhile, due to the limit of storage, calculation and power, a quantization has to be applied. In this case, we naturally desire the distances among different classes to be preserved, or emphasized after quantization. The classical goal of quantization which aims to preserve fidelity to the original signal may not be effective especially when we have restrictions on the number of quantization levels.

In this paper, we provide a mathematical formulation for the optimal quantization of degraded signals in order to preserve L_p distance among different classes, which is called the *Boosting quantization*. As we shall see later, the reason of the naming is from the fact that this particular quantization will not only save time and storage complexity, but also preserve or enhance the fidelity under the L_p norm criterion. We present a quantization algorithm to find the optimal quantization given the statistical information of the signals. We also show several theoretical properties of the boosting quantization. We provide several numerical examples for illustrations and justifications of the proposed algorithm.

2. BOOSTING QUANTIZATION PROBLEM AND ITS QUANTIZATION ALGORITHM

2.1. Problem Formulation

Consider two real signals $\{X_1(t)\}$ and $\{X_2(t)\}$ as discrete-time random processes, where t is the time index. We model the observed signals $Y_i(t)$ as $Y_i(t) = X_i(t) + N_i(t)$ *for* $i = 1, 2$, where $N_i(t)$ is the noise. Let $Q_t(x)$ denote a quantization scheme at each time instant t. $Q_t(x)$ is specified by a set of region boundaries $\{(\mu_l, \mu_{l+1}]\}_l$, and a set of reconstruction values $\{q_l\}_l$, with $l = 1, \ldots, k$. k is the number of quantization levels. In the case of L_p fidelity criterion, we propose our boosting quantizer as the optimal solution of the following optimization problem:

$$\max \sum_t \mathbb{E}(\|QD(t)\|_p - \|ND(t)\|_p) \tag{1}$$

where

$$\begin{aligned} ND(t) = &\quad \|Y_1(t) - Y_2(t)\|_p - \|X_1(t) - X_2(t)\|_p \\ QD(t) = &\quad \|Q_t(Y_1(t)) - Q_t(Y_2(t))\|_p - \|X_1(t) - X_2(t)\|_p \end{aligned}$$

In particular, the terms $\|ND(t)\|_p$ and $\|QD(t)\|_p$ correspond to the fidelity of distances under L_p of the observed and quantized signals to the original signals, respectively. The intuition is that the quantized signals should emphasize or at least maintain the differences among different signal classes than the observed ones. In the paper, we denote the objective function in Eq. (1) given the quantization Q as $f(Q)$ or simply f. If $f(Q) > 0$ for some optimal quantization scheme Q, then it is worthwhile to employ that quantization Q. In that case, we call the optimal quantization scheme as the *Boosting quantization* since the quantization not only quantizes the signal but also boosts the characterizing distance features for different classes of signals. The final goal is to find the quantization scheme optimizing the problem in (1).

Notice that it is enough to consider the quantization scheme separately for each time instant. We will only consider the optimization in (1) at one time instant for the rest of paper.

2.2. Boosting Quantization Algorithm

In order to evaluate the optimal quantization scheme, the optimization problem of Eq. (1) is solved with respect to the parameters of Q_k, i.e., $(\mu_l, \mu_{l+1}]$ and q_l with $l = 1, \ldots, k$. We propose the following quantization algorithm. Unfortunately, since the optimization process involves the maximization of a non-convex function, the algorithm may not converge to the global optimal quantization: it only guarantees to converge to a local optimal point. Since a closed form for the final solution is not available, a gradient descent with line search is used.

If the noise-free signal and noise statistics are known, the ND term in Eq. (1) can be considered constant. Thus, the initial optimization problem becomes

$$\max \mathbb{E}(\|(\|Q(y_1) - Q(y_2)\|_p - \|x_1 - x_2\|_p)\|_p) \tag{2}$$

The optimization is initialized with a uniform quantization scheme, so that the regions are defined by equally spaced μ_l, and the reconstruction levels q_{l+1} are the centroids of each region. Our problem involves a multivariable optimization (i.e., $2k + 1$ variables, k number of quantization levels) which is solved by computing an iterative single-variable optimization. In other words, the optimization is solved for each single variable, and considering all the others fixed: when optimizing the first region boundary μ_1, all the boundaries μ_l with $l = 2, \ldots, k$ and the reconstruction levels q_m with $m = 1, \ldots, k + 1$ are kept fixed. The iteration process over the considered $2k + 1$ variables is stopped when a predefined decreasing rate is achieved between successive function evaluations. The pseudo-code of proposed algorithm is listed in Algorithm 1. Here, the quantities $\Phi_i(i)$, $i = X_1, X_2, N_1, N_2$ are the *pdf* functions of the noise-free signals and the corrupting noises, respectively.

Algorithm 1 Pseudo code for the optimization

Require: $\Phi_{X_1}(X_1)$, $\Phi_{X_2}(X_2)$, $\Phi_{N_1}(N_1)$, $\Phi_{N_2}(N_2)$, k, $Stop$
Ensure: Optimal μ_i, q_j with $i = 1, \ldots, k-1$, $j = 1, \ldots, k$
  $OldVal = 0$, $NewVal = 0$
  Uniform quantizer variables initialization
  while $100 \times \frac{Oldval - NewVal}{Oldval} \geq STOP$ **do**
    $OldVal \leftarrow NewVal$
    for $i = 1$ to k-1 **do**
      Fix all μ_l with $l = 1, \ldots, k-1 \wedge l \neq i$
      Fix all q_m with $m = 1..k$
      Find μ_i maximizing Eq. (2)
    end for
    for $j = 1$ to k **do**
      Fix all μ_l with $l = 1, \ldots, k-1$
      Fix all q_m with $m = 1, \ldots, k \wedge m \neq j$
      Find q_j maximizing Eq. (2)
    end for
    $NewVal \leftarrow$ Maximal value for Eq. (2) for this iteration
  end while

3. THEORETICAL PROPERTIES OF BOOSTING QUANTIZATION

As for the theoretical properties of boosting quantization, we are interested in the following problems in proposed boosting quantization:

1. When should we use the boosting quantization? i.e., is it always worthwhile to use a boosting quantization scheme?

2. Does the performance always increase with the increasing of the quantization levels?

In this section, we present answers to these problems by providing theoretical properties of boosting quantization. First of all, the following theorem provides a sufficient condition for applicability of boosting quantization.

Theorem 1. *If the second or fourth order moment of noise is sufficiently large, then we have $f(Q) > 0$ for some Q, i.e., it is worthwhile to apply the boosting quantization.*

In particular, such bound is determined by the noise properties only. In other words, if the variance, i.e., the power of noise is large enough, the fidelity of differences among quantized signals will be higher than the unquantized version.

Concerning the role of number of the quantization levels, we show by following theorem that under the assumption that all probability distribution functions are strictly increasing, then $f(Q)$ is strictly increasing with the number of quantization levels as expected.

Theorem 2. *Let k_1 and k_2 denote the number of quantization levels of two quantization scheme Q_1 and Q_2 respectively. If $k_2 > k_1$, for any optimal quantization Q_1, there exists a quantization Q_2 such that $f(Q_2) > f(Q_1)$.*

Although we have shown that the performance of boosting quantization is increasing with the number of quantization levels, the robustness of the boosting quantization may deteriorate, even ignoring the increasing complexity of the algorithm. The following results show that under certain circumstances, if the estimate of the distribution is not accurate, thus not correctly reflecting the real one, then the error increases with the number of quantization intervals.

Theorem 3. *There exists a form of perturbation on the distribution function of signal such that if $k_2 > k_1$, for optimal quantizations Q_1 and Q_2 for k_1 and k_2, we have $f'(Q_2) \leq f'(Q_1)$, where f' denotes perturbed f.*

4. NUMERICAL EXAMPLES

In this section, we provide various numerical experiments for illustrations and justifications of our proposed boosting quantization and its properties. The validation of the proposed framework has been carried out in the context of trajectory analysis by adopting the Australian Sign Language [8] datasets. The proposed boosting quantization is also applied on the selected datasets to validate its applicability in both classification and retrieval applications.

In order cope with 2 dimensional time series, and define a proper quantization scheme, we applied the Centroid Distance Function (CDF) transformation as described in [9]. The number of quantization levels has been fixed to 7 for each cluster. After estimating the *pdfs* for each considered trajectory class, we applied the optimization routine described in Section 2 for the evaluation of the best quantization parameters for each cluster. In this phase, the optimization is initialized with a uniform quantizer centered about the mean of the cluster distribution. The noise is considered to be additive and Gaussian. 25 different levels of noise power have been considered for validation, ranging from 10% to 98% of the original signal variance. Each level results in a specific quantization scheme for each class.

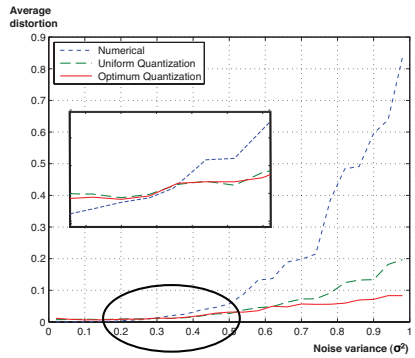

Fig. 1. Average distortion for the original representation, and uniform and optimum quantization, as function of the noise variance (ASL dataset).

Figure 1 depicts the average distortion for the considered configurations (y-axis) w.r.t the noise power (x-axis). As it can be observed, the best performance is achieved, in case of low noise, by the numerical representation; on the contrary, as the noise power increases, the average distortion for the numerical representation rapidly grows, while the distortion for the optimal quantized representation increases more slowly.

We apply the proposed quantization to a simple classification problem involving the three different trajectory classes. The classification accuracy in the three configurations (i.e., numerical, uniform quantization, and optimum quantization) has been computed at different noise levels. The classification process works as follows. The optimum and uniform quantization schemes are obtained for each class, as described in Section 2, for 25 noise levels. For each trajectory class a test set of 69 samples is considered and the centroid is selected as the template for the cluster. Given an incoming path corrupted by noise n, the distance between the trajectory and each cluster template is evaluated in the three considered configurations: the label of the template with minimal distance from the incoming path is selected as the best match. Since the original trajectory is corrupted by artificially generated noise, to guarantee the statistics reliability of our results, the algorithm is run 100 times for each configuration, and considers the average value as the outcome of the analysis. Figure 2 reports the average classification accuracy for the three classes as function of the noise variance, confirming that for low noise power, the best classification results are through a numerical representation. On the contrary, as the variance of

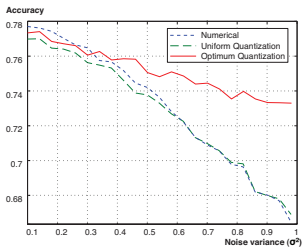

Fig. 2. Classification performance: accuracy *vs.* noise variance (ASL).

the additive noise increases, the numerical representation performance reduces dramatically, highlighting the advantages of the symbolic representation.

In order to give experimental evidence to Theorem 2 of Section 2, we solved the same classification problem considering quantization schemes with increasing number of levels. Figure 3 depicts the curve for the classification accuracy in case of 5, 7, and 10 quantization levels. It is possible to notice how the values of the function f are strictly increasing with the number of levels considered.

Finally, the same dataset has been used for retrieval purposes. In this phase we considered all the trajectories from all the classes as queries (a total of 69*3 = 207 queries), obtaining a ranked list for all of them. Leave-one-out strategy has been chosen. The process has been iterated over the 25 noise levels, as before. In order to show the retrieval performances at increasing noise power, at each noise variance level the Mean Average Precision (MAP) of all queries has been evaluated. The MAP is generally referred to as the geometrical area under the Precision-Recall curve and it's evaluated according the formula in Eq. (3).

$$MAP = \frac{\sum_{r=1}^{N}(P(r) \cdot W(r))}{\#\quad of \quad relevant \quad documents} \tag{3}$$

where N is the number of retrieved documents, $W()$ is a binary function weighting the relevance at a given rank, and $P(r)$ is the precision evaluated at rank r. The retrieval results are reported in Fig. 4 as MAP value (y-axis) *vs.* noise variance (x-axis). Also in this case, the performances in terms of MAP confirm that for low noise power, the numerical representation outperforms the symbolic representation, while for increasing noise variance the optimal quantization offers the best performances.

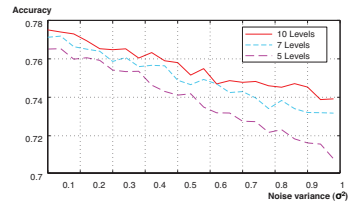

Fig. 3. Classification accuracy for symbolic representation in different quantization level configurations.

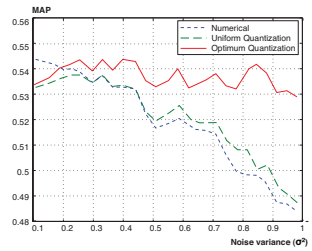

Fig. 4. Retrieval performance in terms of Mean Average Precision *vs.* noise variance.

5. REFERENCES

[1] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory,*, vol. 44, no. 6, pp. 2325 –2383, oct 1998.

[2] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory,*, vol. 28, no. 2, pp. 129 – 137, march 1982.

[3] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory,*, vol. 6, no. 1, pp. 7 –12, march. 1960.

[4] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications,*, vol. 28, no. 1, pp. 84–95, 1980.

[5] M. Broilo, N. Piotto, G. Boato, N. Conci, and F. G. B. De Natale, "Object trajectory analysis in video indexing and retrieval applications," in *Video Search and Mining*. 2010, vol. 287, pp. 3–32, Springer-Verlag.

[6] W. Hu, D. Xie, Z. Fu, W. Zeng, and S. Maybank, "Semantic-based surveillance video retrieval," *IEEE Transactions on Image Processing,*, vol. 16, no. 4, pp. 1168 –1181, apr. 2007.

[7] B.K. Natarajan, "Filtering random noise from deterministic signals via data compression," *IEEE Transactions on Signal Processing,*, vol. 43, no. 11, pp. 2595–2605, 1995.

[8] S. Hettich and S. D. Bay, "The uci kdd archive [http://kdd.ics.uci.edu]," *University of California, Department of Information and Computer Science*, 1999.

[9] F. Bashir, A. Khokhar, and D. Schonfeld, "View-invariant motion trajectory-based activity classification and recognition," *Multimedia Systems*, vol. 12, pp. 45–54, 2006.