A Mixture of Experts Approach to Multi-strategy Image Quality Assessment

Peng Peng and Ze-Nian Li

School of Computing Science, Simon Fraser University,
8888 University Drive, Burnaby, B.C. Canada V5A 1S6
pengp@sfu.ca, li@cs.sfu.ca

Abstract. The success of some recently proposed multi-strategy image quality metrics supports the hypothesis that the Human Visual System (HVS) uses multiple strategies when assessing image quality, where the effect from each strategy on the final quality prediction is conditioned on the quality level of the test image. To date, how to optimally combine multiple strategies into a final quality prediction remains an unsolved problem, especially when more than two strategies are involved. In this paper, we present a data-driven combination method based on a conditional Bayesian Mixture of Experts (BME) model. This method provides an effective way to model the interaction of a flexible number of strategies. Extensive evaluation on three publicly-available image quality databases demonstrates the potential of our method.

Keywords: Image quality assessment, multi-strategy approach, Bayesian mixture of experts (BME), support vector regression (SVR).

1 Introduction

Automatic assessment of image quality is of high significance in multimedia services. Depending on the presence of reference images, there are three types of objective image quality metrics: full-reference, reduced-reference and no-reference. For the first type, a perfect-quality reference image is available during the assessment of its distorted versions. For the other two, we have access to partial information or no information about the reference image. This research focus on the first type.

In the literature, many full-reference image quality metrics have been proposed [15][16]. Some of them achieve good correlation with the human perception of image quality, e.g. the Structural SIMilarity (SSIM) index [13] and the Visual Information Fidelity (VIF) [14]. Most full-reference metrics use a single strategy to evaluate the quality of images. Recently, Larson and Chandler suggest that the human vision system (HVS) uses multiple strategies when assessing image quality. For high-quality images, the HVS employs a "detection-based strategy" to locate and measure distortions that are not readily visible. For low-quality images containing clearly-visible distortions, the HVS uses an "appearance-based strategy" to determine image quality mainly based on how well the image content can be recognized. Relying on these assumptions, they propose the Most

A. Campilho and M. Kamel (Eds.): ICIAR 2012, Part I, LNCS 7324, pp. 123–130, 2012.

Apparent Distortion (MAD) algorithm [1], in which the overall image quality is obtained by adaptive combination of the two strategies. Specifically, high-quality images obtain their rating mostly from the "detection-based strategy", and low-quality images from the "appearance-based strategy". With this multi-strategy scheme, the MAD algorithm achieves state-of-the-art performance. In our previous work, we present a new multi-strategy approach (R-SSIM [2]) which incorporates three much more efficient strategies, and produces comparable results, compared with the MAD algorithm. The success of MAD and R-SSIM strongly supports the idea of using multiple strategies for image quality assessment. However, how to optimally combine multiple strategies remains unsolved. Particularly, when the number of strategies increases, the adaptive combination used in the MAD and R-SSIM methods is not sufficient to model the interaction between the strategies.

In this paper, we present a data-driven approach to multiple-strategy image quality assessment, based on a conditional Bayesian Mixture of Experts (BME) model. Mixture of Experts (MoE) is a well-know model for regression and classification, which learns a probabilistic partitioning of the feature space, as well as the mapping in each sub-region [3,4,5,6]. Generally, in an MoE model, the distribution of the output variables is given by a mixture of component distributions, where the components as well as the mixing coefficients, are conditioned on the input variables. This suits well with the fact that the strategies of a multi-strategy metric should interact differently in different sub-regions of the feature space. Moreover, due to the nature of MoE models, the proposed approach is very flexible on the number of component strategies used. When evaluated on three publicly-available databases, it demonstrates good effectiveness in predicting image quality.

2 Related Work

2.1 Multi-strategy Approaches Using Adaptive Combination

As we mentioned before, the MAD algorithm [1] adaptively combines a "detection-based strategy" with an "appearance-based strategy" to predict image quality. Specifically, the final quality prediction is given by

$$MAD = [Q_{detect}]^{\alpha} \cdot [Q_{appear}]^{(1-\alpha)}, \tag{1}$$

where Q_{detect} relies on HVS characteristics to measure near-threshold distortions, Q_{appear} relies on changes in log-Gabor statistics to capture changes in visual appearance, and the weight $\alpha \in [0, 1]$ is chosen based on the overall level of distortion. In their implementation, α is computed via a function of Q_{detect}:

$$\alpha = \frac{1}{1 + \beta_1 (Q_{detect})^{\beta_2}}, \tag{2}$$

where β_1 and β_2 are set empirically: $\beta_1 = 0.467$ and $\beta_2 = 0.130$.

In our previous work, we present a computationally efficient R-SSIM index which incorporates three strategies. For simplicity, we adopt the adaptive combination method in the MAD algorithm:

$$R\text{-}SSIM = [Q_{ssim}]^{\alpha} \cdot [Q_e \cdot Q_s]^{(1-\alpha)}, \tag{3}$$

where Q_{ssim} is the SSIM index [13], Q_e is an edge-quality measure based on analysis of edge directions, and Q_s is the mean value of the structure component of SSIM over the whole image. The weight α is computed via a function of Q_{ssim}:

$$\alpha = 1 - \frac{1}{1 + \beta_1 (Q_{ssim})^{\beta_2}}, \tag{4}$$

where $\beta_1 = 3$ and $\beta_2 = 5$ are selected.

With the adaptive combination of some carefully designed component strategies, both MAD and R-SSIM have achieved good correlation with human perception of image quality. Naturally, it is of interest to produce even better results. There are two ways to achieve this goal. One is to improve the quality-prediction ability of the component strategies. The other is to employ a more effective combination method. In this work, we focus on the later.

2.2 Mixture of Experts

In [3], Jordan and Jacobs propose a Hierarchical Mixture of Experts (HME) model, for which there exists an Expectation-Maximization (EM) algorithm to learn the parameters. In [4], Bishop and Svensen describe a fully Bayesian treatment of the HME model based on variational inference. This model is employed to estimate speech quality where some encouraging results are reported [7]. Recently, a conditional Bayesian Mixture of Experts model (BME) is proposed in [5,6], which produces superior results on human pose estimation. In this paper, we demonstrate the applicability of the conditional BME model to multi-strategy image quality assessment.

3 Proposed Method

In this section, we describe an approach to perceptual image quality assessment based on a data-driven combination of multiple component strategies.

3.1 Selection of Component Strategies

Instead of designing new component strategies, we adopt some existing strategies, including the two strategies of MAD [1] and the three strategies of R-SSIM [2]. These component strategies are employed on each test image, yielding intermediate quality predictions. The selected intermediate predictions are then fed into a data-driven combination model to obtain the final image quality prediction.

3.2 Conditional Bayesian Mixture of Experts Model

We use a conditional BME model to form the core of our combination method [5,6]. Given the feature vector $\boldsymbol{x}$ yielded by component strategies, we aim to predict the overall image quality y using a model $\mathcal{M}$ with parameters $\boldsymbol{\theta}$. The distribution of the output y conditioned on $\boldsymbol{x}$ is given by a mixture of M component distributions:

$$P(y|\boldsymbol{x}) = \sum_{j=1}^{M} g_j(\boldsymbol{x}, \boldsymbol{\lambda_j}) p_j(y|\boldsymbol{x}, \boldsymbol{w_j}, \sigma_j^2), \tag{5}$$

where

$$g(\boldsymbol{x}, \boldsymbol{\lambda_j}) = \frac{e^{\boldsymbol{\lambda_j^T x}}}{\sum_k e^{\boldsymbol{\lambda_k^T x}}} \tag{6}$$

and

$$p_j(y|\boldsymbol{x}, \boldsymbol{w_j}, \sigma_j^2) = \mathcal{N}(f_i(\boldsymbol{w_j}, \boldsymbol{x}), \sigma_j^2 \mathbf{I}) \tag{7}$$

Here, the mixing coefficients $g_j(\boldsymbol{x}, \boldsymbol{\lambda_j})$ are known as "gating functions", and are normalized to 1. The individual component distributions p_j are called "experts". Specifically, they are Gaussian distributions with mean $f_i(\boldsymbol{w_j}, \boldsymbol{x})$ and covariance matrix $\sigma_j^2\mathbf{I}$. In this work, $f_j(.)$ are linear regressors with weights $\boldsymbol{w_j}$, i.e. $f_j(\boldsymbol{w_j}, \boldsymbol{x}) = \boldsymbol{w_j^T x}$. The weights of experts have Gaussian prior controlled by hyperparameters $\boldsymbol{\alpha}$. Note that both the mixing coefficient and the experts are conditioned on the feature vector $\boldsymbol{x}$. This means that, for different sub-regions in the feature space, the mixing coefficients, as well as the weights of experts, are different. We store the model parameters collectively in $\boldsymbol{\theta} = \{(\boldsymbol{w_j}, \boldsymbol{\alpha_j}, \sigma_j, \boldsymbol{\lambda_j})|j = 1, \ldots, M\}$.

There exists an efficient approximate Bayesian Expectation-Maximization (EM) algorithm to learn the model parameters [6]. In the E-step, the posteriors that expert j has generated datapoint i are estimated. In the M-step two optimization problems are solved, one for each expert and the other for its gate. In [6], a forward feature selection approach is used to deal with high input dimension when training the experts, and a bound optimization method is employed to make the fitting of the gates faster. Since the input space of our work has low input dimensionality ($<= 5$), we do not employ forward feature selection when learning the experts. A Matlab package (*f*BME) for fast training BME models is available online[1].

4 Experiments

We use three image quality databases to evaluate the performance of the proposed multi-strategy image quality metric based on a conditional BME model. To the best of our knowledge, there are seven publicly-available image quality databases with ground-truth subjective quality scores [15]. We choose to evaluate our method on the three largest and also most challenging databases, namely,

[1] *f*BME: `http://www.cs.washington.edu/homes/lfb/software/fBME.htm`

TID2008 [9,10] and CSIQ [1], and LIVE [11,12,13] . Specifically, the TID2008 database contains 25 reference images, 17 distortion types at four levels of distortion, and 1700 distorted images in total. The CSIQ database contains 30 reference images, six types of distortions at four to five levels of distortion, and 866 distorted images. The LIVE databse is comprised of 29 reference images, five distortion types and 779 distorted images.

Because different quality-prediction models produces results in different scales, a non-linear fitting operation between the predicted results and subjective quality scores is performed. The nonlinear mapping is given by a five-parameter logistic function:

$$Quality(x) = \beta_1 logistic(\beta_2, x - \beta_3) + \beta_4 x + \beta_5 \tag{8}$$

$$logistic(a, x) = \frac{1}{2} - \frac{1}{1 + exp(ax)} \tag{9}$$

where x is the score obtained from an objective metric. There are four criteria commonly used for performance comparison, namely: Spearman rank correlation coefficient (SRCC), Kendall rank correlation coefficient (KRCC), Pearson linear correlation coefficient (PLCC) and Root mean square error (RMSE). The first two measure prediction monotonicity, and the other two measure prediction accuracy.

The proposed multi-strategy approach based on a conditional BME model is compared with the following methods:

- SSIM [13] and VIF [14]: two prominent single-strategy metrics as reported in [15];
- MAD [1] and R-SSIM [2]: Two state-of-the-art multi-strategy metrics using adaptive combination;
- Two baseline data-driven multi-strategy methods based on support vector regression (SVR), using the same component strategies as the proposed approach.

On each database, half number of the distorted images are randomly selected to train the conditional BME model, and the remaining half are used for test. We also use the same training set and test set to evaluate the SVR-based baseline data-driven approaches, in which two types of kernels are employed, namely linear kernel and polynomial kernel. The number of experts in the conditional BME model, as well as the parameters of the SVR models, are selected by cross-validation on the training set. A Matlab implementation of SVR is provided in the LIBSVM package [8].

The test results are shown in Table 1, Table 2 and Table 3. We use “baseline-1” and “baseline-2” to denote the two data-driven approaches based on SVR with liner kernel and polynomial kernel, respectively. The notations “[CS2]”, “[CS3]” and “[CS5]” indicate the component strategies employed to produce the results in that row. Specifically, “[CS2]” corresponds to the two component strategies of MAD [1], “[CS3]” corresponds to the three component strategies of R-SSIM [2], and “[CS5]” corresponds to a combination of “[CS2]” and “[CS3]”. Since the

Table 1. Evaluation on the TID Database

Metrics	SRCC	KRCC	PLCC	RMSE
SSIM	0.7629	0.5647	0.7673	0.8647
VIF	0.7361	0.5752	0.7991	0.8107
MAD [CS2]	0.8255	0.6325	0.8214	0.7692
Baseline-1 [CS2]	0.7875	0.5838	0.7851	0.8353
Baseline-2 [CS2]	0.7924	0.5892	0.7877	0.8308
Our Method [CS2]	0.8068	0.6086	0.8163	0.7789
R-SSIM [CS3]	0.7705	0.5799	0.7944	0.8190
Baseline-1 [CS3]	0.7577	0.5691	0.7834	0.8380
Baseline-2 [CS3]	0.7900	0.5942	0.8038	0.8022
Our Method [CS3]	0.8300	0.6303	0.8415	0.7285
Baseline-1 [CS5]	0.8581	0.6641	0.8660	0.6742
Baseline-2 [CS5]	0.8716	0.6817	0.8774	0.6470
Our method [CS5]	**0.8882**	**0.7053**	**0.8958**	**0.5994**

Table 2. Evaluation on the CSIQ Database

Metrics	SRCC	KRCC	PLCC	RMSE
SSIM	0.8737	0.6862	0.7708	0.1613
VIF	0.9152	0.7493	0.9190	0.0999
MAD [CS2]	0.9450	0.7931	0.9469	0.0814
Baseline-1 [CS2]	0.8988	0.7211	0.9137	0.1029
Baseline-2 [CS2]	0.8990	0.7217	0.9127	0.1035
Our Method [CS2]	0.9180	0.7503	0.9351	0.0898
R-SSIM [CS3]	0.9309	0.7651	0.9248	0.0963
Baseline-1 [CS3]	0.9317	0.7644	0.9328	0.0912
Baseline-2 [CS3]	0.9310	0.7652	0.9346	0.0901
Our Method [CS3]	0.9374	0.7773	0.9426	0.0845
Baseline-1 [CS5]	0.9380	0.7838	0.9503	0.0789
Baseline-2 [CS5]	0.9454	0.7954	0.9557	0.0745
Our method [CS5]	**0.9573**	**0.8182**	**0.9650**	**0.0664**

results on the three databases show very similar patterns, we do not explicitly mention any single database in the following analysis.

In the case of "[CS2]", our method outperforms the two SVR-based baseline methods, but does not perform as well as the original MAD metric [1].

Table 3. Evaluation on the LIVE Database

Metrics	SRCC	KRCC	PLCC	RMSE
SSIM	0.9496	0.8014	0.9526	8.5862
VIF	0.9646	0.8313	0.9459	9.1552
MAD [CS2]	0.9639	0.8348	0.9663	7.2663
Baseline-1 [CS2]	0.9291	0.7699	0.9380	9.7752
Baseline-2 [CS2]	0.9294	0.7700	0.7406	18.9549
Our Method [CS2]	0.9366	0.7791	0.9441	9.2988
R-SSIM [CS3]	0.9610	0.8255	0.9626	7.6469
Baseline-1 [CS3]	0.9558	0.8146	0.9578	8.1072
Baseline-2 [CS3]	0.9585	0.8205	0.9597	7.9319
Our Method [CS3]	0.9648	0.8339	0.9661	7.2831
Baseline-1 [CS5]	0.9613	0.8280	0.9649	7.4101
Baseline-2 [CS5]	0.9623	0.8312	0.9661	7.2807
Our method [CS5]	**0.9711**	**0.8535**	**0.9729**	**6.5228**

This implies that, when only two component strategies are employed, a carefully designed adaptive combination method could be more effective than data-driven methods. However, when the number of component strategies increases, our data-driven method become a more appealing option. In the case of "[CS3]", our method outperforms both the R-SSIM metric [2] and the two baseline data-driven methods. This indicates that the trained conditional BME model can better capture the interaction of the three component strategies than the adaptive combination method (see Eq. 3) and the SVR models. Moreover, in the case of "[CS5]", our method achieves the best performance among all evaluated metrics by taking advantages of employing more component strategies and the superior BME-based combination model.

5 Conclusions

The success of some recently proposed multi-strategy image quality metrics supports the assumption that the HVS combines multiple strategies to determine image quality. However, as the number of strategies increases, the existing combination method is not sufficient to model the interaction of them. In this paper, we present a data-driven approach to combine multiple strategies, based on a conditional Bayesian Mixture of Experts model. When evaluated on three publicly-available image quality databases, the proposed method produces very promising results. Moreover, since the BME model is capable of handling input with any number of dimensions, it provides great flexibility in designing (or selecting) component strategies for finer multi-strategy image quality assessment.

References

1. Larson, E.C., Chandler, D.M.: Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy. J. Electron. Imaging 19(1), 11006 (2010)
2. Peng, P., Li, Z.N.: Incorporating Structural Edge Quality to Regularize the Structural Similarity Index (submitted for publication)
3. Jordan, M.I., Jacobs, R.A.: Hierarchical Mixtures of Experts and the EM Algorithm. Neural Computation, 181–214 (1994)
4. Bishop, C.M., Svensén, M.: Bayesian Hierarchical Mixtures of Experts. In: Nineteenth Conference on Uncertainty in Artificial Intelligence, pp. 57–64 (2003)
5. Sminchisescu, C., Kanaujia, A., Metaxas, D.N.: BME : Discriminative Density Propagation for Visual Tracking. IEEE Trans. Pattern Anal. Mach. Intell., 2030–2044 (2007)
6. Bo, L., Sminchisescu, C., Kanaujia, A., Metaxas, D.N.: Fast Algorithms for Large Scale Conditional 3D Prediction. In: CVPR(2008)
7. Mossavat, S.I., Amft, O., de Vries, B., Petkov, P.N., Kleijn, W.B.: A Bayesian Hierarchical Mixture of Experts Approach to Estimate Speech Quality. In: Second International Workshop on Quality of Multimedia Experience (QoMEX) (2010)
8. Chang, C.W., Lin, C.J.: LIBSVM: A Library for Support Vector Machines. ACM Trans. on Intell. Syst. Technol. 2(3), 27 (2011), software `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
9. Ponomarenko, N., Battisti, F., Egiazarian, K., Astola, J., Lukin, V.: Metrics Performance Comparison for Color Image Database. In 4th International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale (2009)
10. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID 2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics. Advances of Modern Radioelectronics 10, 30–45 (2009)
11. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE Image Quality Assessment Database Release 2, `http://live.ece.utexas.edu/research/quality`
12. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. IEEE Trans. Image Process. 15(11), 3440–3451 (2006)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Measurement to Structural Similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
14. Sheikh, H.R., Bovik, A.C.: Image Information and Visual Quality. IEEE Trans. Image Process. 15(2), 43–444 (2006)
15. Lin, W., Kuo, C.C.J.: Perceptual Visual Quality Metrics: A Survey. Journal of Visual Communication and Image Representation 22(4), 297–312 (2011)
16. Moorthy, A.K., Bovik, A.C.: Visual Quality Assessment Algorithms: What Does The Future Hold? Multimedia Tools Appl. 51(2), 675–696 (2011)