

Fuzzy Regression for Perceptual Image Quality Assessment

Kit Yan Chan^a, Ulrich Engelke^b

^aKit Yan Chan is with the Department of Electrical and Computer Engineering, Curtin University, Australia.

^bUlrich Engelke with the Digital Productivity Flagship of the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Hobart, Australia

Abstract - Subjective image quality assessment (IQA) is fundamentally important in various image processing applications such as image/video compression and image reconstruction, since it directly indicates the actual human perception of an image. However, fuzziness due to human judgment is neglected in current methodologies for predicting subjective IQA, where the fuzziness indicates assessment uncertainty. In this article, we propose a fuzzy regression method that accounts for fuzziness introduced through human judgment and the limitations of widely-used psychometric quality scales. We demonstrate how fuzzy regression models provide fuzziness information regarding subjective IQA. We benchmark the fuzzy regression method against the commonly used explicit modeling method for subjective IQA namely statistical regression by considering three real situations involving subjective image quality experiments where: a) the number of participants is insufficient; b) an insufficient amount of data is used for modelling; and c) variant fuzziness is caused by human judgment. Results indicate that fuzzy regression models achieve more effective data fitting and better generalization capability when predicting subjective IQA under different types and levels of image distortion.

Index Terms – Fuzzy regression, mean opinion scores (MOS), subjective image quality assessment, objective image quality metric.

I. INTRODUCTION

The number of digital images we take each year has soared. In 2013, 27,800 digital images were uploaded to Instagram every minute and 208,300 digital images were uploaded to Facebook every minute. Yahoo

expected that 880 billion digital images would be captured in 2014 [26]. In the past decade, we have experienced great technological advancements of new devices for capture, storage, compression, transmission, and display of digital images, mostly resulting in significant increases of image quality. The raw visual information typically passes through multiple steps in an imaging pipeline, each of which affects the quality of the image at the receiver end. With these increasingly complex multimedia applications, perceived image quality evaluation has been receiving significant attention as a means of ensuring certain levels of quality of service. Given the abundance of visual data, it is essential to develop efficient computational prediction models to automatically evaluate image quality and to control the perceptual quality of the visual content by tuning the multi-parameters of the imaging pipeline.

However, it is challenging to develop prediction models that accurately represent image quality perceived by a human. Subjective image quality assessment (IQA) is typically used as a ground truth to develop computational image quality prediction models [6] as humans are considered to be the observers and consumers of most systems and products involving digital images. In subjective IQA, a group of interviewers typically scores the quality of a number of images. Subjective IQA provides a useful tool for evaluating the visual effect of a wide range of artifacts which are imposed on digital images for image acquisition, processing, transportation, compression, and storage [22]. However, it is not possible to implement subjective IQA in real time or as a systematic evaluator for image enhancement. It cannot be incorporated into the design and optimization of image processing algorithms in order to enhance image quality. For this reason, there has been an increasing interest in correlating subject IQA with the objective IQA metrics in order to automatically predict or estimate the perceived image quality [6], where the objective IQA metrics range from simple numerical measures [8] such as the signal-to-noise ratio and the bit error rate [24] to highly complex models incorporating those characteristics of the human visual system that are considered crucial for visual quality perception [11,21,27,34,35]. These prediction models aim to automatically predict perceived image quality as obtained in subjective experiments. Currently, there are no image quality prediction models that work well across a wide range of visual content and distortion types; typically, they perform well only on subsets of the above [6].

To develop quality prediction models, implicit modelling methods based on artificial intelligence have been used based on experimental data of subjective image quality experiments, which are typically based on n-point psychometric scales, such as the Likert scale [19], to record responses from a number of participants who judge the opinion scores of images presented to them. The opinion scores are then combined into Mean Opinion Scores (MOS) for all images, which in turn are instrumental in the training and validation of computational image quality prediction models. Neural networks [7,9] have been used to develop models for predicting subjective IQA, but these approaches lack transparency since they are black-box in nature. Explicit information cannot be indicated in the neural networks. Also, the training time required by the neural networks is much longer compared with the statistical regression, when the network size is large. Fuzzy modeling-based approaches have also been applied to develop prediction models for IQA [4]. However, compared with statistical regression methods, more explicit information can be found in statistical regression models which are in a polynomial form. Hence, variable significances and variable interactions can be determined in the polynomial of the regression models [28].

To generate explicit models, statistical regression [28] is commonly used. Engineers, in general, prefer to use statistical methods because more explicit information can be found than using the fuzzy modeling-based approaches or the neural networks. However, subjective image quality experiments involve human opinion judgments which are inherently imprecise, inconsistent over time, and often non-consensual when involving a group of individuals [5]. Hence, the assessment represents a source of uncertainty that is typically neglected in the development of quality prediction models that correlate subjective IQA and objective IQA. Therefore, the judging process inherently represents a source of uncertainty and bias that is neglected in statistical regression used to match predicted quality with MOS [6,23]. Also, the regression models may not be performed accurately, as they can only be applied accurately within the range for which they are developed [14]. They can be applied only if the given experimental data is normally distributed according to the developed regression model. They can represent a crisp relationship only between the objective image quality metrics and subjective image quality measure, while the uncertainty of the relationship cannot be addressed. Instead, in this paper, we propose to use fuzzy regression to overcome these shortcomings.

In new product development, fuzzy regression [3,13,15,18,29,30] has commonly been used to model correlations between subject customer satisfaction and objective engineering characteristics of new products, where settings of engineering characteristics can affect customer satisfaction with the product. Based on the correlation models, the engineering characteristics can be specified by optimizing customer satisfaction. Fuzzy regression has a distinct advantage over statistical regression as it can address the fuzziness of subjective judgments and it can perform effectively using a small or even incomplete data set [32]. In this article, we propose a novel image quality assessment technique based on fuzzy regression that attempts to account for the ‘fuzziness’ of human judgment introduced through subjective IQA. Indeed, the approach of fuzzy regression is the first time to be developed in order to model the relationship between objective image quality metrics and subjective image quality measure, where the fuzzy regression model attempts to address the fuzziness caused by the subjective IQA. Three validations with three conditions in MOS data sets were performed in order to evaluate whether fuzzy regression outperforms statistical regression in term of generalization capability: 1) varying number of participants; 2) varying data sizes; and 3) varying amount of fuzziness. These three conditions simulated the real situations in subjective image quality experiments where: insufficient numbers of participants are involved; insufficient amount of MOS data is used for modelling; and different amounts of fuzziness are caused by human evaluation of MOS. Experimental results shows that the proposed method overcomes the shortcomings of more widely adopted statistical regression techniques which disregard fuzziness of human judgment and require large data sets with normal distribution assumption.

The rest of the article can be organized as follows: Section II and III discuss fuzziness in IQA and fuzzy regression in developing prediction models for MOS. Section IV validates the fuzzy regression on an extensive image quality database and benchmarks it against statistical regression. Section V concludes the article.

II. FUZZINESS IN SUBJECTIVE IQA

Perception of subjective IQA is inherently imprecise as, typically, only an approximate judgment is made. The widely used N -point psychometric scales, however, usually map qualitative judgments onto opinion scores [19]. It has been shown in [33] that on a continuous rating scale with N opinion scores, people tend to judge

quality around the integers with some degree of uncertainty. One may refer to judgments ‘about’ a particular integer X on psychometric scales. For instance, one may judge quality to be ‘about 2’ on a 5-point scale when one feels that the image quality is ‘Poor’ and ‘about 3’ when one feels that the image quality is ‘Fair’. The question then arises: what does ‘about X ’ actually mean in opinion scores for image quality?

Based on the fuzzy set theory [32], the linguistic term ‘about y^c ’ can be explained by a fuzzy number, $\tilde{y} = (y^c, y^r, y^l)$, with a fuzzy membership function, $\mu_{\tilde{y}}(y)$:

$$\mu_{\tilde{y}}(y) = \begin{cases} 1 & y = y^c \\ \frac{y - y^c}{y^c - y^l} & y^l \leq y < y^c \\ \frac{y^r - y}{y^r - y^c} & y^c \leq y < y^r \\ 0, & \text{otherwise} \end{cases}$$

where $\mu_{\tilde{y}}(y)$ indicates the membership function of the linguistic term ‘about y^c ’; y^c is the center of the fuzzy member of which the center indicates the degree of opinion score for ‘about y^c ’; y^l and y^r are the left and right spreads respectively which indicates the fuzziness of opinion score. When y is exactly equal to y^c , the membership, $\mu_{\tilde{y}}(y)$, is 1 and thus y is a full membership of ‘about y^c ’. When y is within the value between y^l and y^r , y is a membership of ‘about y^c ’. When y^l is close to y^r , the fuzziness of ‘about y^c ’ is low. Here, a fuzzy triangle function is used as it requires uncomplicated fuzzy arithmetical operations compared with a Gaussian or trapezoidal function [5].

Figure 1 shows two fuzzy numbers for the opinion scores namely ‘about 2’, $\tilde{y}_1 = (2, 0.25, 0.25)$, and ‘about 3’, $\tilde{y}_2 = (3, 1, 0.5)$. In this example, the fuzziness of ‘about 2’ is smaller than that of ‘about 3’, as the difference between the left and right spreads of \tilde{y}_2 is larger than that of \tilde{y}_1 .

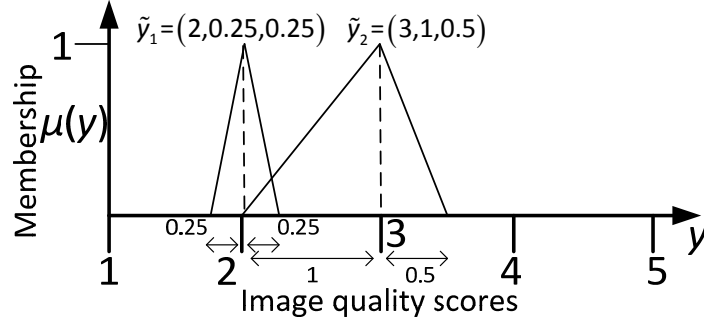


Fig. 1: Two fuzzy numbers, ‘about 2’ and ‘about 3’

In a typical image quality experiment, N_I images are evaluated by N_c participants and the opinion score for the i -th image evaluated by the p -th participant is given by a fuzzy number as

$$\tilde{y}_i(p) = (y_i^c(p), y_i^l(p), y_i^r(p)) \quad (1)$$

Based on $\tilde{y}_i(p)$, we define here the fuzzy mean opinion score (fuzzy MOS), $\tilde{y}_i = (y_i^c, y_i^l, y_i^r)$, for the i -th image with $i = 1, 2, \dots, N_I$ as:

$$\tilde{y}_i = \left(\frac{1}{N_c} \sum_{p=1}^{N_c} y_i^c(p), \frac{1}{N_c} \sum_{p=1}^{N_c} y_i^l(p), \frac{1}{N_c} \sum_{p=1}^{N_c} y_i^r(p) \right) = \frac{1}{N_c} \sum_{p=1}^{N_c} \tilde{y}_i(p) \quad (2)$$

The commonly used MOS for subjective IQA namely conventional MOS [6], \bar{y}_i , for the i -th image can be computed by (3) which accounts only the crisp opinion scores (i.e. center of the fuzzy number, $y_i^c(p)$) for all participants.

$$\bar{y}_i = \frac{1}{N_c} \sum_{p=1}^{N_c} y_i^c(p) \quad (3)$$

Unlike the fuzzy MOS, it can be noted from (3) that the fuzziness of the opinion scores for the image quality cannot be determined by the conventional MOS, which is its limitation. Without determining the fuzziness, the uncertainty in human judgments cannot be indicated.

III. SUBJECTIVE IQA MODELS

A. Statistical regression for MOS

The statistical regression model in (4) can be developed using the collected MOS data of the N_I image, \bar{y}_i , in order to predict the conventional MOS.

$$\hat{y}_i = \beta_0 + \beta_1 x_1^i + \beta_2 x_2^i + \dots + \beta_M x_M^i + \varepsilon_i \quad (4)$$

where β_j are the $M+1$ regression coefficients and x_j^i with $j=1,2,\dots,M$ are the M normalized objective IQA metrics for the i -th image. The random error term ε_i is assumed to be normal distributed with zero mean, constant variance and covariance.

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i, j \text{ with } i \neq j. \quad (5)$$

However, statistical regression suffers from two limitations which are discussed in [16]. Firstly, the fuzziness of human judgment (i.e. y_i^l and y_i^r) cannot be addressed through statistical regression but only the conventional MOS (i.e. \bar{y}_i) is accounted for. Secondly, confidence intervals can be constructed for the estimates of \hat{y}_i only if the following three assumptions are satisfied [17]: 1) all x_j^i are normal distributed on the regression line \hat{y}_i ; 2) the deviations of \bar{y}_i are normal distributed; and 3) the relationship between \bar{y}_i and x_j^i , is continuous over the data range.

The normal distribution assumption is usually violated for very high or low ratings, due to the inherent limits of quality scales. It is also only satisfied for large data sets and often results in misleading confidence intervals otherwise. As a result, one needs to balance a trade-off between the accuracy, and cost and time in collecting data when using statistical regression [1], as a large data size is usually required for the developing the models. Also, the fuzziness of the subjective IQA (i.e. y_i^l and y_i^r illustrated in (3)) cannot be addressed by the regression model which only accounts the MOS crisp (i.e. \bar{y}_i in (2)).

B. Fuzzy regression for fuzzy MOS

Fuzzy regression attempts to overcome the limitations of statistical regression as it accounts for fuzziness of human judgments. It is also less error-prone when only a small number of imprecise data are used to estimate the fuzzy MOS [10]. The fuzzy MOS, $\hat{y}_i = (\hat{y}_i^c, \hat{y}_i^l, \hat{y}_i^r)$ for the i -th image can be estimated using fuzzy regression [16] as follows:

$$\hat{y}_i = (\hat{y}_i^c, \hat{y}_i^l, \hat{y}_i^r) = (a_0^c, a_0^l, a_0^r) + \sum_{j=1}^M (a_j^c, a_j^l, a_j^r) x_j^i = \left(a_0^c + \sum_{j=1}^M a_j^c \cdot x_j^i, a_0^l + \sum_{j=1}^M a_j^l \cdot x_j^i, a_0^r + \sum_{j=1}^M a_j^r \cdot x_j^i \right) \quad (6)$$

where (a_j^c, a_j^l, a_j^r) is the j -th fuzzy coefficient.

Apart from estimating the crisp of the conventional MOS \hat{y}_i as in (4), the fuzzy regression in (6) estimates uncertainty or vague phenomena based on the left and right spreads, \hat{y}_i^l and \hat{y}_i^r . The fuzzy coefficients (a_j^c, a_j^l, a_j^r) address the residual of the estimates to the real fuzzy MOS data that is assumed to be due to human fuzzy judgments [10,16]. All (a_j^c, a_j^l, a_j^r) are determined by minimizing the sum of residual errors [3] between the real fuzzy MOS data, \tilde{y}_i , and the fuzzy MOS estimates, \hat{y}_i , as:

$$\begin{aligned} Error &= \sum_{i=1}^{N_i} (\hat{y}_i - \tilde{y}_i)^2 \\ &= \sum_{i=1}^{N_i} \left((a_0^c, a_0^l, a_0^r) + \sum_{j=1}^M (a_j^c, a_j^l, a_j^r) \cdot x_j^i - (y_i^c, y_i^l, y_i^r) \right)^2 \end{aligned} \quad (7)$$

The expression in (7) can be expanded based on the weighted fuzzy arithmetic operations for asymmetric triangular membership functions as (8). The detailed steps of the expression can be referred to [3].

$$\begin{aligned} Error &= \sum_{k=1}^{N_i} \left(a_0^c + \sum_{j=1}^M a_j^c x_j - y_k^c \right)^2 + \\ &\frac{1}{3} \sum_{k=1}^{N_i} \left(a_0^c + \sum_{j=1}^M a_j^c x_j - y_k^c \right) \left[\left(a_0^r + \sum_{j=1}^M a_j^r x_j - y_k^r \right) - \left(a_0^l + \sum_{j=1}^M a_j^l x_j - y_k^l \right) \right] + \\ &\frac{1}{12} \sum_{k=1}^{N_i} \left[\left(a_0^r + \sum_{j=1}^M a_j^r x_j - y_k^r \right)^2 + \left(a_0^l + \sum_{j=1}^M a_j^l x_j - y_k^l \right)^2 \right] + \sum_{k=1}^{N_i} \left(a_0^c + \sum_{j=1}^M a_j^c x_j - y_k^c \right)^2 \end{aligned} \quad (8)$$

The *Error* is minimized by deriving (8) with respect to each a_j^c , a_j^r , and a_j^l , with $j=1,2,\dots,M$, and then solving the derivatives which are set to zero. The derivatives of (8) with respect to a_j^c , a_j^r , and a_j^l for $j=1,2,\dots,M$ are given by equations (9.1), (9.2) to (9.M) respectively. The detailed steps of expanding the derivatives can be referred to [3].

$$\frac{\partial(\text{Error})}{\partial a_0^*} = M \cdot a_0^* + \left(\sum_{i=1}^{N_I} x_1^i \right) \cdot a_1^* + \left(\sum_{i=1}^{N_I} x_2^i \right) \cdot a_2^* + \dots + \left(\sum_{i=1}^{N_I} x_M^i \right) \cdot a_M^* - \sum_{i=1}^{N_I} y_i^* = 0, \text{ with } j=0 \quad (9.1)$$

$$\frac{\partial(\text{Error})}{\partial a_1^*} = \left(\sum_{i=1}^{N_I} x_1^i \right) \cdot a_0^* + \left(\sum_{i=1}^{N_I} (x_1^i)^2 \right) \cdot a_1^* + \left(\sum_{i=1}^{N_I} x_1^i \cdot x_2^i \right) \cdot a_2^* + \dots + \left(\sum_{i=1}^{N_I} x_1^i \cdot x_M^i \right) \quad (9.2)$$

$$- \sum_{i=1}^{N_I} (x_1^i \cdot y_i^*) = 0, \text{ with } j=1$$

: : : : : :

$$\frac{\partial(\text{Error})}{\partial a_M^*} = \left(\sum_{i=1}^{N_I} x_M^i \right) \cdot a_0^* + \left(\sum_{i=1}^{N_I} x_M^i \cdot x_1^i \right) \cdot a_1^* + \left(\sum_{i=1}^{N_I} x_M^i \cdot x_2^i \right) \cdot a_2^* + \dots + \left(\sum_{i=1}^{N_I} (x_M^i)^2 \right) \quad (9.M)$$

$$- \sum_{i=1}^{N_I} (x_M^i \cdot y_i^*) = 0, \text{ with } j=M$$

where '*' ∈ ['c', 'r', 'l'] for both y_i^* and a_j^* .

In (9.1) to (9.M), all a_j^* with $j=1,2,\dots,M$ are unknowns, and all y_i^* and x_j^i are constants. Therefore, the formulation in (9.1) to (9.M) is identical to the least squares regression with M unknowns and M equations. With '*' = 'c', the least squares regression can be used to determine a_j^c , $j=1,2,\dots,M$, based on the centers of the real fuzzy MOS data, y_i^c , $i=1,2,\dots,N_I$. Similarly, a_j^l and a_j^r can be obtained respectively, based on left, y_i^l , and right, y_i^r , spreads of the real fuzzy MOS data. With the fuzzy coefficients, (a_j^c, a_j^l, a_j^r) , the fuzzy regression formulated in (6) can be used to estimate fuzzy MOS $\hat{\tilde{y}} = (\hat{y}^c, \hat{y}^l, \hat{y}^r)$. As the center values, \hat{y}^c , and the vagueness, \hat{y}^l and \hat{y}^r , are indicated on the fuzzy MOS estimates, these three numbers can be used to estimate the values of the MOS through the defuzzification given in (8). Based on the second term of (8), the potential bias in human judgment due to the fuzziness can be determined.

$$\mathfrak{B}(\hat{\tilde{y}}) = \hat{y}^c + \frac{1}{6} \cdot (\hat{y}^r - \hat{y}^l) \quad (10)$$

IV. ALGORITHMIC PERFORMANCE

This section discusses how the fuzzy regression can be used to indicate the fuzziness of IQA and also this section evaluates the algorithmic performance of the fuzzy regression by benchmarking it against the commonly used statistical regression for developing subjective IQA models [6].

A. Image quality database

Both approaches are evaluated using Zaric et al. [37]'s extensive image quality database which is developed for evaluating the effectiveness of image quality metrics. The database is based on subjective opinion of human observers for undistorted and distorted images, where the subjective evaluations were conducted in accordance with Recommendation ITUR BT.500-11 [12]. The database is the subsequent work from the frequently used Sheikh et al. [31] and Ponomarenko et al.'s [25] image quality databases. Here, we used Zaric et al. [37]'s extensive image quality database in this research, as Zaric et al. [37]'s database records each individual image quality evaluation for each participant, while the averages of the image quality evaluations for all participants are only available for both Sheikh et al. [31] and Ponomarenko et al. [25] databases. Based on the individual image quality evaluation for each participant, the fuzzy MOS for each image can be determined using Zaric et al. [37]'s database.

The database contains 575 images, 23 of which are original color images, referred to as O^i , $i=1,2,\dots,23$, and 552 distorted images, referred to as $D_{j,k}^i$. Specifically, $D_{j,k}^i$ represents a distorted version of O^i with the j -th distortion type and the k -th degradation level. The database covers four different distortion types: Additive White Gaussian Noise (AWGN), Gaussian Blur (GB), JPEG2000 coding and JPEG coding. Six degradation levels are available for each distortion type, with larger k relating to larger distortions.

The opinion scores for each image were collected by 118 naïve participants between 20 and 30 years of age. Each image was scored between 14 and 36 times with an average number of scores of approximately 20. Figure 2 illustrates a 5-point psychometric scale with the labels: 0=bad, 0.25=poor, 0.50=fair, 0.75=good and 1.00=excellent [36]. These labels can be related to five fuzzy numbers [20]: $(0,0_f)$, $(0.25_f, f)$, $(0.50_f, f)$,

$(0.75, f, f)$, and $(1.00, f, 0)$, where f is the scoring fuzziness. The fuzzy MOS of each image is computed based on (2). In this work, we considered four levels of fuzziness: $f = 0.15, 0.20, 0.25$, and 0.28 .

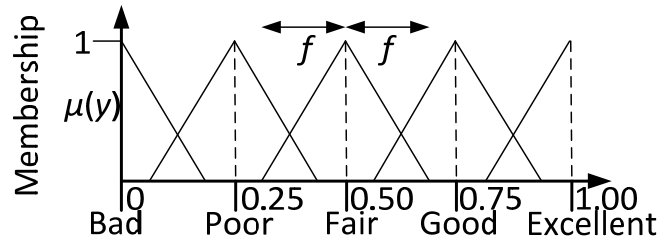


Fig. 2: Fuzzy 5-point psychometric quality scale

B. Model performance

To demonstrate the effectiveness of the fuzzy regression method, we optimized the statistical regression models individually for the four distortion types. For this purpose, we chose five feature metrics that were previously shown [6] to capture well these image distortions namely, block boundary differences (x_1), edge smoothness (x_2), edge-based image activity (x_3), gradient-based image activity (x_4), and image histogram statistics (x_5).

The performance of the fuzzy regression is evaluated by the following mean absolute error (MAE),

$$e^{FR} = \sum_{\exists i \in I} \left| \mathfrak{F}(\hat{y}_i) - \bar{y}_i \right| \quad (11)$$

which quantifies the difference between the actual MOS, \bar{y}_i , and the predicted MOS, $\mathfrak{F}(\hat{y}_i)$, of the i -th image, with \mathfrak{F} denoting the defuzzification of the fuzzy regression outcome defined in (10), where all the i -th image in the image set, I , is considered. Similarly, the following MAE for the statistical regression,

$$e^{SR} = \sum_{\exists i \in I} \left| \hat{y}_i - \bar{y}_i \right| \quad (12)$$

is computed as a benchmark for the fuzzy regression, where e^{SR} is used to evaluate the difference between the actual MOS, \bar{y}_i , and the predicted MOS, \hat{y}_i , of the statistical regression model.

Table 1 illustrates the four fuzzy regression models for the four distortion types, along with the MAE and Pearson linear correlation coefficients (PLCCs) between the actual and predicted MOS. The PLCCs attempt to

indicate the correlation between the actual and predicted MOS, which gives a value between +1 and -1, where 1 is totally positive correlated, 0 is not correlated, and -1 is totally negative correlated. The PLCCs has widely used in the statistics as a measure of the degree of linear dependence between the actual observations and the predictions.

For comparison, the statistical regression performance based on MOS is also reported. In both cases, all images in the Zaric database were used in the training. The smaller training errors and higher correlations indicate that the fuzzy regression outperforms the statistical regression for the four distortion types considered in this work. For further illustration, we present in Figures 3 (a)-(d) the first four original images, O^1 , O^2 , O^3 , and O^4 , and the first four distorted images, $D_{1,6}^1$, $D_{2,6}^2$, $D_{3,6}^3$, and $D_{4,6}^4$ corrupted by AWGN, GB, JPEG2000 and JPEG, respectively, with the strongest degradation ($k=6$). Figures 4 (a)-(d) show the actual fuzzy MOS and predicted fuzzy MOS for the four images. They exemplify that the degree of fuzziness in terms of the spreads depends on the distortion type of the respective image. Furthermore, the left and right spreads are not necessarily symmetrical, but are skewed towards the centre of the quality scale. This effect dominates for high and low fuzzy MOS and is largely an artifact of the quality scale limits that is taken into account in fuzzy regression but not in statistical regression.

Table 1: Fuzzy regression models for each distortion type, MAE for both fuzzy regression and statistical regression and Pearson linear correlation coefficients between observed MOS and predicted MOS

| Distortion types | Fuzzy regression models | MAE for fuzzy regression | MAE for statistical regression | PLCC for fuzzy regression | PLCC for statistical regression |
|------------------|---|--------------------------|--------------------------------|---------------------------|---------------------------------|
| AWGN | $\tilde{y} = (0.14, 0.27, 0.41) + (0.89, -0.01, -0.35) \cdot x_1 + (0.49, 0.00, -0.17) \cdot x_2 + (-0.68, -0.37, 0.04) \cdot x_3 + (-0.42, 0.12, 0.13) \cdot x_4 + (0.30, -0.02, -0.13) \cdot x_5$ | 0.150 | 0.168 | 0.972 | 0.897 |
| GB | $\tilde{y} = (0.69, 0.13, 0.08) + (-0.80, -0.07, 0.21) \cdot x_1 + (-0.00, 0.21, 0.16) \cdot x_2 + (0.262, -0.12, -0.16) \cdot x_3 + (0.03, 0.03, 0.03) \cdot x_4 + (0.19, 0.06, -0.03) \cdot x_5$ | 0.118 | 0.120 | 0.983 | 0.864 |
| JPEG2000 | $\tilde{y} = (1.06, 0.27, 0.06) + (-1.02, -0.29, 0.11) \cdot x_1 + (-0.61, -0.07, 0.13) \cdot x_2 + (0.75, -0.00, -0.24) \cdot x_3 + (-0.22, 0.08, 0.16) \cdot x_4 + (-0.02, -0.01, -0.03) \cdot x_5$ | 0.286 | 0.297 | 0.986 | 0.846 |
| JPEG | $\tilde{y} = (0.98, 0.28, 0.12) + (0.00, 0.05, 0.01) \cdot x_1 + (-0.09, -0.01, -0.01) \cdot x_2 + (0.11, 0.06, 0.02) \cdot x_3 + (-1.03, -0.22, 0.20) \cdot x_4 + (-0.05, -0.01, -0.03) \cdot x_5$ | 0.129 | 0.146 | 0.979 | 0.906 |



Fig. 3a Original O^1 , (upper) and distorted with AWGN, $D_{1,6}^1$, (lower)



Fig. 3b Original O^2 (upper) and distorted with GB $D_{2,6}^2$ (lower)



Fig. 3c Original O^3 (upper) and distorted with JPEG2000 $D_{3,6}^3$ (lower)

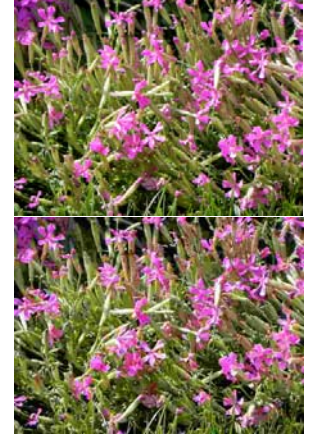


Fig. 3d Original O^4 (upper) and distorted with JPEG $D_{4,6}^4$ (lower)

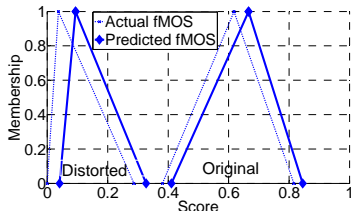


Fig. 4a Fuzzy MOS for O^1 and $D_{1,6}^1$.

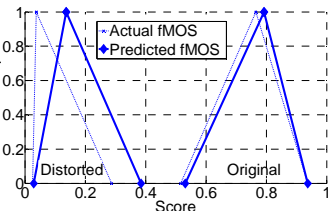


Fig. 4b Fuzzy MOS for O^2 and $D_{2,6}^2$.

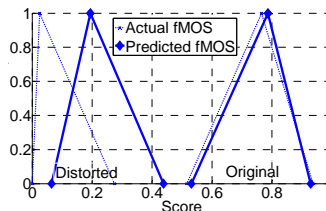


Fig. 4c Fuzzy MOS for O^3 and $D_{3,6}^3$.

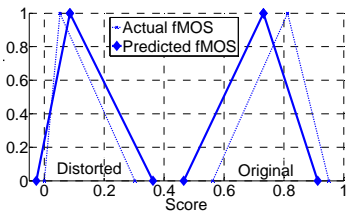


Fig. 4d Fuzzy MOS for O^4 and $D_{4,6}^4$.

C. Model generalization capability

We evaluate the generalization capability of the fuzzy regression models in predicting MOS in order to estimate the model accuracies with respect to untrained datasets. Each distortion type was considered individually. Leave-one-out cross-validation (LOOCV) was conducted, in order to validate the generalization capability of statistical and fuzzy regression models, where it uses one single observation as the validation set and the remaining observations as the training set. LOOCV is repeated for each of the observations as a validation set and for the remaining observations as the training set. The MAE and PLCC are again used as performance indicators.

Three validations with three conditions in the MOS data sets were performed in order to evaluate whether fuzzy regression outperforms statistical regression in term of generalization capability: 1) varying number of participants contributing to the MOS; 2) varying amount of MOS data; and 3) different amount of fuzziness of the MOS data. These three conditions simulated the real situations in subjective image quality experiments where insufficient numbers of participants are involved; an insufficient amount of MOS data is used for modelling; and different amounts of fuzziness are caused by human evaluation of MOS. These validations further determine whether fuzzy

regression outperforms statistical regression when the assumption of normal distribution in MOS data cannot be ensured and also small and uncertain data sets are used for modelling.

1) Varying number of participants contributing to the MOS

We evaluate whether the number of participants influences the relative performance of fuzzy regression compared with statistical regression, as the number of participants affects the standard error to the mean of the IQA, and thus the standard error affects the confidence intervals of the statistical regression on estimating the MOS. When the number of participants is large, the standard error is generally small. When the number of participants is small, the standard error is generally large. Based on the data sets generated by different numbers of participants, we can evaluate whether fuzzy regression can outperform statistical regression in different standard errors. For this purpose, we compute the conventional MOS and fuzzy MOS based on 5 participants and 14 participants. We note that we were limited to a maximum of 14 participants as that was the highest number of ratings for a subset of the images.

In the LOOCV, the MOS of 22 original images and their corresponding distorted images are used for training, and the remaining image and its corresponding distorted images are used for validation. The MAE and PLCC are again used as performance indicators. Figures 5 and 6 show the results for the MOS and PLCC based on 5 and 14 participants, respectively. These results confirm that the fuzzy regression performs consistently better than the statistical regression on a validation set. This holds across a range of distortion types and is independent of the number of participants involved to compute the MOS and PLCC.

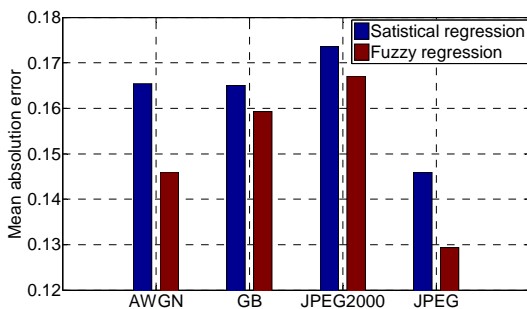


Fig. 5a: MAE for MOS based on 5 participants

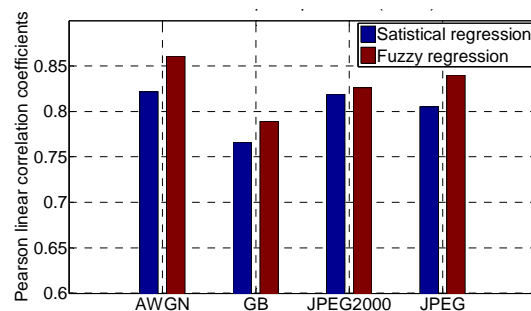


Fig. 5b: PLCC for MOS based on 5 participants

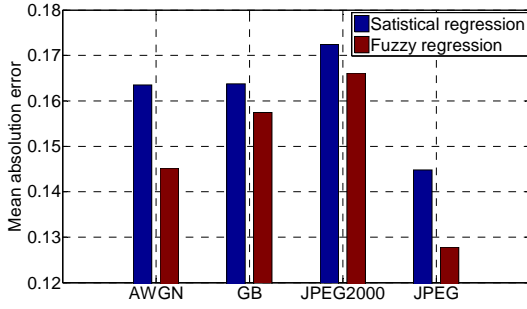


Fig. 6a: MAE for MOS based on 14 participants

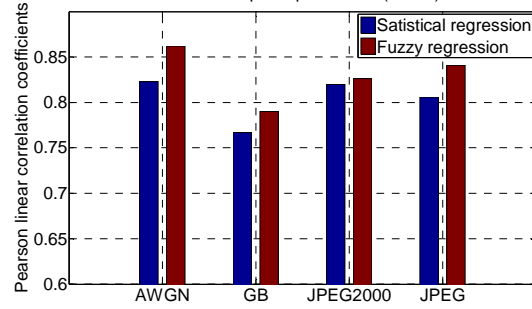


Fig. 6b: PLCC for MOS based on 14 participants

2) Varying amount of MOS data

We evaluate whether different MOS data sizes impact on the training of the models. It attempts to validate whether the fuzzy regression outperforms the statistical regression when small or large MOS data sizes were used on developing the fuzzy regression models. In particular, the assumption of normal distribution of the MOS data sets cannot be ensured when the data sizes are small. Two MOS data sets with two difference sizes were used to develop the MOS models. The four distortion types were considered individually.

The first MOS data set comprised the first 12 original images, O^i with $i=1,2,\dots,12$, and the corresponding corrupted images, $D_{j,k}^i$, distorted from O^i with the six distortion levels, $k=1,2,\dots,6$ and the four distortion types, $j=1,2,\dots,4$. Hence, the first MOS data size is 84, comprising 12 original images and 72 distorted images; the second MOS data set comprised all the 23 original images, O^i , and the 138 corresponding distorted images $D_{j,k}^i$ with $i=1,2,\dots,23$. Hence, the second MOS data size is 161, comprising 23 original images and 138 distorted images. LOOCV was conducted to validate the generalization capability of the MOS models when the three MOS data sets are used.

Figures 7 and 8 show the results for the first and second MOS data sets respectively, where the number of MOS data for the first set is smaller than that of the second set. Figures 7a and 7b show respectively that the MAEs obtained by the fuzzy regression are generally smaller than those of the statistical regression for the four distortion types and the PLCCs obtained by the fuzzy regression are generally larger than those of the statistical regression. Figure 8 shows similar results in that the smaller MAEs and larger PLCCs were obtained by the fuzzy regression for

the four distortion types. Hence, better generalization capabilities can be achieved by the fuzzy regression than those achieved by the statistical regression, when the smaller and larger MOS data sizes were used.

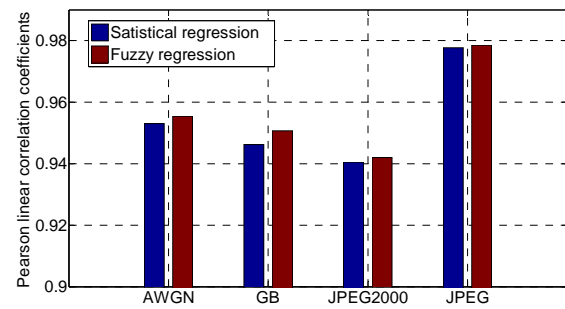
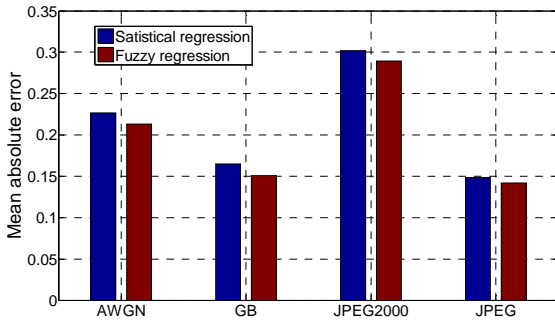


Fig. 7a: MAE for the second MOS data set with 84 pieces of MOS data

Fig. 7b: PLCC for the second MOS data set with 84 pieces of MOS data

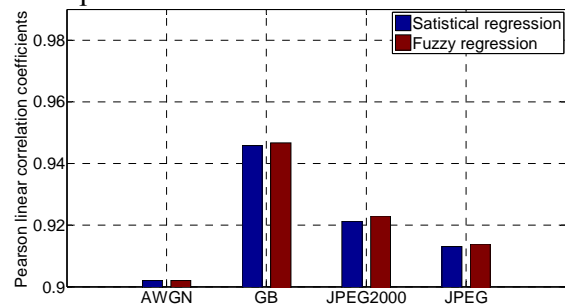
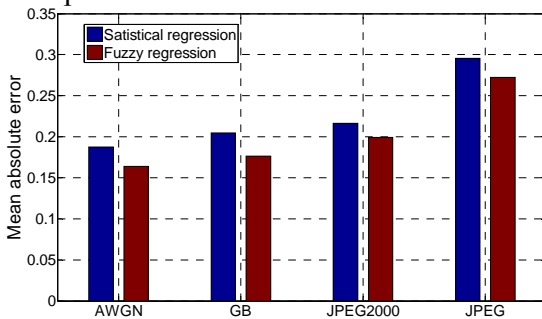


Fig. 8a: MAE for the third MOS data set with 161 pieces of MOS data

Fig. 8b: PLCC for the third MOS data set with 161 pieces of MOS data

3) *Difference amount of fuzziness involved on the MOS data*

We evaluate whether the performance of the fuzzy regression is affected by different amounts of fuzziness according to the psychometric quality scale. It attempts to validate whether the fuzzy regression can outperform statistical regression when different amount of fuzziness was involved on the MOS data. The four different levels of fuzziness with $f = 0.15, 0.20, 0.25, 0.28$ were considered in the 5-point psychometric quality scale. In the LOOCV, the MOS of 22 original images and their corresponding distorted images are used for training and the remaining image and its corresponding distorted images are used for validation. Figures 9, 10, 11 and 12 show the results for the models addressing the four types of distortion namely AWGN, JPEG2000, JPG and GB distortion, when different fuzziness was considered in the 5-point psychometric quality scale. Figures 9a and 9b show the results for the AWGN distortion. They show respectively that the MAEs obtained by the fuzzy regression are generally smaller than those of the statistical regression for the four distortion types, and the PLCCs obtained by the fuzzy regression are generally larger than those of the statistical regression. Figures 10, 11 and 12 show the results for the JPEG2000,

JPG and GB distortion respectively. Similar results can be found in the figures where smaller MAEs and larger PLCCs are obtained by the fuzzy regression models. Hence, better generalization capabilities can be achieved by the fuzzy regression than those achieved by the statistical regression, when different amounts of fuzziness are accounted on the 5-point psychometric quality scale.

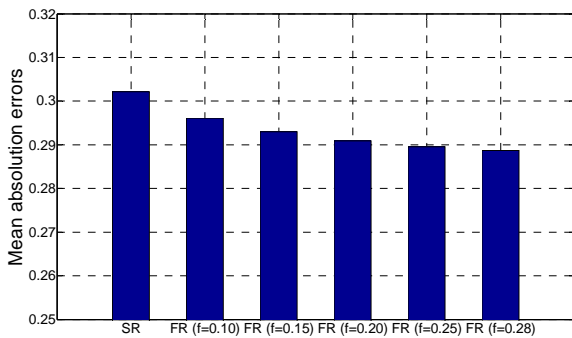


Fig. 9a: MAE for AWGN distortion when different amounts of fuzziness are considered

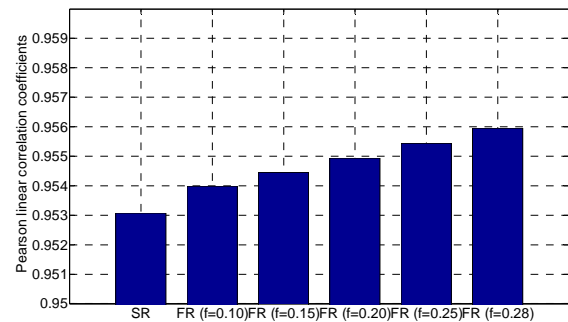


Fig. 9b: PLCC for AWGN distortion when different amounts of fuzziness are considered

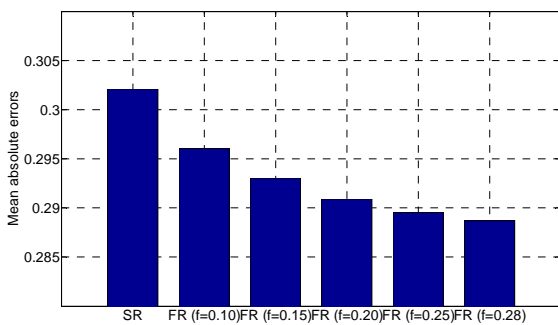


Fig. 10a: MAE for JPEG2000 distortion when different amounts of fuzziness are considered

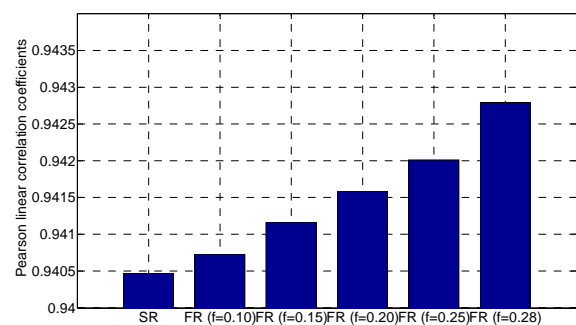


Fig. 10b: PLCC for JPEG2000 distortion when different amounts of fuzziness are considered

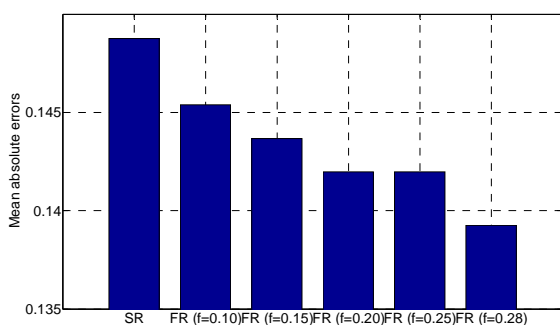


Fig. 11a: MAE for JPG distortion when different amounts of fuzziness are considered

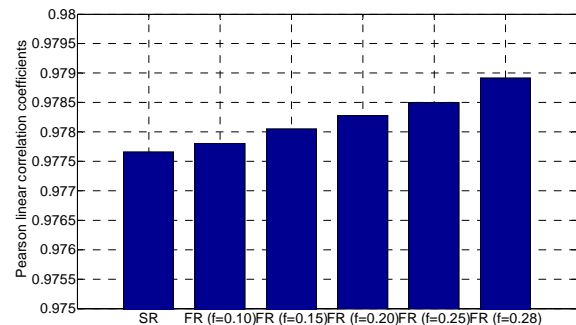


Fig. 11b: PLCC for JPG distortion when different amounts of fuzziness are considered

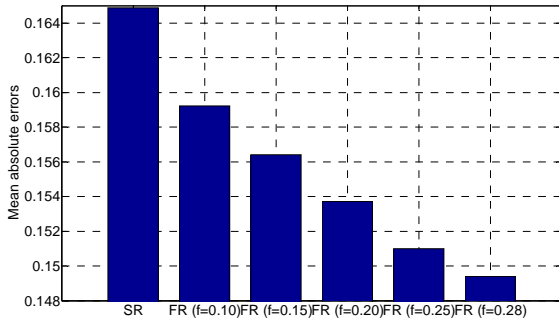


Fig. 12a: MAE for GB distortion when different amounts of fuzziness are considered

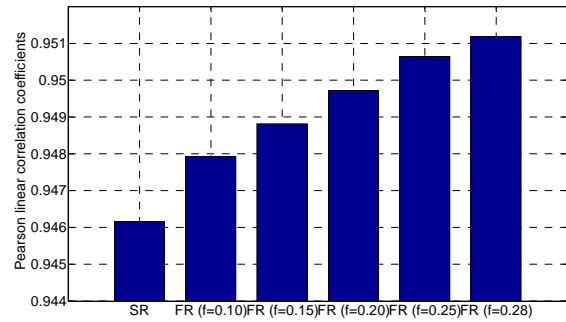


Fig. 12b: PLCC for GB distortion when different amounts of fuzziness are considered

V. CONCLUSION AND FUTURE WORK

In this article, we present a fuzzy regression method to address fuzziness in subjective IQA, which is neglected in current research that considers only the crisp values for MOS. We demonstrate how the fuzziness in the MOS can be predicted from the fuzzy regression models. Also, we have benchmarked the fuzzy regression against the commonly used statistical regression method. MOS data was collected under three different: 1) different numbers of participants; 2) different sizes of the MOS data; and 3) varying amounts of fuzziness. These three conditions usually exist in the collection of the MOS data as the number of participants, data sizes and fuzziness of the data are correlated to the collection cost of MOS data. The validation results indicate that better data fitting and generalization capabilities can be achieved with fuzzy regression models. Therefore, the fuzzy regression outperforms statistical regression when small data sets containing fuzziness are used for modelling the MOS.

In future work, we will further validate the generalization capability of the fuzzy regression using a mix of several image databases which have less distortion bias, and larger samples of image data. Based on the mixed database, more convincing validation can be conducted. We will also develop an image quality database based on the quantity and the fuzziness for the subjective image quality experiments. Using this image quality database, the effectiveness of the fuzzy regression can be further validated when the inconstant fuzziness is used. Also, we will study the correlation between fuzziness in human quality assessment and human observer confidence, and a two-dimensional fuzzy regression will be developed based on interval type-two fuzzy mechanisms.

REFERENCES

- [1] D.G. Bails and L.C. Peppers, *Business Fluctuations: Forecasting Techniques and Applications*, Prentice-Hall, New York, 1982.
- [2] Y.H. Chang, Hybrid fuzzy least squares regression analysis and its reliability measures, *Fuzzy Sets and Systems*, vol. 119, pp. 225-246, 2001.
- [3] L.H. Chen, C.C. Hsueh, and C.J. Chang, A two-stage approach for formulating fuzzy regression models, *Knowledge-Based Systems*, vol. 52, pp. 302-310, 2013.
- [4] I. De, and J. Sil, No-reference image quality assessment using interval type 2 fuzzy sets, *Applied Soft Computing*, vol. 30, pp. 441-453, 2015.
- [5] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Mathematics in Science and Engineering, Vol. 144. Academic Press, New York, 1980.
- [6] U. Engelke, T.M. Kusuma, H. Zepernick and M. Caldera, Reduced-reference metric design for objective perceptual quality assessment in wireless imaging, *Signal Processing: Image Communication*, vol. 24, no. 7, pp. 525-547, 2009.
- [7] U. Engelke and H. Zepernick, An artificial neural network for quality assessment in wireless imaging based on extraction of structural information, *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 1249-1252, 2007.
- [8] A.M. Eskicioglu and P.S. Fisher, Image quality measures and their performance, *IEEE Transactions on Communication*, vol. 43, no. 12, pp.2959–2965, 1995.
- [9] K. Gu, G. Zhai, X.K. Yang and W.J. Zhang, Deep learning network blind image quality assessment, *IEEE International Conference on Image Processing*, pp. 511-515, 2014.
- [10] B. Hesehmatty and A. Kandel, Fuzzy linear regression and its applications to forecasting in uncertain environment, *Fuzzy Sets and Systems*, vol. 15, pp. 159-191, 1985.
- [11] Y. He, Y. Xuan, W. Chen, and X. Fu, Subjective image quality assessment: A method based on signal detection theory, *IEEE International Conference on Systems, Man and Cybernetics*, pp. 4915-4919, 2009.

- [12] ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union/ITU Radiocommunication Sector, January 2002.
- [13] H. Jiang, C.K. Kwong, W.H. Ip, Z. Chen, Chaos-Based Fuzzy Regression Approach to Modeling Customer Satisfaction for Product Design, *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 5, pp. 926-936, 2013.
- [14] J.D. Jobson, *Applied Multivariable Data Analysis*, Springer Verlag, New York, 1991.
- [15] E.E. Karsak, Robot selection using an integrated approach based on quality function deployment and fuzzy regression, *International Journal of Product Research*, vol. 46, no. 3, pp. 723-738, 2008.
- [16] K.J. Kim, H. Moskowitz and M. Kokaslan, Fuzzy versus statistical linear regression, *European Journal of Operational Research*, vol. 92, pp. 417-434, 1996.
- [17] D.G. Kleinbaum and L.L. Kupper, *Applied regression analysis and other multivariable methods*, Duxbury Press, 1978.
- [18] C.K. Kwong, Y. Chen, K.Y. Chan, and X. Luo, A generalised fuzzy least-squares regression approach to modelling relationships in QFD, *Journal of Engineering Design*, vol. 21, no. 5, pp. 601-613, 2010.
- [19] R. Likert, A technique for the measurement of attitudes, *Archives of Psychology*, vol. 22, no. 140, pp. 55, 1932.
- [20] M. Mares, *Computation over Fuzzy Quantities*, CRC Press Inc., 1994.
- [21] J.B. Martens, Multidimensional modeling of image quality, *Proceedings of IEEE*, vol. 90, no. 1, pp.133-153, 2002.
- [22] M. Miyahara, K. Kotani and V.R. Algazi, Objective picture quality scale (PQS) for image coding, *IEEE Transactions on Communications*, vol. 46, no. 9, pp. 1215-1226, 1998.
- [23] A. Mittal, G.S. Muralidhar, J. Ghosh and A. Bovik, Blind image quality assessment without human training using latent quality factors, *IEEE Signal Processing Letters*, vol. 19, no. 2, pp. 75-78, 2012.
- [24] A.F. Molisch, *Wireless Communications*, Wiley, IEEE Press, New York, 2005.
- [25] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics, *Advances of Modern Radio electronics*, vol. 10, pp. 30-45, 2009., <http://www.ponomarenko.info/tid2008.htm>

- [26] Popphoto, How many photos are uploaded to the internet every minute? PopPhoto.com, May 27, 2013.
(<http://www.popphoto.com/news/2013/05/how-many-photos-are-uploaded-to-internet-every-minute>).
- [27] A.W. Rix, A. Bourret, M.P. Hollier, Models of human perception, *BT Technology Journal*, vol. 17, no. 1, pp. 24–34, 1999.
- [28] G.A.F. Seber, *Linear regression analysis*, Wiley, 2003.
- [29] Z. Sener and E.E. Karsak, A decision model for setting target levels in quality function development using nonlinear-programming-based fuzzy regression and optimization, *International Journal of Advanced Manufacturing Technologies*, vol. 48, pp. 1173-1184, 2010.
- [30] Z. Sener and E.E. Karsak, A combined fuzzy linear regression and fuzzy multiple objective programming approach for setting target levels in quality function deployment, *Expert Systems with Applications*, vol. 38, pp. 3015-3022, 2011.
- [31] H.R. Sheikh, Z. Wang, L. Cormack and A.C. Bovik, LIVE Image Quality Assessment Database Release 2, <http://live.ece.utexas.edu/research/quality>.
- [32] . Tanaka and J. Watada, Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets and Systems*, vol. 272, pp. 275-289, 1998.
- [33] Q. H. Thu, M. N. Garcia, F. Speranza, P. Corriveau, and A. Raake, Study of rating scales for subjective assessment of high-definition video, *IEEE Trans. on Broadcasting*, vol. 57, no. 1, pp. 1-14, Mar 2011.
- [34] B.A. Wandell, *Foundations of Vision*, Sinauer Associates, 1995.
- [35] J. Wu, W. Lin, G. Shi, and A. Liu, , Reduced-reference image quality assessment with visual information fidelity, *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700-1705, 2013.
- [36] L.A. Zadeh, The Concept of a linguistic variable and its application to approximate reason – I, *Information Science*, vol. 8, pp. 199-249, 1975.
- [37] A. Zaric, N. Tatalovic, N. Brajkovic, H. Hlevnjak, M. Loncaric, E. Dumic and S. Grgic, VCL@FER Image Quality Assessment Database, *Automatika*, vol. 53, no. 4, pp. 344-354, 2012.