# Information Distance based Photoshop Metric

Nima Nikvand & Zhou Wang
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada
nnikvand@uwaterloo.ca
zhou.wang@uwaterloo.ca

Wisam Farjow & Xavier Fernando
Department of Electrical & Computer Engineering
Ryerson University
Toronto, Ontario, Canada
wfarjow@hotmail.com
fernando@ee.ryerson.ca

***Abstract*—Objective predictor of perceived modifications to images, or the so-called Photoshop metric, finds many applications and is particularly desirable in the fashion industry. A promising solution to the problem is information distance, which measures the minimal number of bits required to transform one object to another. The application of information distance measures to photo editing assessment, however, is perplexed by human visual system characteristics, such as the difference in attention levels to different parts of an image, and the varying sensitivity at detecting different types of image distortions. Here we make one of the first attempts to develop an information distance based measure for Photoshop metric, where information distance-based features are used to train a Support Vector Regressor (SVR). Experimental results show that the proposed metric is well correlated with mean human observer ratings.**

## I. Introduction

Editors of fashion magazines have long been criticized for publishing altered images of celebrities and fashion models, often making them look flawlessly attractive. Such modifications, while appealing to the average reader, have been linked to mental and physical disorders in younger generations, and have given rise to calls by politicians in many countries such as UK, France, and Israel to provide labels for altered images of this kind [1]–[3]. In similar cases, Photoshop scandals of celebrities on the Internet have shocked Hollywood community, and have raised the question of how much editing is considered too much in this business.

Inspired by dire need for an automatic rating system of photoshopped images for the fashion industry, Farid and Kee proposed an algorithm which represents human perception of photo manipulations [4]. In this work, we propose a different approach by introducing a set of information theoretic features which are capable of predicting a perceptual score for a pair of original and modified images, such that the predicted score is correlated with the human perception of modifications.

Information distance features are based on measuring the minimal number of bits required to transform the original image to the edited image and vice versa. While these features are capable of describing the modifications in detail, in some cases, the Human Visual System (HVS) does not distinguish subtle modifications that might translate into a long information distance that require a large number of bits to describe the modifications. In our previous works we have shown promising applications of information distances in image similarity and quality assessment [5], [6] Here we opt to a machine learning algorithm built upon such features and trained by a human observer rated database to predict human ratings. In effect, the machine learning algorithm gives the information distance features the flexibility they require to predict scores that are not only consistent with the amount of modifications, but also consider characteristics of the HVS in distinguishing these modifications.

## II. Related Work

Kee and Farid proposed a measure that quantifies image modifications perceptually [4]. Eight summary statistics features extracted from 468 pairs of original and altered images rated by human observers were used to train a Support Vector Regressor (SVR) to predict the Mean Opinion Scores (MOS) of the images. The features are divided into two groups, and are intended to measure the impact of geometric modifications and photometric modifications on the modified picture.

Geometric distortions between the original and edited images are modeled using a 6-parameter affine model along with two extra parameters to model the brightness and contrast changes [7]. The photometric features are designed to capture the effect of sharpening or blurring, and other structural distortions that might occur during modification of images. In order to quantify the effect of photometric modifications that are not captured by previous steps, SSIM [8] is calculated on the luminance channel of the warped before and after images, and the mean and standard deviation of SSIM are used as two additional features which embody basic blurring, sharpening and special effects by various photoshop filters [4]. These features are then scaled into the range [-1,1] and are fed into a nu-SVR tool with Gaussian radial basis kernel [9] along with human observer ratings for training. The training set consists of a diverse set of original and retouched images collected from online sources. 390 users were paid to rate these images on a scale of 1 to 5 on Amazon's Mechanical Turk website. Each user was shown a total of 50 images, including a random set of five images three times to measure the consistency of responses. The SVR parameters $\gamma$ and $c$ are fine tuned using a 2D grid search to maximize the correlation coefficient of each training set. The algorithm is tested using leave-one-out cross-validation by training the SVR using 467 images and predicting the score for the remaining image. This process is repeated 468 times and the score for each image is predicted. The authors report 80% correlation between the predicted scores and the mean observer scores with a mean/median absolute error of 0.3/0.24, a maximum absolute error of 1.19 and a standard deviation of 0.249.

The implementation of the aforementioned algorithm was not published. Two groups of researchers attempted to reproduce the results but with limited success. The first group applied a reimplementation of the algorithm to a subset of images in the database due to computational limitations. Their result has less than six percent correlation between the predicted scores and mean observer ratings using 5-fold cross validation [10]. The second group developed a dataset of 137 images and applied a modified version of the original algorithm. They report a 65.2 percent correlation between the predicted scores and mean observer ratings [11].

Both groups published their reimplementations online. Applying them to the original dataset using leave-one-out cross-validation, we obtain a correlation of less than 30% with the mean observer ratings in both cases. We have also attempted to reimplement the original algorithm and obtained similar results. Therefore, we modify upon the algorithm by enforcing symmetry in learning the filter used for feature extraction, and by projecting the vector fields onto a gradient of luminance channel before computing the related features. The predicted scores of our modified algorithm have 57.4% correlation with the mean observer ratings, with a mean/median absolute error of 0.42/0.37, a maximum error of 1.7063, and a standard deviation of 0.3.

## III. Information Distance Method

### A. Feature Extraction

In this section, we introduce a set of information distance based features which will be used for training the SVR using human observer data.

The first set of features are selected to measure the effect of geometric modifications. We start by aligning the before and after images using the registration technique [4]. Assuming that $f_a$ and $f_b$ are local regions of the luminance channel of the original and edited images, respectively, we have:

$$cf_a(x,y)+b=f_b(m_1x+m_2y+t_x,m_3x+m_4y+t_y) \quad (1)$$

where the $m_i$ terms are affine parameters, and $c$ and $b$ are contrast and luminance change parameters. Using this registration, a two dimensional vector field of geometric transformations can be built as follows [7]:

$$\overrightarrow{v}(x,y)=\begin{pmatrix} m_1x+m_2y+t_x-x \\ m_3x+m_4y+t_y-y \end{pmatrix}. \quad (2)$$

The mean and standard deviation of the magnitude of the vector field in (2) projected onto the gradient vector field of the image, and computed over the face and body regions are taken to be the geometric features. These features are defined for face, hair, and body regions respectively, and are computed using motion vectors in the direction of $X$ and $Y$ extracted from the vector field (2). The motion vectors are inserted into column vectors of a matrix $M$ such that:

$$\mathrm{M}=\begin{pmatrix} M_x & M_y \end{pmatrix}. \quad (3)$$

Assuming that motion vectors in $X$ and $Y$ directions are following a joint Gaussian distribution, $M_x, M_y \sim N(\mu_M, \Sigma_M)$ we have:

$$h(M_x,M_y)=\frac{1}{2}\log(2\pi e)^2|\Sigma_M|, \quad (4)$$

where $|\Sigma_M|$ is the determinant of the covariance of $M$, and we take $\log(|\Sigma_M|)$ computed over face, body, and hair regions as our geometric features.

The aligned images are then transformed into the $La^*b^*$ color space, [12], and the second set of features are selected to be $La^*b^*$ differential entropy features. These features are designed to quantify modifications to $La^*b^*$ pixel domain such as color change, noise and structural changes not captured by registration. These features are defined for face, body and hair separately by subtracting each channel of before and after images directly and inserting it into column vectors $\Delta L$, $\Delta a^*$, and $\Delta b^*$. The vectors are then inserted in the columns of a matrix $\Delta$:

$$\Delta=\begin{pmatrix} \Delta L & \Delta a^* & \Delta b^* \end{pmatrix}. \quad (5)$$

Since the differential entropy of joint Gaussian distribution is a higher bound for the differential entropy of all distributions of the same covariance structure [13], we assume that the observation vectors follow a joint Gaussian distribution. In this case the differential entropy of a joint Gaussian distribution $\Delta L, \Delta a^*, \Delta b^* \sim N(\mu_\Delta, \Sigma_\Delta)$ is:

$$h(\Delta L,\Delta a^*,\Delta b^*)=\frac{1}{2}\log(2\pi e)^3|\Sigma_\Delta|, \quad (6)$$

where $\Sigma$ represents the covariance matrix of $\Delta$, and $|\Sigma_\Delta|$ represents its determinant. We take $\log(|\Sigma_\Delta|)$ computed over the regions of face, hair and body, respectively, as our $La^*b^*$ entropy features.

### B. Results and Discussion

Both sets of features are carefully selected to quantify a higher bound on the average number of bits required to describe the modifications carried out on the original image to reach the edited image. While these features show some correlation with the mean observer ratings, it must be taken into consideration that the visual system may not react evenly across different types of image modifications in proportion to the mathematical information distance measures. For example, certain regions such as human faces may attract more visual attention, and some subtle changes in image may not be noticeable by the visual system. Specifically, we used a nu-SVR with Gaussian kernel implemented in LIBSVM [9] to train a predictor using information distance features as the input. The parameters, $\gamma$ and $c$ were selected to maximize the correlation of the predicted scores and the mean observer ratings in both leave-one-out cross-validation and 1-fold cross-validation (train-all/test-all) schemes by an exhaustive search over a range of possible values. The predicted scores have 76 percent correlation with mean observer ratings in this case. Figures 1 and 2 show examples of the original and edited images with their corresponding mean observer scores and predicted scores by the proposed method. It can be seen that the proposed method is capable of predicting the scores with higher accuracy when the nature of modifications are geometric, and the performance degrades when more complicated editing filters are used. Figure 3 shows the result of leave-one-out cross validation of the method on (a) 234 and (b) 100 randomly selected images. The Pearson linear correlations between the predicted scores and mean observer ratings in this case are 72% and 74%, respectively. The algorithm was also verified by randomly dividing the dataset into training and testing sets, and averaging the correlations of the predicted scores with mean observer scores for 1000 times. The average correlation of the predicted scores with MOS in this case is 65% with a standard deviation of 13%.

(MOS: 3.78 / P: 3.78) (MOS:3.64 / P:3.89) (MOS: 3.42 / P: 3.42)

Fig. 1: Good Examples: Geometric modifications

In comparison, the prediction performance of the proposed method built upon information distance features is significantly better than our reimplementation of the algorithm described in [4], and is similar to that reported in [4]. Analysis shows that the proposed method and the method in [4] are both overcompensating when observer scores are in the lower range and undercompensating when the observer scores are in the higher range. The results of both methods stay in the range of plus/minus one standard deviation of the observer ratings. The distribution of mean observer scores and the predicted scores by individual observer ratings are shown in Figure 4. As expected, the density of individual observer ratings are higher in the intervals where the mean observer ratings and the predicted ratings are consistent. This suggests the soundness of both the subjective test and the predicted results.

(MOS: 4.62 / P: 3.31) (MOS:3.56 / P:2.53) (MOS: 4.08 / P: 2.89)

Fig. 2: Bad Examples: Photometric modifications

## IV. Conclusion

We propose a set of information distance features, based on which we develop a perceptual image similarity measure for photo retouching applications. The features were tested on an observer rated dataset of original and modified images. It was shown that the features are effective at predicting the perceptual distortions of retouched images. Compared with the existing method [4], the proposed method has significantly lower complexity in feature extraction and competitive performance in prediction accuracy.

## Acknowledgment

Authors would like to thank NSERC, SOSCIP, and Ontario Centres of Excellence for their continued support and funding of this project.

## References

[1] J.K. Thmpson and L.J. Heinberg, “The media’s influence on body image disturbance and eating disorders: We’ve reviled them, now can we rehabilitate them?,” *Journal of Social Issues*, vol. 55, pp. 339–353, Summer 1999.

[2] P.N. Myers and F.A. Biocca, “The elastic body image: The effect of television advertising and programming on body image distortions in young women.,” *Journal of Communication*, vol. 42, pp. 108–133, 1992.

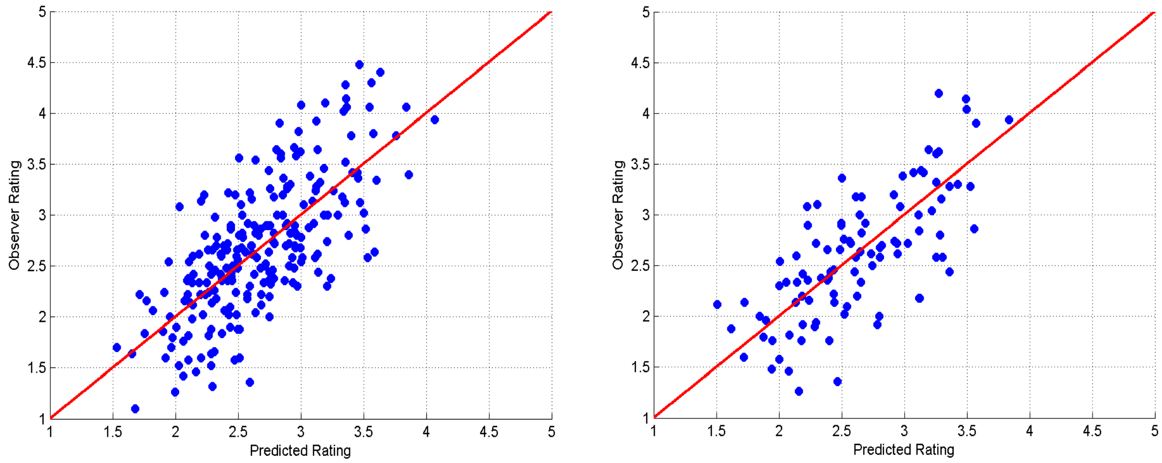
[3] D. Smeesters, T. Mussweiler, and N. Mandel, “The effects of thin and heavy media images on overweight and underweight consumers: social comparison processes and behavioral implications.,” *Journal of Consumer Research*, vol. 36, pp. 930–949, Apr. 2010.

[4] E. Kee and H. Farid, “A perceptual metric for photo retouching,” *Proceedings of the National Academy of Sciences*, vol. 108, November 2011. http://www.pnas.org/content/early/2011/11/21/1110747108.abstract.

[5] N. Nikvand and Z. Wang, “Generic image similarity based on kolmogorov complexity,” in *Proceedings of International Conference on Image Processing (ICIP) 2010*, 2010.

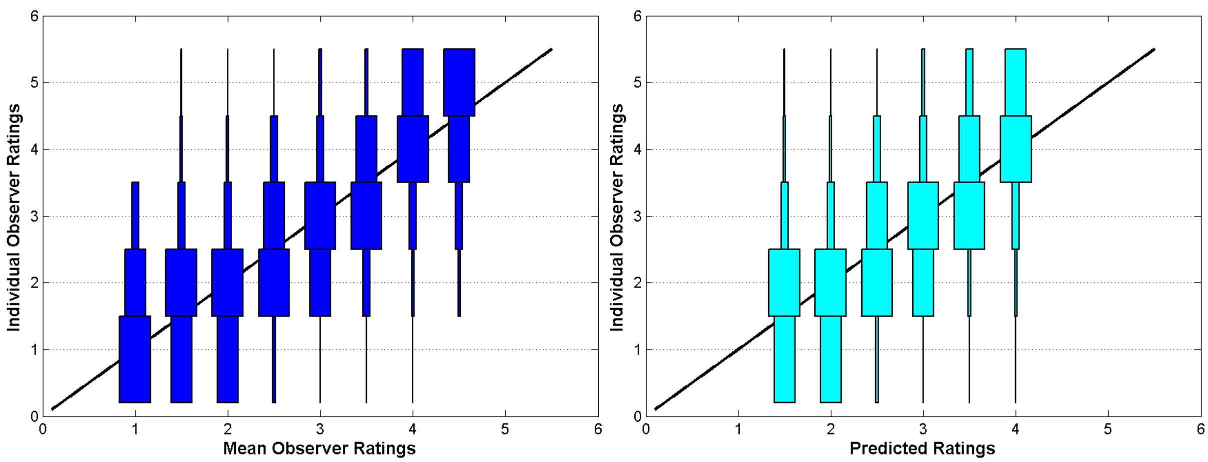
[6] N. Nikvand and Z. Wang, “Image distortion analysis based on normalized perceptual information distance,” *Signal, Image and Video Processing*, vol. 7, no. 3, pp. 403–410, 2013.

[7] S. Periaswamy and H. Farid, “Elastic registration in the presence of intensity variations,” *IEEE Trans. Medical Imaging*, vol. 22, pp. 865–874, July 2003.




(a) (b)

Fig. 3: Leave-one-out cross-validation on a set of randomly selected images: (a) 234 images with 72% correlation (b) 100 images with 74% correlation




(a) (b)

Fig. 4: Distributions: (a) Mean observer ratings vs. individual observer ratings; (b) Predicted ratings vs. individual observer ratings

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.
[9] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[10] B. Collins, A. Danowitz, and A. Zvinakis, "Photo retouching metric," 2012. http://white.stanford.edu/teach/index.php/CollinsZvinakisDanowitz.
[11] A. Gellineau, J. Huang, and C. Lee, "Photo retouching metric," 2012. http://white.stanford.edu/teach/index.php/GellineauHuangLee.
[12] M. Tkalcic and J. F. Tasic, "Colour spaces: perceptual, historical and applicational background," vol. 1, pp. 304–308 vol.1, IEEE, 2003.
[13] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.