Use of Face Detection to Qualify Image Processing Algorithms

Steven Simske, Hewlett-Packard Labs, 3404 E. Harmony Rd., MS 36, Fort Collins CO 80528, USA
Dalong Li, Hewlett-Packard Co., 11445 Compaq Center Dr W, Houston TX 77070, USA
Darryl Greig, Hewlett-Packard Labs, Longdown Avenue, Stoke Gifford, Bristol BS34 8QZ UK

Abstract

Image processing algorithms always affect the quality of the image to which they are applied. Traditionally, the effects of these algorithms are determined using broad image quality measurements such as PSNR, SSIM, VIF and MSE. These quantitative measures, however, do not provide a functional assessment of the image processing algorithms. In this paper, we propose and introduce the use of facial detection as such a functional measurement of image quality after performing image processing transformations on the image. To assess this, we downsampled images containing a combined 953 faces, then upsampled the images using five different image processing operations including interpolation and filtering with different parameters. This process resulted in the misdetection of 12-20 images (and a 14-21% increase in error rate) in comparison to the original set, illustrating the deleterious effect of downsampling and subsequent upsampling on the images.

Introduction

Image transformations are a critical element in most modern printing and imaging systems. Transformations such as lossless and lossy compression, interpolation and super-resolution are necessary for the raster image processing (RIP) unit of the digital printing system to create the appropriate set of instructions for transferring the media (ink, toner, etc.) to the substrate. On the imaging side, different capture devices and technologies—line and 2D scanners, mobile phone cameras and other digital cameras, inspection cameras, surveillance cameras, etc.—have different imaging capabilities; that is, different color, modulation transfer function, contrast, vignetting, etc.

In order to predict or measure print quality, a variety of image quality (IQ) metrics have been defined. Most of these provide a **quantitative measure** of IQ, and variants on these measurements are often touted as better representing the relative rating that would be given by human IQ experts. However, we believe that replacing such IQ measurements with a **qualitative measure** provides a more direct analogy to how a human IQ expert would relatively rate the images. In order to qualitatively relatively compare different image sets, therefore, we introduce the use of face detection as an image inspection method.

Signal-to-Noise and Inspection Methods

Many image processing algorithms are assessed based on quantitative evaluations of the image. These include peak signal to noise ratio (PSNR) of the image in addition to image inspection techniques such as structural similarity (SSIM), Visual Image Fidelity (VIF) and Mean Squared Error (MSE). These metrics [1][2] are broad, taking into account the entire image at once. For example, PSNR is defined by:

$$PSNR(\hat{f}) = 10\log_{10}\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}255^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(f(i,j)-\hat{f}(i,j))^2}$$

where $\hat{f}(i,j)$ is the transformed image, and $f(i,j)$ is the original high-resolution image. The size of the images are M × N.

While these methods provide, in many cases, useful comparative metrics for IQ [3], they do not provide important absolute IQ information. It seems reasonable that a direct measurement of the **functional capacity** of the image will provide a more predictive, quantitative measure of IQ.

Face Detection and Image Processing Transformations

In this study, images are processed as described herein below. First, the images are downsampled by a factor of 2. Anti-aliasing filtering is used before downsampling. From the downsampled image, we next interpolate the image with a bicubic kernel. After the interpolation, we apply super-resolution filtering [3] with various settings regarding the distance metric used in the object function in training the super-resolution filter and the size of the filter. For example, the P1filter33 mentioned in Table 1 implies that the super-resolution filter is trained with the 1-norm distance (city-block or Manhattan distance) and that the size of the filter is 3 by 3. The P1filter33 is the filter that minimized the distance/error between the filtered image and the ground-truth image in terms of 1-norm distance in the training set. Similarly, the P2filter55 means that the distance function is based on 2-norm (Euclidean distance) and the size of filter is 5 by 5. Self training [3] on each image in a dataset can be time-consuming; in this work, to speed up the processing, we pre-trained the filters on the USC-SIPI image dataset [5]. Once trained, these filters were used throughout for the image transformation and filtering.

We believe face detection [4] is a strong candidate for a quantitative measure of IQ for images that include human faces. For our experiments, we considered a data set containing 953 ground-truthed faces; that is, 953 faces were clearly visible to human observation. This is termed ∑P, for the actual number of positives (true faces) in the datasets for Table 2.

A face detection algorithm used elsewhere [4] was used to analyze the images. As shown in Table 1, the face detector was able to positively identify 873 of the 953 (TP, or number of true positives—faces positively identified—is equal to 873 in the **Originals** row and TP column) faces. Only 7 False Positives (FP; non-faces identified as faces) were identified by the face detection algorithm. Since 80 faces in the ground truthed set were not found, the False Negatives, or FN, column has the value 80 in Table 1 for the **Originals** row.

IMAGE SET	FP	FN	TP	TP/∑P	∑P
Originals	7	80	873	0.916	953
Bicubic	9	92	861	0.903	953
P1filtered33	7	95	858	0.900	953
P1filtered55	5	100	853	0.895	953
P2filtered33	12	92	861	0.903	953
P2filtered55	6	94	859	0.901	953
Downsampled	1	93	860	0.902	953

Table 1. Data for the Original, Downsampled and Filtered (Bicubic, P1filtered33, P1filtered55, P2filtered33 and P2filtered55) after downsampled images. FP = false positives (non-faces detected as faces), FN=false negatives (faces not found), TP=true positives (actual faces found), ∑P=sum of positives (actual number of faces), and TP/∑P = percent of true faces actually found. The number of errors (FP+FN) was similar for all five filtered groups, and all of these groups were statistically significantly more error-prone than the Originals. The Downsampled group cannot be compared to the rest since 70 images were too small after downsampling to apply facial recognition.

The ratio TP/(TP+FN), the fourth data column, is therefore 873/953 = 0.916. Thus, the face detector is able to find 91.6% of the original faces. This is traditionally termed recall. The precision, given in Table 2, is TP/(TP+FP), and is 99.2%. Finally, the accuracy, which is the harmonic mean of precision and recall, is 95.2% (Table 2).

These values were computed for the Downsampled images, and again after upsampling these images using five different upsampling approaches (Bicubic, P1filtered33, P1filtered55, P2filtered33 and P2filtered55 columns in Tables 1 and 2). After downsampling, face detection accuracy drops by 0.4%. After subsequent upsampling, another 0.3-0.6% drop in accuracy is observed. This means an increase in the error rate of 8% due to downsampling, and an increase in error rate of from 14%-21% after both downsampling and upsampling. The differences in error rate for the 5 filtering approaches are 17%, 17%, 21%, 19% and 14%, which is 18% ± 3%, a z-score of 6 and thus very significant (less than 1 chance in a billion this difference is due to chance).

IMAGE SET	PRECISION (p)	RECALL (r)	ACCURACY (2pr/[p+r])
Originals	0.992	0.916	0.952
Bicubic	0.990	0.903	0.944
P1filtered33	0.992	0.900	0.944
P1filtered55	0.994	0.895	0.942
P2filtered33	0.986	0.903	0.943
P2filtered55	0.993	0.901	0.945
Downsampled	0.998	0.902	0.948

Table 2. Precision (p = TP/[TP+FP]), Recall (r = TP/∑P) and Accuracy (2pr/[p+r]) for the data in Table 1.

Discussion and Conclusions

These data show that face detection is a reasonable means of assessing IQ. Downsampling followed by upsampling results in a 14-21% increase in face detection error rate. This is a large relative change, and indicates face detection is a promising method for determining IQ in image sets that contain some faces. By extension, analogous methods such as specific shape or pattern detection, and more advanced methods such as face recognition, should also prove valuable as functional measurements for IQ.

In future work, we will evaluate face detection results for other salient image processing transformations, such as super-resolution inverse transformations for improved compression [3]. The face detection algorithm and system [4], however, is only a starting point. A direct comparison of the original images with the images after each of the image transformations – computing precision and accuracy – is an effective means of assessing which image transformations are not harmful to functional imaging; that is, image analysis tasks which rely on image recognition. More importantly, in future work, we will compare the more sensitive functional measurement, face detection, to the usually-employed qualitative measurements, such as PSNR, SSIM and MSE.

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[2] H. R. Sheikh and A. C. Bovik , "A Visual Information Fidelity Approach to Video Quality Assessment" (Invited Paper), The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, January 23-25, 2005.

[3] Li D, Simske SJ: Fast Single Image Super-Resolution by Self-Trained Filtering, ICIC 2011 and Lecture Notes in Computer Science, 2012, Volume 6839/2012, 469-475, DOI: 10.1007/978-3-642-25944-9_61.

[4] Greig D: Video Object Detection Speedup Using Staggered Sampling, IEEE WACV, Snowbird UT, 7-8 Dec 2009, 10.1109/WACV.2009.5403129.

[5] University of Southern California - Signal & Image Processing Institute - The USC-SIPI Image Database, http://sipi.usc.edu/services/database/Database.html, last accessed on 11 June 2012..

Author Biography

Steve Simske is an HP Fellow and the Director and Chief Technologist of the Document Ecosystem portfolio in Hewlett-Packard Labs. Steve is currently on the IS&T Board. He is also an IS&T Fellow and a member of the World Economic Forum's Global Agenda Council on Illicit Trade. Steve has advanced degrees in Biomedical, Electrical and Aerospace Engineering, and has more than 50 granted US patents.