Framework for Applying Full Reference Digital Image Quality Measures to Printed Images

Tuomas Eerola, Joni-Kristian Kämäräinen*, Lasse Lensu, and Heikki Kälviäinen

Machine Vision and Pattern Recognition Research Group (MVPR)
*MVPR/Computational Vision Group, Kouvola
Department of Information Technology
Lappeenranta University of Technology (LUT), Finland
firstname.lastname@lut.fi

Abstract. Measuring visual quality of printed media is important as printed products play an essential role in every day life, and for many "vision applications", printed products still dominate the market (e.g., newspapers). Measuring visual quality, especially the quality of images when the original is known (full-reference), has been an active research topic in image processing. During the course of work, several good measures have been proposed and shown to correspond with human (subjective) evaluations. Adapting these approaches to measuring visual quality of printed media has been considered only rarely and is not straightforward. In this work, the aim is to reduce the gap by presenting a complete framework starting from the original digital image and its hard-copy reproduction to a scanned digital sample which is compared to the original reference image by using existing quality measures. The proposed framework is justified by experiments where the measures are compared to a subjective evaluation performed using the printed hard copies.

1 Introduction

The importance of measuring visual quality is obvious from the viewpoint of limited data communications bandwidth or feasible storage size: an image or video compression algorithm is chosen based on which approach provides the best (average) visual quality. The problem should be well-posed since it is possible to compare the compressed data to the original (full-reference measure). This appears straightforward, but it is not because the underlying process how humans perceive quality or its deviation is unknown. Some physiological facts are know, e.g., the modulation transfer function of the human eye, but the accompanying cognitive process is still unclear. For digital media (images), it has been possible to devise heuristic full-reference measures, which have been shown to correspond with the average human evaluation at least for a limited number of samples, e.g., the visible difference predictor [1], structural similarity metric [2], and visual information fidelity [3]. Despite the fact that "analog" media (printed images) have been used for a much longer time, they cannot overcome certain limitations, which on the other hand, can be considered as the strengths of

A.-B. Salberg, J.Y. Hardeberg, and R. Jenssen (Eds.): SCIA 2009, LNCS 5575, pp. 99–108, 2009.

digital reproduction. For printed images, it has been considered to be impossible to utilise a similar full-reference strategy since the information undergoes various non-linear transformations (printing, scanning) before its return to the digital form. Therefore, the visual quality of printed images has been measured with various low-level measures which represent some visually relevant characteristic of the reproduced image, e.g., mottling [4] and the number of missing print dots [5]. However, since the printed media still dominate in many reproduction forms of visual information (journals, newspapers, etc.), it is intriguing to enable the use of well-studied full-reference digital visual quality measures in the context of printed media.

For digital images, the relevant literature consists of full-reference (FR) and no-reference (NR) quality measures according to whether a reproduced image is compared to a known reference image (FR), or a reference does not exist (NR). Where the NR measures stand out as a very challenging research problem [6], the FR measures are based on a more stronger rationale. The current FR measures make use of various heuristics and their correlation to the human quality experience is tested usually with a limited set for pre-defined types of distortions. The FR measures, however, posses an almost unexplored topic for printed images where the subjective human evaluation trials are often much more general. By closing the gap, completely novel research results can be achieved. An especially intriguing study where a very comprehensive comparison between the state-of-the-art FR measures was performed for digital images was published by Sheikh et al. [7]. How could this experiment be replicated for the printed media?

The main challenges in enabling the use of the FR measures with the printed media are actually those completely missing from the digital reproduction: image correspondence by accurate registration and removal of reproduction distortions (e.g., halftone patterns). In this study, we address these problems with known computer vision techniques. Finally, we present a complete framework for applying the FR digital image quality measures to printed images. The framework contains the full flow from a digital original and printed hard-copy sample to a single scalar representing the overall quality computed by comparing the corresponding re-digitised and aligned image to the original digital reference. The stages of the framework, the registration stage in particular, are studied in detail to solve the problems and provide as accurate results as possible. Finally, we justify our approach by comparing the computed quality measure values to an extensive set of subjective human evaluations.

The article is organised as follows. In Sec. 2, the whole framework is presented. In Sec. 3, the framework is tested and improved, as well as, some full reference measures are tested. Future work is discussed in Sec. 4, and finally, conclusions are devised in Sec. 5.

2 The Framework

When the quality of a compressed image is analysed by comparing it to an original (reference) image, the FR measures can be straightforwardly computed, cf., computing "distance measures". This is possible as digital representations are in correspondence, i.e., there exists no rigid, partly rigid or non-rigid (elastic) spatial shifts between the images and compression should retain photometric equivalence. This is not the case with printed media. In modern digital printing, a digital reference exists, but it will undergo various irreversible transforms, especially in printing and scanning, until another digital image for the comparison is established. The first important consideration is the scanning process. Since we are not interested in the scanning but printing quality, a scanner must be an order of magnitude better than a printing system. Fortunately, this is not difficult to achieve with the available top-quality scanners in which sub-pixel accuracy of the original can be used. It is important to use sub-pixel accuracy because this prevents the scanning distortions to affect the registration. Furthermore, to prevent photometric errors from occurring, the scanner colour mapping should be adjusted to correspond to the original colour map.This can be achieved by using a scanner profiling software that comes along with the high-quality scanners. Secondly, a printed image contains halftone patterns, and therefore, descreening is needed to remove high halftone frequencies and form a continuous tone image comparable to the reference image. Thirdly, the scanned image needs to be very accurately registered with the original image before the FR image quality measures or dissimilarity between the images can be computed. The registration can be assumed to be rigid since non-rigidity is a reproduction error and partly-rigid correspondence should be avoided by using the high scanning resolution.

Based on the above general discussion, it is possible to sketch the main structure for our framework of computing FR image quality measures from printed images. The framework structure and data flow are illustrated in Fig. 1. First, the printed halftone image is scanned using a colour-profiled scanner. Second, the descreening is performed using a Gaussian low-pass filter (GLPF) which produces a continuous tone image. To perform the descreening in a more psychophysically plausible way, the image is converted to the CIE L*a*b* colour space where all the channels are filtered separately. The purpose of CIE L*a*b* is to span a perceptually uniform colour space and not suffer from the problems related to, e.g., RGB where the colour differences do not correspond to the human visual system [8]. Moreover, the filter cut-off frequency is limited by the printing resolution (frequency of the halftone pattern) and should not be higher than 0.5 mm which is the smallest detail visible to human eyes when unevenness of a print is evaluated from the viewing distance of 30 cm [4]. To make the input and reference images comparable, the reference image needs to be filtered with the identical cut-off frequency.

2.1 Rigid Image Registration

Rigid image registration was considered as a difficult problem until the invention of general interest point detectors and their rotation and scale invariant descriptors. These methods provide unparametrised methods which yield accurate and robust correspondence essential for the registration. The most popular method which combines both the interest point detection and description is David Lowe's SIFT [9]. Registration based on the SIFT features has been utilised, for example,

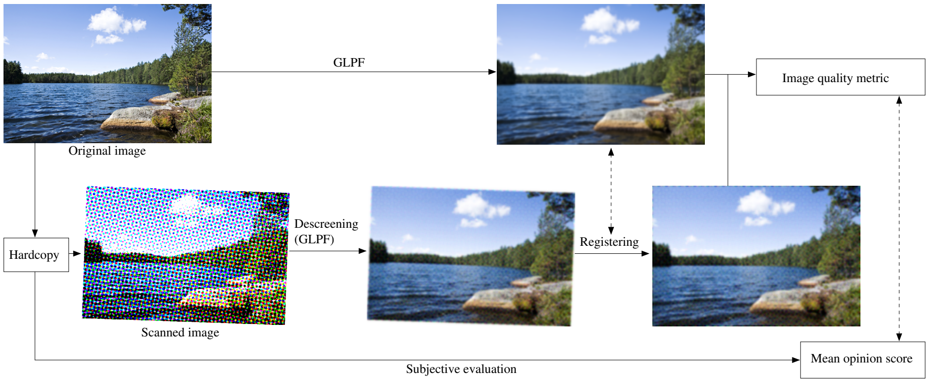


Fig. 1. The structure of the framework and data flow for computing full-reference image quality measures for printed images

in mosaicing panoramic views [10]. The registration consists of 4 stages: extract local features from both images, match the features (correspondence), find a 2D homography between correspondence and finally transform one image to another for comparison.

Our method performs a scale and rotation invariant extraction of local features using the scale-invariant feature transform (SIFT) by Lowe [9]. The SIFT method includes also the descriptor part which can be used for matching, i.e., the correspondence search. As a standard procedure, the random sample consensus (RANSAC) principle presented in [11] is used to find the best homography using exact homography estimation for the minimum number of points and linear estimation methods for all "inliers". The linear methods are robust and accurate also for the final estimation since the number of correspondences is typically quite large (several hundreds of points). The implemented linear homography estimation methods are Umeyama for isometry and similarity [12], a restricted direct linear transform (DLT) for affinity and the standard normalised DLT for projectivity [13]. The only adjustable parameters in our method are the number of random iterations and the inlier distance threshold for the RANSAC which can be safely set to 2000 and 0.7 mm, respectively. This makes the whole registration algorithm parameter free. In image transformation, we utilise standard remapping using bicubic interpolation.

2.2 Full Reference Quality Measures

Simplest FR quality measures are mathematical formulae for computing element-wise similarity or dissimilarity between two matrices (images), such as, the mean squared error (MSE) or peak signal-to-noise ratio (PSNR). These methods are widely used in signal processing since they are computationally efficient and have a clear physical meaning. These measures should, however, be restricted by the known physiological facts to bring them in correspondence with the human visual system. For example, the MSE can be generalised to colour images by computing Euclidean distances in the perceptually uniform CIE L*a*b* colour space as

$$LabMSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [\Delta L^*(i,j)^2 + \Delta a^*(i,j)^2 + \Delta b^*(i,j)^2] \quad (1)$$

where $\Delta L^*(i,j)$, $\Delta a^*(i,j)$ and $\Delta b^*(i,j)$ are differences of the colour components at point (i,j) and M and N are the width and height of the image. This measure is known as the L*a*b* perceptual error [14]. There are several more exotic and more plausible methods surveyed, e.g., in [7], but since our intention here is only to introduce and study our framework, we utilise the standard MSE and PSNR measures in the experimental part of this study. Using any other FR quality measure in our framework is straightforward.

3 Experiments

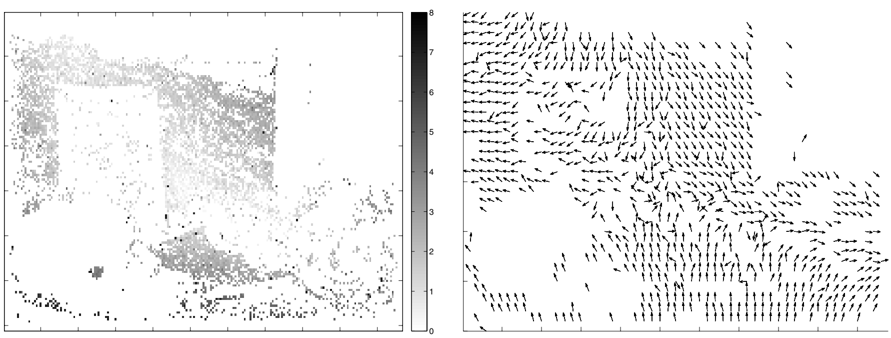
Our "ground truth", i.e., the dedicatedly selected test targets (prepared independently by a media technology research group) and their extensive subjective evaluations (performed independently by a vision psychophysics research group) were recently introduced in detail in [15,16,17]. The test set consisted of natural images printed with a high quality inkjet printer on 16 different paper grades. The printed samples were scanned using a high quality scanner with 1250 dpi resolution and 48-bit RGB colours. A colour management profile was derived for the scanner before scanning, scanner colour correction, descreening and other automatic settings were disabled, and the digitised images were saved using lossless

Fig. 2. The reference image

compression. Descreening was performed using the cut-off frequency of 0.1 mm which was selected based on the resolution of the printer (360 dpi). The following experiments were conducted using the reference image in Fig. 2, which contains different objects generally considered as most important for quality inspection: natural solid regions, high texture frequencies and a human face. The size of the original (reference) image was 2126 × 1417 pixels.

3.1 Registration Error

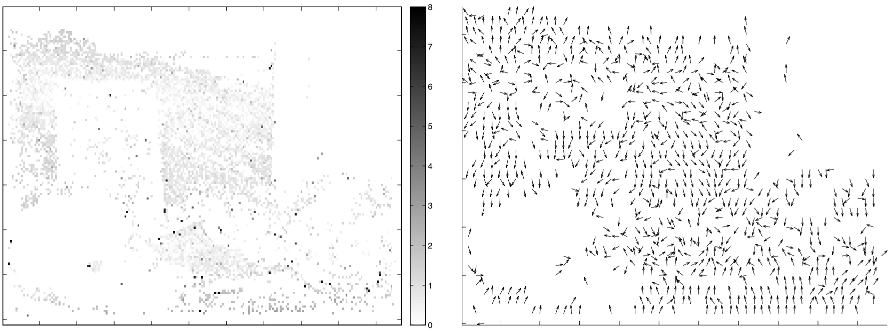
The success of the registration was studied by examining error magnitudes and orientations in different parts of the image. For a good registration result in general, the magnitudes should be small (sub-pixel) and random, and similarly their orientations should be randomly distributed. The registration error was estimated by setting the inlier threshold, used by the RANSAC, to relatively loose and by studying the relative locations of accepted local features (matches) between the reference and input images after registration. This should be a good estimate of the geometrical error of the registration. Despite the fact that the loose inlier threshold causes a lot of false matches, the most of the matches are still correct, and the trend of distances between the correspondence in different parts of the image describes the real geometrical registration error.



(a) (b)

Fig. 3. Registration error of similarity transformation: (a) error magnitudes; (b) error orientations

In Fig. 3, the registration errors are visualised for similarity as the selected homography. Similarity should be the correct homography as in the ideal case, the homography between the original image and its printed reproduction should be similarity (translation, rotation and scaling). However, as it can be seen in Fig. 3(a), the registration is accurate to sub-pixel accuracy only in the centre of the image where the number of local features is high. However, the error magnitudes increase to over 10 pixels near the image borders which is far from sufficient for the FR measures. The reason for the spatially varying inaccuracy



(a) (b)

Fig. 4. Registration error of affine transformation: (a) error magnitudes; (b) error orientations

can be seen from Fig. 3(b), where the error orientations are away from the centre on the left- and right side of the image, and towards the centre on the top and at the bottom. The correct interpretation is that there exists small stretching in the printing direction. This stretching is not fatal for the human eye, but it causes a transformation which does not follow similarity. Similarity must be replaced with another more general transformation, affinity being the most intuitive. In Fig. 4, the registration errors for affine transformation are visualised. Now, the registration errors are very small over the whole image (Fig. 4(a)) and the error orientations correspond to a uniform random distribution (Fig. 4(b)).

In some cases, e.g., if the paper in the printer or imaging head of the scanner do not move at constant speed, registration may need to be performed in a piecewise manner to get accurate registration results. One noteworthy benefit of the piecewise registration is that after joining the registered image parts, the falsely registered images are clearly visible and can be either re-registered or eliminated from biasing further studies. In the following experiments, the images are registered in two parts.

3.2 Full Reference Quality Measures

The above presented experiment was already a proof-of-concept for our framework, but we wanted to briefly apply some simple FR quality measures to test the framework in practise.

The performance of the FR quality measures was studied against the subjective evaluation results (ground truth) introduced in [15]. In brief, all samples (same image content) were placed on a table in random order. Also the numbers from 1 to 5 were presented on the table. An observer was asked to select the sample representing the worst quality of the sample set and place it on the number 1. Then, the observer was asked to select the best sample and place it on the number 5. After that, the observer was asked to place the remaining samples on numbers 1 to 5 so that the quality grows regularly from 1 to 5. The final ground truth was formed by computing mean opinion scores (MOS) over all observers. Number of the observers was 28.

In Fig. 5, the results for the two mentioned FR quality measures, PSNR and LabMSE are shown, and it is evident that even with these most simple pixel-wise measures, a strong correlation to such an abstract task as the "visual quality experience" was achieved. It should be noted that our subjective evaluations are on a much more general level than in any other study presented using digital images. Linear correlation coefficients were 0.69 between PSNR and MOS, and -0.79 between LabMSE and MOS. These are very promising and motivating future studies on more complicated measures.

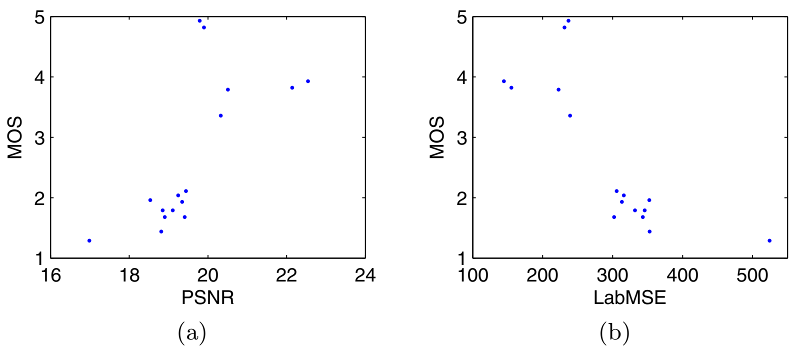


Fig. 5. Scatter plots between simple FR measures computed in our framework and subjective MOS: (a) PSNR; (b) LabMSE

4 Discussion and Future Work

The most important consideration in the future work is to find FR measures which are more appropriate for printed media. Although our registration method works very well, sub-pixel errors still appear and they always affect simple pixel-wise distance formula, such as the MSE. In other words, we need FR measures which are less sensitive to small registration errors. Another notable problem arises from the nature of subjective tests with printed media: The experiments are carried out using printed (hard-copy) samples and the actual digital reference (original) is not available for the observers and not even interesting; the visual quality experience is not a task of finding differences between the reproduction and original, but a more complex process of what is seen as excellent, good, moderate or poor quality. This point has been wrongly omitted in many digital image quality studies, but it must be embedded in FR measures.

In the literature, several approaches have been proposed to enhance the FR algorithms to be more consistent with the human perception: mathematical distance formulations (e.g., fuzzy similarity measures [18]), human visual system (HVS) model based (e.g., Sarnoff JNDmetrix [19]), HVS models combined application specific modelling (DCTune [20]), structural (structural similarity metric [2]), and information theoretic (visual information fidelity [3]). It will be interesting to evaluate these more advanced methods in our framework. Proper statistical evaluation, however, requires a larger amount of samples and several different image contents. Another important aspect is the effect of the cut-off frequency in the descreening stage. What is the suitable cut-off frequency and does it depend on the used FR measure?

5 Conclusions

In this work, we presented a framework to compute full reference (FR) image quality measures, common in digital image quality research field, for printed natural images. The work was first of its kind in this extent and generality, and it will provide a new basis for future studies on evaluating visual quality of printed products using methods common in the field of computer vision and digital image processing.

Acknowledgement

The authors would like to thank Raisa Halonen from the Department of Media Technology in Helsinki University of Technology for providing the test material and Tuomas Leisti from the Department of Psychology in University of Helsinki for providing the subjective evaluation data. The authors would like to thank also the Finnish Funding Agency for Technology and Innovation (TEKES) and partners of the DigiQ project (No. 40176/06) for support.

References

1. Daly, S.: Visible differences predictor: an algorithm for the assessment of image fidelity. In: Proc. SPIE, San Jose, USA. Human Vision, Visual Processing, and Digital Display III, vol. 1666, pp. 2–15 (1992)
2. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
3. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Transactions On Image Processing 15(2), 430–444 (2006)
4. Sadovnikov, A., Salmela, P., Lensu, L., Kamarainen, J., Kalviainen, H.: Mottling assessment of solid printed areas and its correlation to perceived uniformity. In: 14th Scandinavian Conference of Image Processing, Joensuu, Finland, pp. 411–418 (2005)
5. Vartiainen, J., Sadovnikov, A., Kamarainen, J.K., Lensu, L., Kalviainen, H.: Detection of irregularities in regular patterns. Machine Vision and Applications 19(4), 249–259 (2008)
6. Sheikh, H.R., Bovik, A.C., Cormack, L.: No-reference quality assessment using natural scene statistics: JPEG 2000. IEEE Transactions on Image Processing 14(11), 1918–1927 (2005)
7. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Transactions On Image Processing 15(11), 3440–3451 (2006)

8. Wyszecki, G., Stiles, W.S.: Color science: concepts and methods, quantitative data and formulae, 2nd edn. Wiley, Chichester (2000)
9. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
10. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision 74(1), 59–73 (2007)
11. Fischler, M., Bolles, R.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing 24(6) (1981)
12. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE-TPAMI 13(4), 376–380 (1991)
13. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2003)
14. Avcibaş, I., Sankur, B., Sayood, K.: Statistical evaluation of image quality measures. Journal of Electronic Imaging 11(2), 206–223 (2002)
15. Oittinen, P., Halonen, R., Kokkonen, A., Leisti, T., Nyman, G., Eerola, T., Lensu, L., Kälviäinen, H., Ritala, R., Pulla, J., Mettänen, M.: Framework for modelling visual printed image quality from paper perspective. In: SPIE/IS&T Electronic Imaging 2008, Image Quality and System Performance V, San Jose, USA (2008)
16. Eerola, T., Kamarainen, J.K., Leisti, T., Halonen, R., Lensu, L., Kälviäinen, H., Nyman, G., Oittinen, P.: Is there hope for predicting human visual quality experience? In: Proc. of the IEEE International Conference on Systems, Man, and Cybernetics, Singapore (2008)
17. Eerola, T., Kamarainen, J.K., Leisti, T., Halonen, R., Lensu, L., Kälviäinen, H., Oittinen, P., Nyman, G.: Finding best measurable quantities for predicting human visual quality experience. In: Proc. of the IEEE International Conference on Systems, Man, and Cybernetics, Singapore (2008)
18. van der Weken, D., Nachtegael, M., Kerre, E.E.: Using similarity measures and homogeneity for the comparison of images. Image and Vision Computing 22(9), 695–702 (2004)
19. Lubin, J., Fibush, D.: Contribution to the IEEE standards subcommittee: Sarnoff JND vision model (August 1997)
20. Watson, A.B.: DCTune: A technique for visual optimization of DCT quantization matrices for individual images. Society for Information Display Digest of Technical Papers XXIV, 946–949 (1993)