Mateusz BUCZKOWSKI
POZNAN UNIVERSITY OF TECHNOLOGY
3 Polanka St., 61-655 Poznan, Poland

Measuring the effective coverage of the image databases

Abstract

The paper presents new methods for estimating effective coverage of the image databases in spatial information (SI) vs. colorfulness (CF) space. Existing and commonly used metrics are vulnerable to databases containing very different images in terms of SI and CF, which can maximize the relative ranges, while preserving, in specific cases, high uniformity. The area covered by convex hull created on SI × CF space can be large, but the interior of the convex envelope might be poorly filled. To overcome this flaw, a fill factor is proposed, which determines how well the convex hull area is utilized. In the paper the method is described in details and presented on a few existing databases, as well as on artificial, non-existing databases to prove its robustness in different scenarios. Provided results show that new metric is closer to the actual coverage than existing methods.

Keywords: image quality assessment, coverage analysis, image characteristic analysis.

1. Introduction

As computer vision is used in more and more applications, where multiple processing steps are involved, the final image can be distorted in many ways. It is important then, to assure good image quality for a human observer. To do so, one need more than basic metrics like peak signal to noise ratio (PSNR), which is based only on a difference of reference and distorted image. It is so, because the same value of PSNR can be perceived differently for Gaussian noise, compression artifacts or color change. A lot of effort is put into finding an efficient method, which can predict the perceived image degradation compared to the ideal image.

In order to enable such research, different institutes and universities create image quality assessment (IQA) databases, which contain a set of pristine images, their distorted versions and mean opinion scores (MOSs) assigned by human observers to the particular images. Having such database as a ground truth, one can design and validate an algorithm which based on an image will predict its MOS.

There are multiple available IQA databases in public domain, which can serve as a ground truth for algorithm evaluation. Three of them, namely CIDIQ [1] LIVE2006 [2, 3, 4] and TID [5] were selected to show how proposed methods behave for real datasets. As the process of gathering MOS values for images is time- and cost-consuming, the number of images is limited. Aforementioned databases contain 23, 29 and 25 reference images, respectively.

The variety of possible image content is huge. From dark and shady up to colorful and containing lots of detailed textures. It is important then, to measure coverage of this variety by different databases. The more different scenes and types of images the better, as the database offers more complete overview and can be used to validate behavior of the algorithm in different scenarios.

In the paper currently used metrics for measuring coverage are presented, their weak spot is pointed out and possible solutions to overcome this issue are presented.

2. IQA database coverage

For a single image two metrics are calculated, namely spatial information (SI) and colorfulness (CF). Spatial information is a measure of edge energy and it is defined as follows [6]:

$$SI = \sqrt{\frac{L}{1800}}\sqrt{\frac{\sum s_r^2}{P}}. \tag{1}$$

where s_r is a square root of sum of edges energy in horizontal and vertical direction. Edge energy is obtained from Sobel kernel filtration, L is a vertical resolution and P is a number of pixels in the image.

Colorfulness represents intensity and variety of colors in an image [7]. In order to calculate the CF we need to represent an image in opponent color spaces: $rg = R - G$ and $yb = 0.5(R + G) - B$. Having those we can calculate mean (μ) and standard deviation (σ) in both directions and calculate CF using following formula:

$$CF = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} + 0.3\sqrt{\mu_{rg}^2 + \mu_{yb}^2}. \tag{2}$$

Based on SI and CF for each image the coverage of the database can be calculated. The total relative coverage [8] is calculated as the area of convex envelope of all images in a database. For the area calculation, both SI and CF are normalized by the largest value among all databases, i.e.:

$$CF_{norm} = \frac{CF_i}{CF_{max}} \quad and \quad SI_{norm} = \frac{SI_i}{SI_{max}}. \tag{3}$$

Results of SI and CF calculation for LIVE2006, TID and CIQID databases, without normalization are presented in the Figure 1. The figure contains also calculation for three non-existing, artificially created databases, which are intended to show the weakness of total relative coverage and will be used to present behavior of new proposed metrics. New datasets have following properties: *Random* is a set of random points in range of 0-160 for *CF* and 0-175 for *SI*, which correspond to ranges of real databases. The distribution is uniform in both axes. Next database, *Uniform*, is a uniform grid of points. The 110 points are evenly spaced in both directions. The *X shaped* database contains points only on diagonals, which create large convex hull, with uniform distribution, while considering individual axes, but leaving a lot of uncovered area inside the envelope.

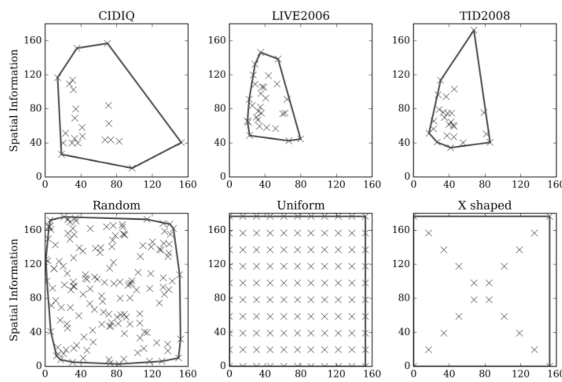

Fig. 1. Colorfulness vs. spatial information for considered databases and corresponding convex hull

As one can see databases cover not only different areas, but they are filled with different number of images and the empty spaces within convex envelope may vary. This is why considering only occupied area is not enough.

Author of [8] proposes to measure uniformity in each of the axis (*CF* and *SI*) individually. Based on last two examples, namely Uniform and X shaped one might notice, that uniformity will be the same, but the completeness of these two databases is quite different.

3. Convex fill factor

In order to overcome identified weakness a method for describing how well the convex hull is filled must be used. For this purpose three approaches are presented, compared and discussed.

3.1. Fixed radius coverage

First proposed approach is based on assumption that single image cover not only precisely computed values of *SI* × *CF*, but it can serve as a representative for neighboring *SI* × *CF* values. Each image is then not a single point on *SI* × *CF* plane, but it creates circle of some radius *r* [9]. The circles from different images do not stack up. It means that adding new image with *SI* and *CF* values similar to existing one does not improve significantly the coverage. Ratio of circles area that fits into convex hull and a whole area of convex envelope is a measure of how well the images fill the *SI* × *CF* space. This method is visualized in the Figure 2 for all considered databases.

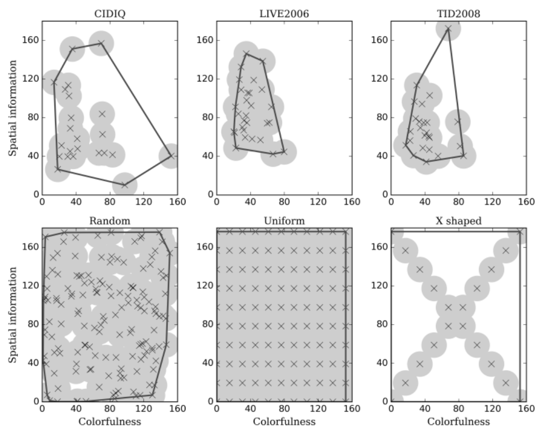

Fig. 2. Fixed radius coverage visualization

3.2. Gaussian radius coverage

The Gaussian radius coverage is closely related to the above method. Instead of constant value in fixed radius Gaussian function is used. This relaxes the assumption, that image covers the larger area of *SI* × *CF* in the same way. With this method, the further we are from the actual *SI* × *CF* value the lower the coverage is. As in previous case, completeness is expressed as a ratio of marked area and whole convex hull area. Figure 3 presents visualization of this method for the same databases.

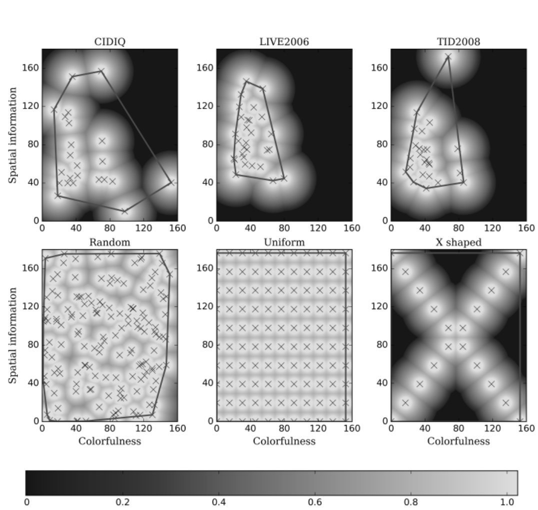

Fig. 3. Gaussian radius coverage visualization

3.3. Delaunay triangulation uniformity

Last of the discussed methods is based on Delaunay triangulation, which from set of points *P* creates triangulation *DT(P)* such that no point *P* is inside the circumcircle of any triangle in *DT(P)* [10]. Then area of each triangle is calculated and normalized in following way:

$$A_i = \frac{A_i}{\sum_k A_k}, \tag{4}$$

where A_i is area of *i*-th triangle.

Having normalized areas one can calculate B-bin histogram of areas and then uniformity of triangulation, which is based on entropy. As the entropy is maximum for the 0.5 and we are interested in minimizing the entropy following formula is used:

$$U = \sqrt{2 \cdot |0.5 + \sum_{k=1}^{B} p_{Ak} \log_B P_{Ak}|}, \tag{5}$$

where p_{Ak} is normalized number of areas in bin *k*.

Uniformity is expressed as a value in range 0…1. The higher the value the more uniform convex hull is. Figure 4 shows Delaunay triangulation for all considered databases, where area is depicted with different colors.

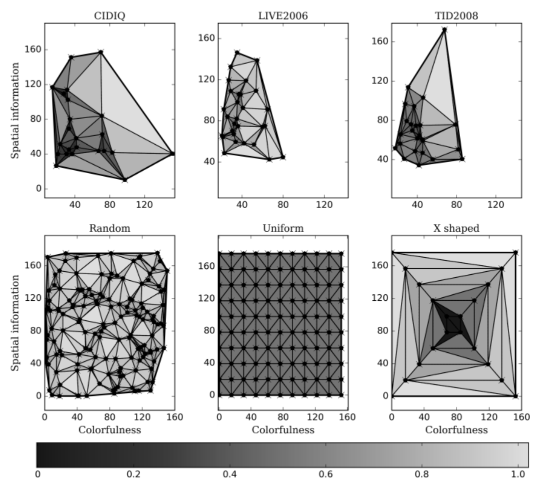

Fig. 4. Delaunay triangulation uniformity visualization

4. Results

The output values of three proposed methods for all considered databases are summarized in the Table 1 and depicted in the Figure 5. While measuring the performance of the IQA method, an author wants to make sure that large variety of images were used during this process. Therefore good IQA database has many images with different values of *SI* and *CF*. The perceptual quality order of the presented databases, regardless of the convex area, but taking into account its completeness is as follows: Uniform, Random, LIVE, TID, CIDIQ and X shaped. Such order is motivated by amount of free, uncovered areas in convex envelope.

Tab. 1. Numerical results for proposed methods

	Fixed radius	Gaussian radius	Delaunay triangulation
CIDIQ	0.42	0.60	0.29
LIVE	0.94	0.90	0.88
TID	0.59	0.72	0.62
Random	0.95	0.92	0.79
Uniform	1.00	0.92	1.00
X shaped	0.34	0.53	0.61

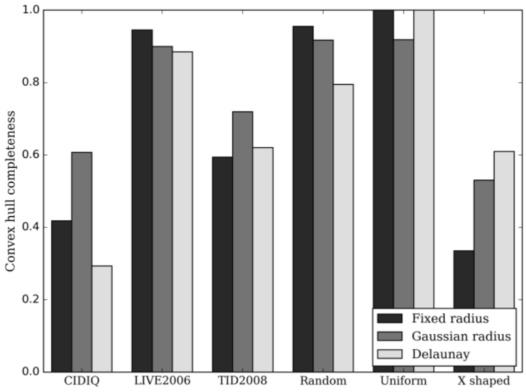

Fig. 5. Results comparison for different methods of computing convex hull completeness

From provided results one can see that both fixed radius and Gaussian radius methods provide such gradation of databases. In case of fixed radius, the differences are more noticeable, as we can easily point out the perfect database (Uniform having value 1.0) and the bad database (X shaped, 0.34). In case of Gaussian radius, none of the considered databases reach the perfect score, as the images would have to be infinitely dense in order to provide 1.0 score in this method. Similarly, none of the databases goes below score 0.5. While comparing these two methods we can see that decision from Gaussian radius method is softer – this method does not promote or reject databases easily.

Delaunay triangulation method provides different ordering of the databases. In this case it is Uniform, LIVE, Random, TID, X shaped and CIDIQ. This method tends to favor high regularity in the database. The more uniform distribution of points in both directions, the higher score from Delaunay triangulation is. The weakness of this method is particularly well visible in case of Random database. Its coverage is very good – area is completely filled with images, yet the score is only 0.79. This is due to the fact that some areas (e.g. around CF = 125, SI = 55) have higher density of images than other regions. Because of this, the database is punished with lower score. Such phenomenon does not occur in case of first two methods. If additional images appear in the database, causing uniformity, the score can only increase while calculating completeness by fixed or Gaussian radius. For Delaunay triangulation adding new images can decrease the score, which can be considered as counterintuitive.

This behavior can be more intuitive if we would consider this metric not as a how well images cover the particular convex hull, but rather how big area this amount of images cover. One would expect that adding new images to the database increase its variety. If it is not the case and the convex hull is not changed after new image is added we might expect the score to be lower.

5. Conclusions

In the paper the measures for calculating IQA databases coverage were presented and a weakness of currently used approaches was pointed out. As in some cases theoretical coverage can be maximized with low variety of images, a new metric is proposed, which shows how well the covered area is utilized. Three possible solutions are presented and discussed, namely fixed radius coverage, Gaussian radius coverage and Delaunay triangulation uniformity.

New metrics are calculated for three selected and publicly available IQA databases as well as for three artificial, non-existing databases. Thanks to this, strength of new methods could be shown.

Two of the proposed metrics can be successfully used in the intended purpose of measuring effective coverage, while the third one does not fully meets the criterion of grading databases in order of perceptual coverage. This metric, however, can be used in different aspect. It shows how uniform and regular the database is.

Work founded by the Ministry of Science and Higher Education for the statutory activity of conduction research and development work and related tasks, contributing to the development of young scientists and doctoral students in 2017 under the project number 08/83/DSMK/4726.

6. References

[1] Liu X., Pedersen M., Hardeberg J.Y.: CID:IQ – A New Image Quality Database, vol. 8509, Springer, 2014.
[2] Sheikh H.R., Wang Z., Cormack L., Bovik A.C.: LIVE Image Quality Assessment Database Release 2, http://live.ece.utexas.edu/research/quality, Mar. 2017 .
[3] 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, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
[4] 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, vol.13, no.4, pp. 600-612, Apr. 2004.
[5] Ponomarenko N., Lukin V., Zelensky A., Egiazaarian K., Carli M., Battisiti F.: TID2008 – A Database for Evaluation Full-Reference Visual Quality Assessment Metrics, Advances of Modern Radioelectronics, vol. 10, pp. 30-45, 2009.
[6] ANSI T1.801.03: Digital transport of one-way video signals – parameters for objective performance assessment. American National Standards Institute, 1996.
[7] Hasler D., Susstrunk S.: Measuring colorfulness in natural images, Proc. SPIE Human Vision and Electronic Imaging vol. 5007, pp. 87-95, 2003.
[8] Winkler S.: Analysis of Public Image and Video Databases for Quality Assessment. IEEE Selected Topics on Signal Processing, vol. 6, no. 6, pp. 616-625, 2012.
[9] Buczkowski M., Stasinski R.: Effective Coverage as a New Metric for Image Quality Assessment Databases Comparison, Submitted to: International Conference on Systems, Signals and Image Processing , 2017, Poznan, Poland.
[10] Delaunay B.: Sur la spheere vide. A la meemoire de Georges Voronoi, Bulletin de l'Academie des Sciences de l'URSS. Classe des sciences mathematiques et na, no. 6, pp. 793–800, 1934.

Received: 22.02.2017 Paper reviewed Accepted: 04.04.2017

Mateusz BUCZKOWSKI, MSc, eng.

Mateusz Buczkowski graduated from Poznan University of Technology in 2012, where he is currently pursuing his PhD. His research interests include image and video processing in general. Current work is strictly related to no-reference image quality assessment. He took part in two EU-FP7 projects in field of telecommunications, which was a great opportunity to gain research experience.

e-mail: mateusz.buczkowski@put.poznan.pl