VQEG eLetter

Volume 2, Issue 1 • February 2016

Contents[1]

Storyboard

VIME co-chairs and issue Editors: Quan Huynh-Thu, Michele Saad and James Goel

Introduction

No-reference perceptual image quality needs to address the effects of image capture and rendering in consumer devices

The Video Quality Experts Group (VQEG) provides an open forum where video quality experts meet to advance the field of video quality assessment. In the early years of VQEG, the group focused on video quality, addressing issues related to video compression and transmission over lossy networks. However, with the latest advances in digital cameras and multimedia devices, and their widespread use to capture photos and videos, visual quality has opened another level of investigation. In particular, image quality of photos captured by such devices is not so much impacted by compression and transmission errors but by capture and rendering. With the current state of multimedia devices, visual quality has expanded to include aspects such as aesthetic quality. Two photos without any visible compression degradation can still be perceived as exhibiting very different image quality due to the possible artifacts introduced by the capture and due to different post-processing algorithms. In this scenario, the notion of reference image has no real meaning anymore. What makes an image more appealing than another one? For a given content, what process is required to enhance the perception of the quality? To answer these questions, no-reference quality assessment methods are required. However, so far, most of the research work on objective no-reference perceptual image quality has been focused on the impact of image compression and transmission. Likewise, subjective assessment methods consider either the notion of image quality fidelity or that processing degrades the quality of the image. The existing methods cannot be used to quantify the impact of a processing applied to enhance the quality of an image or to assess aesthetic quality.

To address this gap, VQEG started a new workgroup and project called VIME (Video and Image Models for consumer content Evaluation) in mid-2014. VIME addresses the no-reference measurement of the quality of images (and videos in the future) captured by consumer cameras, such as DSLRS, compact cameras, mirrorless cameras, and camera-enabled devices, such as mobile phones and tablets. In this context, both subjective and objective quality assessment methods are needed.

Issue Overview

This eLetter issue provides a collection of articles covering image datasets and scientific advances related to the work of VIME. Articles cover image datasets available for research, metrics and descriptors useful to characterize the quality of images.

"The VIME Image Database" by Dr. Quan Huynh-Thu of Canon Information Systems Research Australia (CiSRA) and Dr. Michele Saad of Intel provides an introduction and description of the free dataset of images being developed by the VQEG VIME project. This is an on-going effort and the article describes how anyone can easily contribute new images to the dataset.

"Image Classifiers" by Mikołaj Leszczuk, Remigiusz Baran, Michał Grega, Krzysztof Rusek, Piotr Guzik of AGH University describes several no-reference metrics developed or implemented by their research group.

"VESA Advanced Display Stream Compression" by James Goel of Qualcomm calls for proposals to support the work of VESA Advanced Display Stream Compression (Adv-DSC), which aims at developing algorithms for low-impairment compression of HDR and wide-color gamut images.

"VIME and Subjective Image Quality Tests" by Michele Saad and Philip Corriveau of Intel presents the Consumer Content Resolution and Image Quality (CCRIQ) database.

"Combining HSV Color and rootSIFT for Image Retrieval" by Ahmad Alzu'bi, Abbes Amira, Naeem Ramzan, and Tareq Jaber of the University of West Scotland introduces an optimized image descriptor that combines color and local features for image retrieval.

"Open Source No-Reference Toolset moving image quality forward in new ways!" by Philip Corriveau of Intel offers an open source no-reference image tool set called VIQET (VQEG Image Quality Evaluation tool).

The VIME Image Database

Quan Huynh-Thu and Michele Saad

VIME Charter

"VIME's charter is to address the problem of no-reference quality assessment of consumer images. Attention is drawn to the terms *consumer images* and *no-reference*."

VQEG formed the VIME (Video and Image Models for consumer content Evaluation) workgroup at its meeting in Stockholm in July of 2014. VIME's charter is to address the problem of no-reference quality assessment of consumer images. Attention is drawn to the terms "*consumer images*" and "*no-reference*".

Consumer images are images captured by a consumer device such as a mobile phone, tablet, compact camera, digital single-lens reflex (DSLR) camera, or other type of camera available to a consumer. Thus, VIME addresses distortions introduced by the camera capture process and post-processing (not just compression artifacts), using real-world images taken with digital cameras and smartphones.

"No-reference quality assessment" refers to the problem of evaluating the quality of an image in the absence of a pristine or perfect version of that image. The no-reference paradigm bypasses the need to compare against a reference and performs the image quality evaluation on the test image with no additional ancillary information. In most practical usage such reference is not available. In the case of photography, such notion of reference does not exist, especially when considering aesthetics. The evaluation is based only on the intrinsic features of the test image.

The goals and differentiations from past no-reference image quality work are summarized by the following points:

"VIME is currently collecting images to build a large dataset of photos that can be used to conduct its research work.
… VIME is making this dataset freely available to the research community."

- VIME considers distortions introduced by camera capture and processing (not only compression artifacts) using real-world and natural images taken with digital cameras and smartphones.
- VIME addresses no-reference and aesthetic quality of images and videos.
- VIME addresses image enhancement, i.e. there is no concept of pristine reference image as the processed image can be of higher image quality than the starting image.

Further information can be found on the VIME project website at: http://www.its.bldrdoc.gov/vqeg/projects/vime.aspx.

Subjective studies are central to understanding the human visual response to image quality, and image data sets with corresponding subjective scores are key to image quality model development and test.

VIME is currently building a dataset of photos that can be used to conduct its research work. Although many images exist on the internet, most image datasets available on the web do not meet the needs of VIME. Dedicated to the goal of no-reference image quality model development, the VIME workgroup began the project of creating an open data set of images for use in the VIME research.

A Flickr group called the *VIME Image Database* was created at https://www.flickr.com/groups/vime/. The aim of the Flickr group is to create a dynamically growing data set of images that will eventually be used for subjective testing and model design. People are invited to contribute images to the database under the CC0 license (Public Work Dedication license).

The dataset contains more than 1,100 images to date and is constantly growing.

VIME is making this image dataset freely available to the research community.

Subjective Image Quality Evaluation

Objective model development typically relies on subjective quality evaluation studies for two reasons: 1) to understand the psychophysical response consumers have to image quality; and 2) to use the subjective data produced for model development and model performance evaluation.

The studies in [1] and [2] have begun to address the need for consumer-like images with associated subjective scores for the purpose of no-reference quality evaluation model development. In most previous traditional studies the images used for subjective testing consisted of reference images with versions of them that contain varying degrees of simulated distortions such as blur, additive noise, or compression and transmission artifacts. In contrast, the studies in [1] and [2] emphasize the need for datasets of images with non-simulated distortions; that is, images with real artifacts that are representative of the type of photos seen by consumers. These studies emphasize the need for subjective quality evaluation of images as they are produced by consumer devices.

While these studies are an important step towards building datasets of representative consumer content, the number of scenes in these datasets is limited. On the other hand, the space of consumer images is large, complex, and requires understanding the camera response in a multitude of real life scenarios. To better understand the camera response, the larger the corpus of consumer images available to researchers, the better. The need for a large set of consumer images has motivated the creation of the dynamically growing VIME Image Database which we describe in the following section.

VIME Image Database: Goal

The VIME Image Database was created shortly after the VQEG face-to-face meeting in Santa Clara, California in February of 2015. The goal of this data set is multifold:

1) To create a dataset of images that are freely available for anyone to use with no licensing issues. The images of the VIME Image database are hence all under the CC0 (Public dedication work) license.
2) To create a dataset of images that are representative of the types of images consumers capture. These are images that are captured with mobile devices, compact cameras, and higher end cameras such as DSLR on a regular basis.
3) To gather a large set of images to be used in subjective studies to examine and address questions related to consumer image quality. There are a number of unanswered questions that the VIME group has begun to address. This dataset is expected to provide images for use in such studies.
4) To use the large corpus of images for no-reference objective image evaluation model development and performance evaluation.

The VIME Image Database is a dynamically growing set of photos that will increase in size as people continue to contribute photos to it.

How to Contribute and Upload Photos to the VIME Image Database

VIME has setup a Flickr Group called "**VIME Image Database**". Everyone is invited to contribute by uploading images to this image collection. People already using Flickr can contribute their images right away. For those not so familiar with Flickr, here are a few simple steps to follow.

You will need to:

- Have or create a Flickr account

- Join the VIME Image Database group to upload and share your images with the VIME group

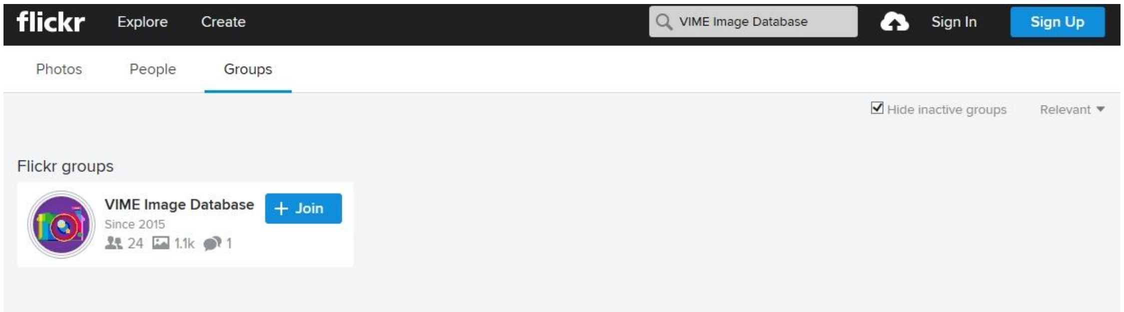


Figure 1: The VIME Image Database can be found by searching for it in the "Groups" tab.

Here are some typical examples of how you can contribute to the VIME image dataset:

- Upload a single image: the image can be of any quality and photographic skills
- Upload several images of the same scene captured by different cameras
- Upload several images of the same scene captured by the same camera using different capture settings (e.g. shutter speed, aperture, ISO) or using different lenses

You are required to **upload your images under the CC0 license (Public Work Dedication)**. This ensures that VQEG and the entire research community can use and exchange the images without any legal issues. This requirement ensures that this image dataset is freely available to the entire research community.

There are currently no specific requirements on scene content, resolutions, or aspect ratios. However, it is recommended that you upload the highest resolution available for your image.

There are no requirements concerning the image quality and level of photographic skills. VIME is interested in images that have good photographic quality as well as images that have bad photographic quality.

To contribute your images, follow these simple steps:

- Sign in to your Flickr account or create a Flickr account if you don't have one.
- Search for "VIME Image Database" in Groups, as shown in Fig. 1, and join.
- Share your photos with the VIME Image Database group using one of the following two options:
  - Share new photos with VIME when you upload them for the first time to your Flickr account, or
  - Share existing photos in our Flickr account with VIME.

Option 1: Upload New Photos and Share them with the VIME Image Database

This will allow you to immediately share the photos that you are uploading to your account with VIME.

- Go to your account homepage click on "Upload" to add a new image to your own collection. See Fig. 2.

Figure 2: Click on "Upload" at the top right hand side corner of your account homepage.

- Add a description to your photos (Optional)
- Add tags, e.g. daytime, nighttime, indoor, outdoor, landscape, portrait, building, architecture, plant, forest, grass, rock, car,… This is shown in Fig. 3.

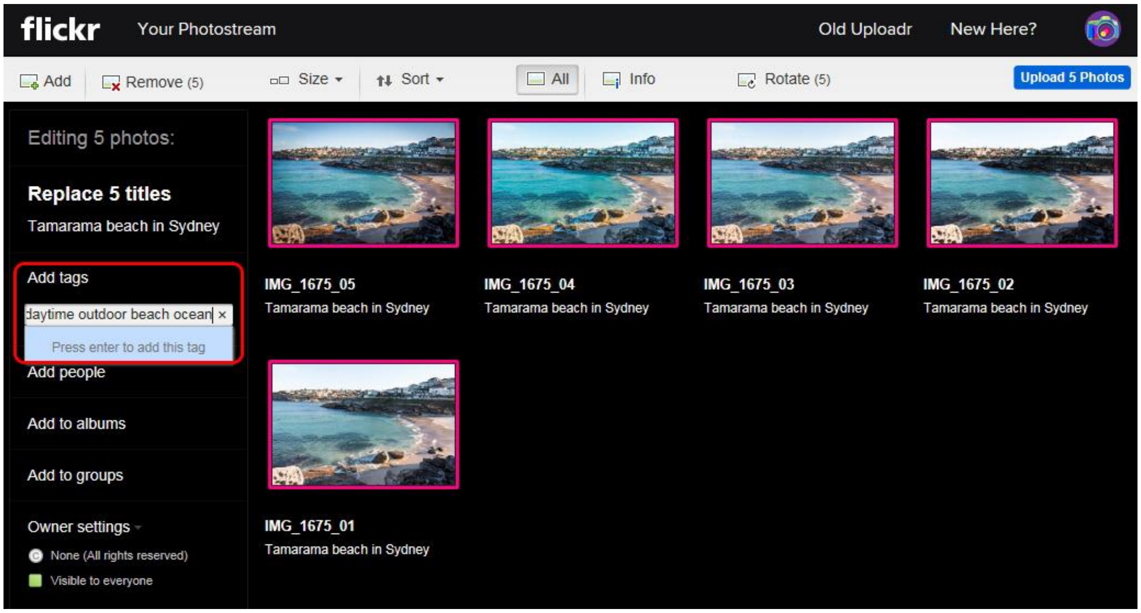


Figure 3: Adding tags to the photos.

- **Add to the VIME group**: Click on "Add to groups" and select "VIME Image Database" from the list of the groups you have joined as shown in Fig. 4.

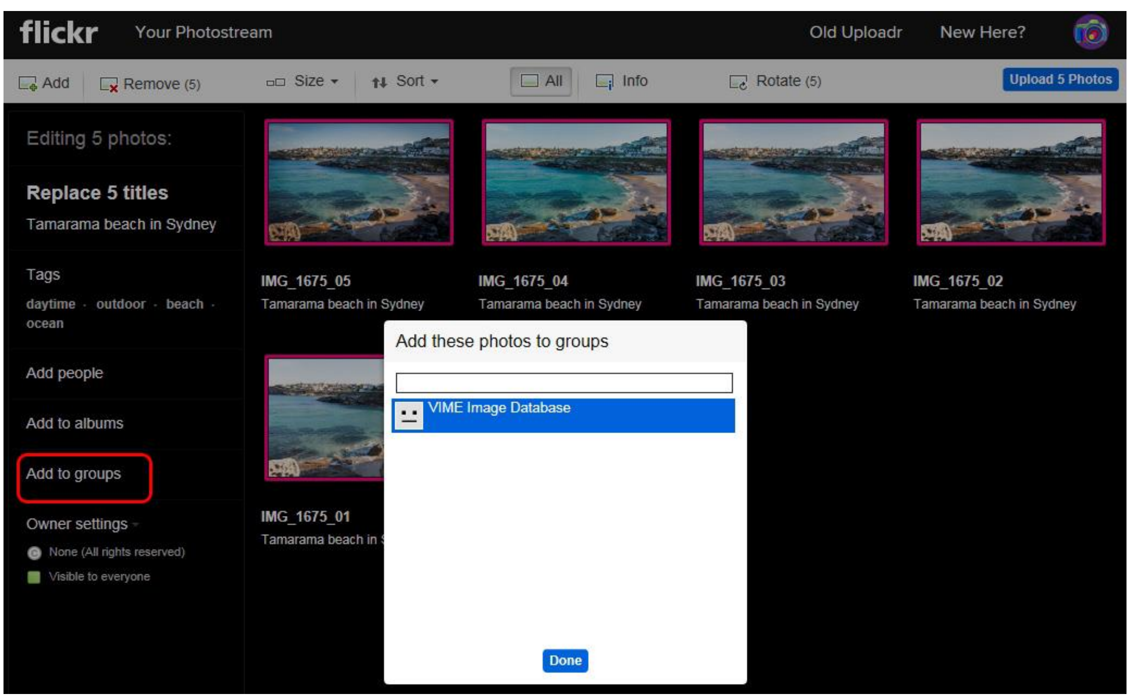


Figure 4: Add the photos to the VIME Image Database group.

- Set **Owner settings to CC0**: by default this setting is set to "None" so you need to change this setting to "Public Domain Dedication (CC0)" – Click 'done' after selection. This is demonstrated in Fig. 5. **This is important as your image will not be used in the VIME Image Database if its license is not set to CC0.**

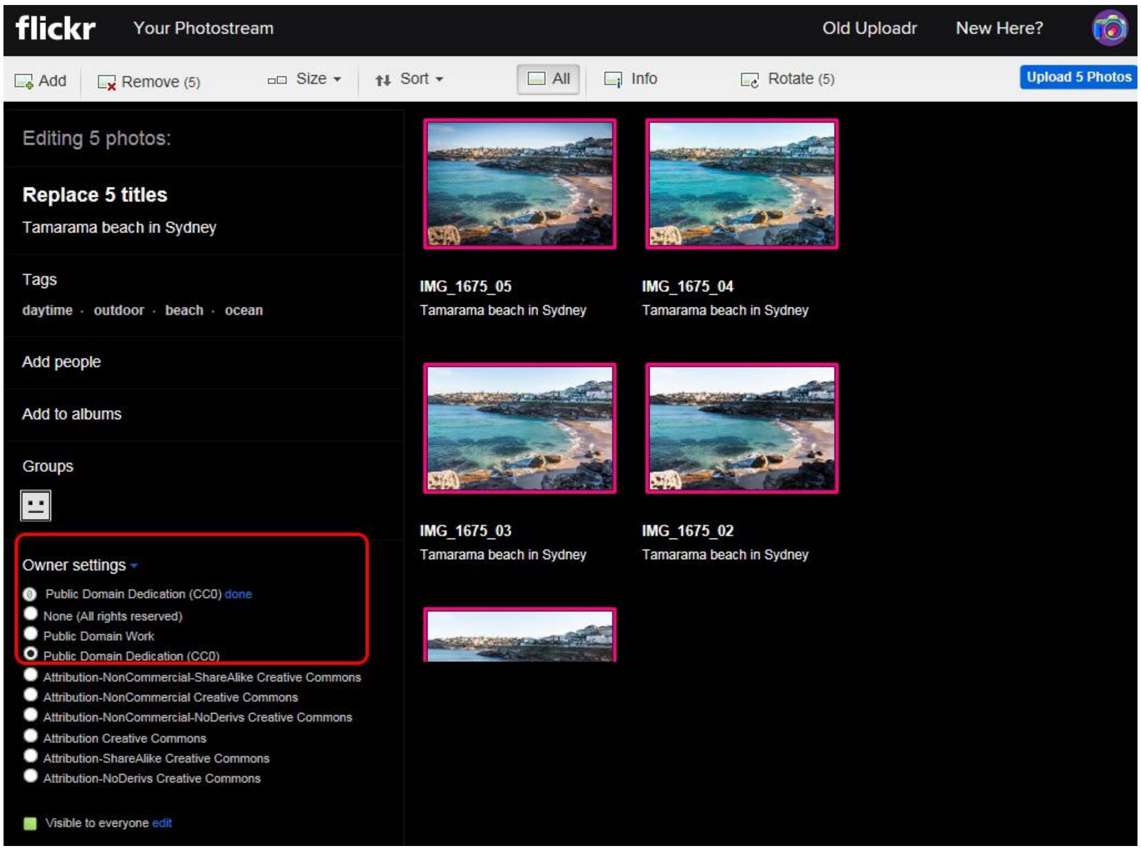


Figure 5: Set the photo licenses to CC0.

- Upload Photos: click on "Upload" in top-right corner. (Fig. 6.)

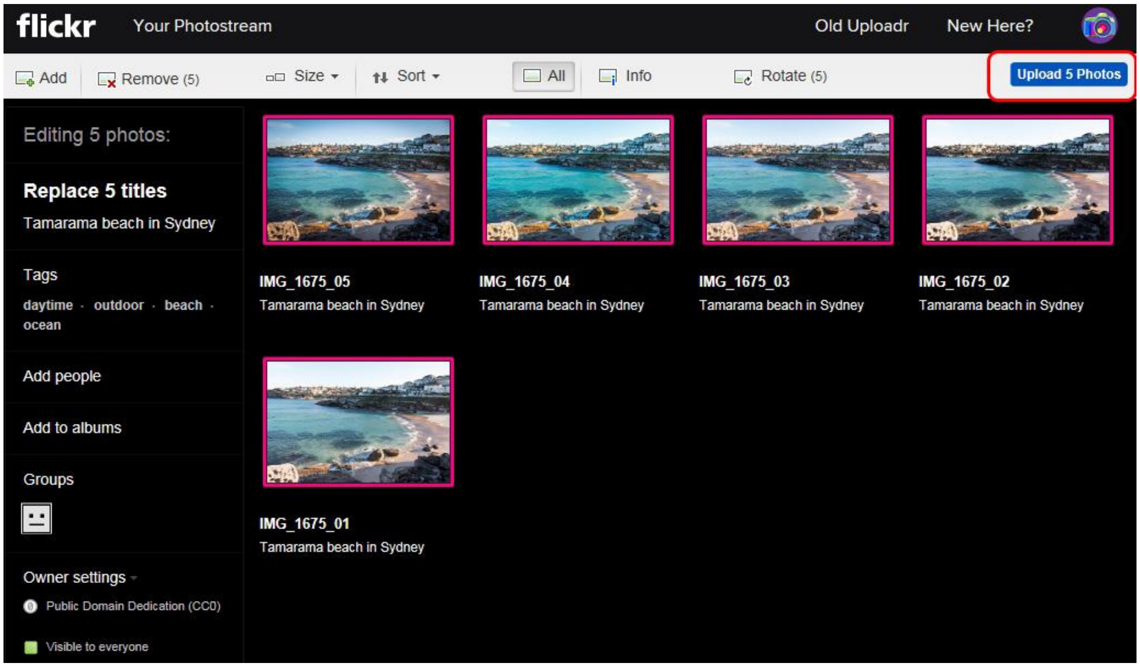


Figure 6: Upload the photos.

Option 2: Share Existing Photos in your Account with the VIME Image Database

This will allow you to share a batch of photos that already exist in your Flickr account with the VIME Image Database.

- Go to the Menu You -> Organize (as in Fig. 7)
- Select all the photos you want to share with the VIME Image Database group
- Make sure you edit the settings similarly to Option 1:
  - Add to VIME Group
  - Set Owner settings to CC0
  - Add tags
- Click "Send to Group"

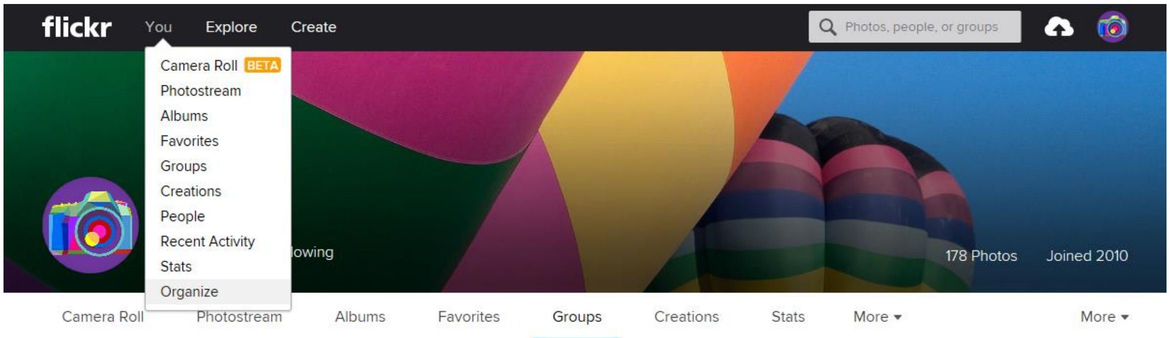


Figure 7: Go to "Organize" under the "You" menu.

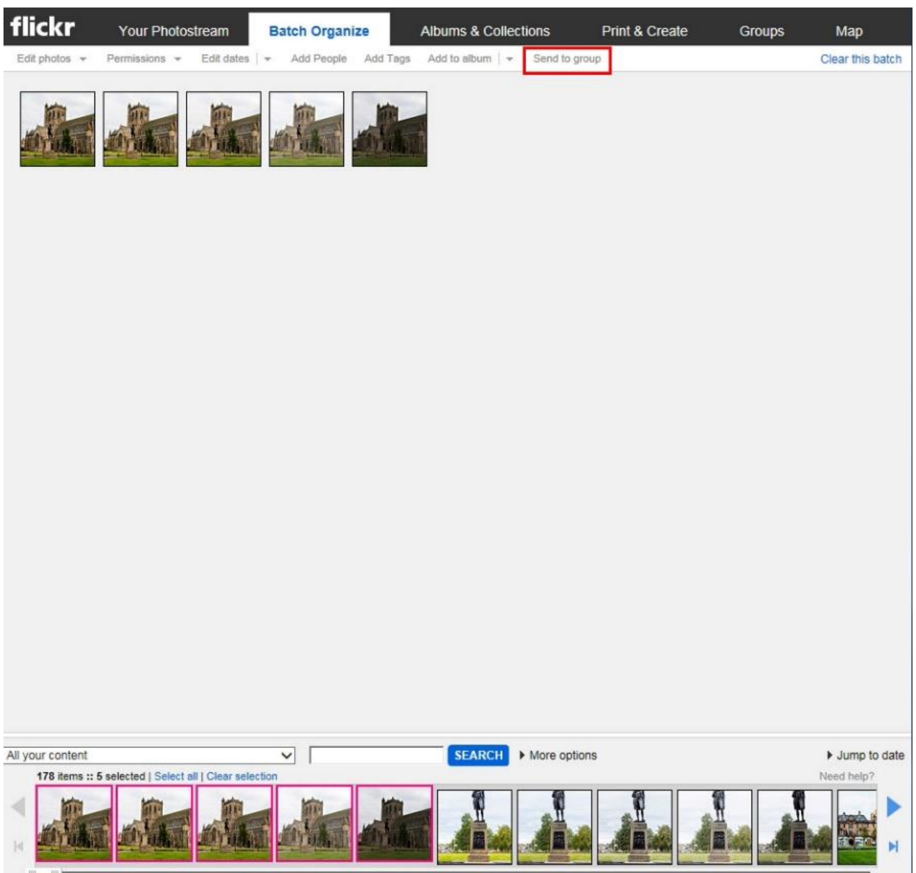


Figure 8: Click "Send to Group" under in the "Batch Organize" tab.

***Michele A. Saad** is a Senior Engineer and Researcher in perceptual image and video quality assessment at Intel. She received her Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2013. Her research interests include statistical modeling of images and videos, motion perception, design of perceptual image and video quality assessment algorithms, and statistical data analysis and mining and machine learning.*

***Quan Huynh-Thu** is a Senior Research Engineer at Canon Information Systems Research Australia (CiSRA). He received his Ph.D. degree in electronic systems engineering from the University of Essex (UK) in 2009. His research interests include psychophysics, the understanding of the human visual system and its application to advanced digital imaging using models of perception.*

References

[1] Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P. and Häkkinen, J. "CID2013: a database for evaluating no-reference image quality assessment algorithms", *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 390-402, Jan. 2015.

[2] M.A. Saad, M.H. Pinson, D.G. Nicholas, N. Van Kets, G. Van Wallendael, R. Da Silva, R.V. Jaladi, and P.J. Corriveau, "Impact of Camera Pixel Count and Monitor Resolution on Perceptual Image Quality", *IEEE Color and Visual Computing Symposium (CVCS)*, August 2015.

Image Classifiers

Mikołaj Leszczuk, Remigiusz Baran, Michał Grega, Krzysztof Rusek, Piotr Guzik

Introduction

"Data analysis is performed in order to enrich the data (mainly images) by extracting their features and classify them according to given criterions"

The IMCOP project—an "Intelligent Multimedia System for Web and IPTV Archiving. Digital Analysis and Documentation of Multimedia Content"—is a joint Polish-Israeli R&D project realized by a consortium of four partners. In general, IMCOP's objectives are twofold: multimedia data analysis and content discovery on one side, and data aggregation, content related binding (finding and assigning content related connections between data) and delivery on the other. Data analysis is performed in order to enrich the data (mainly images) by extracting their features and classifying them according to given criteria, as in Baran et al. (2015). A list of those criteria, as well as the classifiers developed so far in the IMCOP project, is as follows:

- Age classifier
- Bokeh effect detector
- Dominant color of clothes (dresses)
- Dominant color counter
- Nudity identification
- People counter
- Profile/enface classification
- Red eyes detection
- Smile detection
- Unshaved faces

Further sections will explain the classifiers in detail.

Age Classifier

The age classifier decides whether the detected face belongs to a person that is older or younger than 18. The features for classification are histograms of local binary patterns (LBP). Support Vector Machine (SVM) with radial basis kernel (RBF) is used for classification. The algorithm returns both classification result and the probability of the indicated class.

Figure 1. Coarse bokeh on a photo shot with an 85 mm lens and 70 mm **entrance pupil** diameter, which corresponds to *f*/1.2. "Josefina with Bokeh" by carlosluis – http://www.flickr.com/photos/paseodelsur/51805888/. Licensed under CC BY 2.0 via Commons – https://commons.wikimedia.org/wiki/File:Josefina_with_Bokeh.jpg#/media/File:Josefina_with_Bokeh.jpg.

Bokeh Effect Detector

In **photography**, **Bokeh** is the aesthetic quality of the blur produced in the out-of-focus parts of an image produced by a lens. Bokeh occurs for parts of the scene that lie outside the **depth of field** (Figure 1).

The algorithm, still a work in progress, is based on a combination of detection of out-of-focus blur (Leszczuk et al. (2014)) and detection of faces (Rusek et al. (2013), Rusek et al. (2014)).

The input is an RGB image of a face. The output is binary (Bokeh effect, no Bokeh effect) with confidence probability.

Dominant Color of Clothes (Dresses)

This classifier allows identification of the color of the dress of an actress in a so called "red carpet" photo. This scenario is useful for automated generation of content, when the designer of the system wants to achieve high diversity of photos presented to the user.

Technically this solution is based on face detection. The largest face in the image is identified, which allows identification of the sampling region located in the hip area.

For purpose of clustering the color is quantized into one of nine categories: red, green, blue, black, white, magenta, yellow, cyan and colorful.

Dominant Color Counter

This classifier is capable of counting and identifying the dominant colors in the image. It is compliant with the Dominant Color descriptor as described in the MPEG-7 standard ISO/IEC 15938.

Nudity Identification

The algorithm checks whether the input image contains nudity or not. It is based on a statistical model of human skin color and some additional shape information obtained with use of discrete cosine transform (DCT). Support vector machine (SVM) with radial basis function (RBF) kernel is used for classification. The result is the probability that the image contains nudity.

People Counter

The people counter module is a simple wrapper around the face detector from OpenCV Library. Haar cascades are used for face detection, and the number of detected face regions is assumed to be the number of people in the image.

Profile/En Face Classification

Preliminary work on profile classification was published in (Rusek 2011). The current implementation is far more accurate compared to the previous statistical model of the face image.

In this new approach, the face is classified as facing left, right, or front. Support vector machines are used for classification. The features used for classification are the concatenated histograms of local binary patterns (LBP) calculated over a 10 × 10 pixel sliding window. Before feature extraction, the size of a face image is normalized to 100 × 100 pixels.

Red Eye Detection

Taking a picture in bad light condition often ends with the red eye effect. Despite many correcting algorithms, the effect is still possible. Red eye is detected by combining two algorithms. The first algorithm locates the positions of the eyes in the face image. The second algorithm, for each pixel, calculates the probability that the pixel belongs to red eye. Probability is calculated using statistical model of the red eye color based on generalized linear models. A decision is made, based on the average redness, which is the average probability of a red pixel, weighted by the distance to located eye pupils.

Smile Detection

Research on smile detection has gained a lot of attention. Our approach relies on face detection in grayscale images. For each detected face region, smile is detected by support vector machines. The features used in the algorithm are the mean and standard deviation of coefficients of Haar wavelets decomposition of the image. These are the well-known texture descriptors. Such descriptors are collected over three scales ranging from an 8 × 8 to a 64 × 64 pixel sliding window. Before feature extraction, face region in the base image size is rescaled to 128 × 128 pixels.

Unshaved Faces

Detection of unshaved faces can be a complicated task because of the many styles of beard and moustache. In IMCOP, unshaved faces are recognized by support vector machines. The image is pre-processed by calculating skin probability for each pixel. The skin colour model is a Generalized Linear Model of RGB components. A nonlinear model up to third power is constructed. SVM input is the raw value of probability map scaled to 10 × 10 pixels.

Contribution to VIME

As described in the previous article, the VIME (Video and Image Models for consumer content Evaluation) project in VQEG (Video Quality Experts Group) is dedicated to No-Reference image quality and is currently collecting images to build a dataset that can be used to conduct our research work. VIME has set up a Flickr group where contributors can upload images to this dataset.

The indicators can contribute to VIME by automatically adding Flickr **machine tags**. A machine tag or **triple tag** uses a special **syntax** to define extra **semantic** information about the tag, making it easier or more meaningful for interpretation by a computer program. Machine tags comprise three parts: a **namespace**, a **predicate**, and a **value**.

References

[1] Baran, R., Zeja, A., Ślusarczyk, P., "An overview of the IMCOP system architecture with selected intelligent utilities emphasized," *Proceedings of Eight International Conference on*

Multimedia Communications, Services and Security (MCSS 2015), Nov. 2015.

[2] Leszczuk, M., Hanusiak, M., Farias, M. C. Q., Wyckens, E., Heston, G. "Recent developments in visual quality monitoring by key performance indicators.," *Springer Multimedia Tools and Applications*, pp. 1-23, Sept. 2014. DOI: 10.1007/s11042-014-2229-2.

[3] Rusek, K., Orzechowski, T., Dziech, A., "LDA for face profile detection," *Proceedings of Fourth International Conference on Multimedia Communications, Services and Security (MCSS 2011)*, 2011.

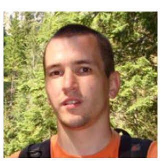
[4] Rusek, K., Guzik, P. "Neural network regression of eyes location in face images," *Proceedings of Sixth International Conference on Multimedia Communications, Services and Security (MCSS 2013)*, June 2013.

[5] Rusek, K., Guzik, P. "Two-Stage Neural Network Regression of Eye Location in Face Images," *Springer Multimedia Tools and Applications*, pp. 1-14, July 2014. DOI: 10.1007/s11042-014-2114-z.

Mikołaj Leszczuk
AGH University of Science and Technology, Poland

Remigiusz Baran
Kielce University of Technology, Poland

Michał Grega
AGH University of Science and Technology, Poland



Krzysztof Rusek
AGH University of Science and Technology, Poland

Piotr Guzik
AGH University of Science and Technology, Poland

VESA Advanced Display Stream Compression

James Goel

This eLetter article calls for proposals to support the perceptual and objective quality assessment of the next generation Video Electronics Standards Association (VESA) Advanced Display Stream Compression (Adv-DSC). VESA is interested in developing academic and industry partnerships to create new techniques for the both the objective and subjective analysis of low-impairment compression of HDR and wide-color gamut images. The VIME group within VQEG would also appreciate the addition of photographs described in this article containing high-dynamic range and wide-color gamut qualities to the VIME Flickr database. Interested universities and companies should contact James Goel (jgoel@qti.qualcomm.com).

> "To meet the growing need for visually lossless data compression, the Video Electronics Standards Association (VESA) has issued a new call-for-proposals for the next generation visually lossless display stream compression (DSC) system named Advanced-DSC (Adv-DSC)"

Adv-DSC Applications

Modern displays and mobile electronics must support high-bandwidth video signals within a low-power envelope and severe EMI constraints. The latest European Broadcasting Union (EBU) proposal for an 8K 10 bit 120Hz display format requires a completely new approach to content generation, production, transmission and display. The extreme data bandwidth required for this application is documented in Table 1 - Video Format Bandwidth CalculationTable 1. It is interesting to note the maximum 10240 × 4320 resolution as the

21:9 aspect ratio of the 16:9 FUHD format running at 120 Hz with up to 4:4:4 16 bit RGB pixel formats.

Visually lossless data compression may be used throughout the content delivery chain to meet the constraints mentioned earlier. The rest of this article discusses the visual quality analysis and subjective trial methodology that may be used to assess visually lossless, low-impairment compression codecs for these applications.

Table 1 - Video Format Bandwidth Calculation

Use-Case	Horizontal Pixels	Vertical Pixels	Frame rate	Pixel Format (4:2:0, 4:2:2, 4:4:4)	Comp per Pixel	Bits Per Component	Bits per Pixel	GPixels/ Sec	Bandwidth in Gbits/sec	Bandwidth in GBytes/s
	1920	1080	60	4:2:0	2	8	16	0.12	2.39	0.24
2K Production @60Hz	1920	1080	60	4:4:4	3	12	36	0.12	5.37	0.54
2K Broadcast @60Hz	1920	1080	60	4:2:0	2	10	20	0.12	2.99	0.30
8K Production @120Hz	7680	4320	120	4:4:4	3	12	36	3.98	171.99	17.20
8K Broadcast @120Hz	7680	4320	120	4:2:0	2	10	20	3.98	95.55	9.56
8K Production @60Hz	7680	4320	60	4:4:4	3	12	36	1.99	86.00	8.60
8K Broadcast @60Hz	7680	4320	60	4:2:0	2	10	20	1.99	47.78	4.78
10K Production @60Hz	10240	4320	60	4:4:4	3	12	36	2.65	114.66	11.47
10K Broadcast @60Hz	10240	4320	60	4:2:0	2	10	20	2.65	63.70	6.37

To meet the growing need for visually lossless data compression, the VESA has issued a new call-for-proposals for the next generation visually lossless DSC system named Adv-DSC. The requirements for Adv-DSC are outlined in Table 2.

Consumers expect to access this content on large displays plugged into the wall and on smaller mobile devices. The power and cost constraints placed on the electronics of both large and mobile displays require a compression solution capable of reducing the video bandwidth from 4:1 up to 6:1 times the original resolution with no perceived loss in visual quality. This last constraint is especially important since this DSC is applied in the last leg of the processing path, right before the data is converted to display pixels.

The first generation DSC succeeded in producing visually lossless results when converting 24 bits per pixel video down to 8 bits per pixel compressed (bpc) constant data rate. VESA adopted the DSC 1.1 standard this year for use in MIPI, embedded DisplayPort and MHL applications. It is based on algorithms optimized for low-cost, high-quality display applications and is described in the following website link (http://www.vesa.org/wp-content/uploads/2014/04/VESA_DSC-ETP200.pdf). DSC 1.1 was optimized for 24 bit to 8 bpc compression and the introduction of the new FUHD 8K standard requires more bandwidth compression to reach power and cost goals for future applications.

The Adv-DSC requirements were selected to give the standard enough capacity to support new products for the next five years, anticipating that in 2020 the Tokyo Olympics will be broadcast in 8K at 120Hz resolution. The large pixel component bit depth is required to maintain visually lossless quality throughout the production chain and support HDR and wide-color gamut.

Table 2 - VESA Advanced Display Compression Requirements

Attribute	Requirement	Comments
Resolutions	Up to 10240x4320	No interlaced format
Frame Rate	Up to 120 Hz	
Component Type	RGB, YCbCr; full-range, i.e. each component ranges from 0 to 2bpc - 1 in integer format	Input type to the encoder shall match the output type of the decoder. Internal color space conversion is permitted, but if used, it shall be specified as

Attribute	Requirement	Comments
		part of the proposal and included in the model.
Max Components	3	
Component bit depth	8, 10, 12, 14, or 16 bits	Referred to as bits per component (bpc) in this document
Sampling	4:4:4, 4:2:2 , 4:2:0	

Table 3 - Advanced DSC Coding Requirements

Attribute	Requirement	Comments
Coding across frames	No, intra-frame only	
Required coded bit rate that ensures visually lossless coding of source content with no chroma sub-sampling (4:4:4)	8 bpc source: 4.8 – 6 bpp 10 bpc source: 6 bpp 12 bpc source: 6 – 9 bpp 14-16 bpc source: 9 bpp	The encoding process shall guarantee that the specified bpp rate is met for all content. This requirement applies for all values of image attributes. For cases where a range of bpp values are given, the lower bpp target may involve some compromises (e.g., additional cost/memory, higher viewing distance for visually lossless quality, etc.) Visually lossless quality shall be maintained through at least two generations (coding/decoding cycles).
Required coded bit rate that ensures visually lossless coding of source content with chroma subsampling (4:2:2 or 4:2:0)	4:2:2: 10% lower than 4:4:4* 4:2:0: 20% lower than 4:4:4*	Bit rate reduction is measured relative to the bit rate required for visually lossless coding of 4:4:4. * Refer to the bit rate targets specified in the previous row of this table.

The most important requirement of Adv-DSC is the constant bit-rate (CBR) that must be maintained throughout the encoding and decoding stages. This CBR requirement requires a lossy compression system capable of appearing visually lossless to ensure consumer adoption and acceptable visual quality. The challenge of evaluating this low-impairment quality requires selecting difficult content that stresses the entire codec. The following Adv-DSC content types are identified by the call-for-technology:

Adv-DSC Section A.3 Content Types

Many types of still images will be evaluated:

- Continuous tone images
- Landscapes
- People portraits
- Animals
- Fine text, web pages and graphics
- Computer screen captures with or without sub-pixel rendering, etc.
- Test patterns such as noise and zone plates will be evaluated, but some visual loss may be tolerated on certain patterns.

Video tests will include movies, television, computer games, graphics, etc. Source video may be compressed using a standard broadcast compression algorithm before compression testing (e.g., MPEG-2, AVC, HEVC, etc.).

Adv-DSC Section A.4 High Dynamic Range (HDR) Testing

Figure 2 - Examples of Challenging DSC Test Material

Testing will include high dynamic range displays and content. Content will include BT.2020 10 and 12 bpc source as well as content sourced from other optical-electrical transfer functions (OETF). To cover different color gamut usage modes, content with sRGB, BT.709, and BT.2020 primaries will be tested.

The following example content is suggested by the ISO 291720-2 low-impairment standard.

The ISO 29170-2 Standard Protocol Description

Advanced DSC methodology follows the ISO/IEC 29170-2 subjective test methodology. This dual stimulus protocol alternates the reconstructed image with the original on one side while the other side remains static. The subject chooses either the left or right image as the image with a perceived flicker. If the subject cannot detect any flicker, they are asked to try their best and select either left or right image. The test cannot advanced unless a choice is made. Over a large number of observers, typically 40, the statistical analysis indicates whether a reconstructed image is visually lossless or not. The protocol uses a visually lossless JND threshold of 75%—in other words, if all 40 users are below the 75% detection threshold, the reconstructed image is deemed visually lossless.

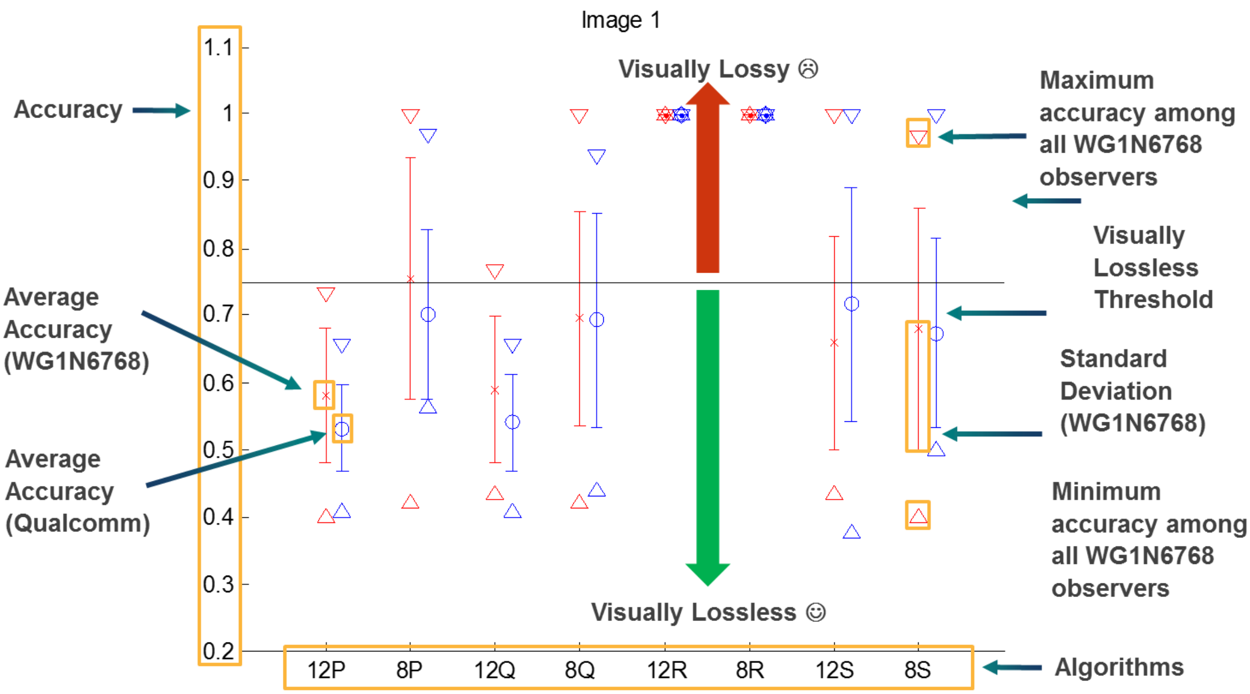


Figure 3 - Visually Lossless Subjective Trial Results

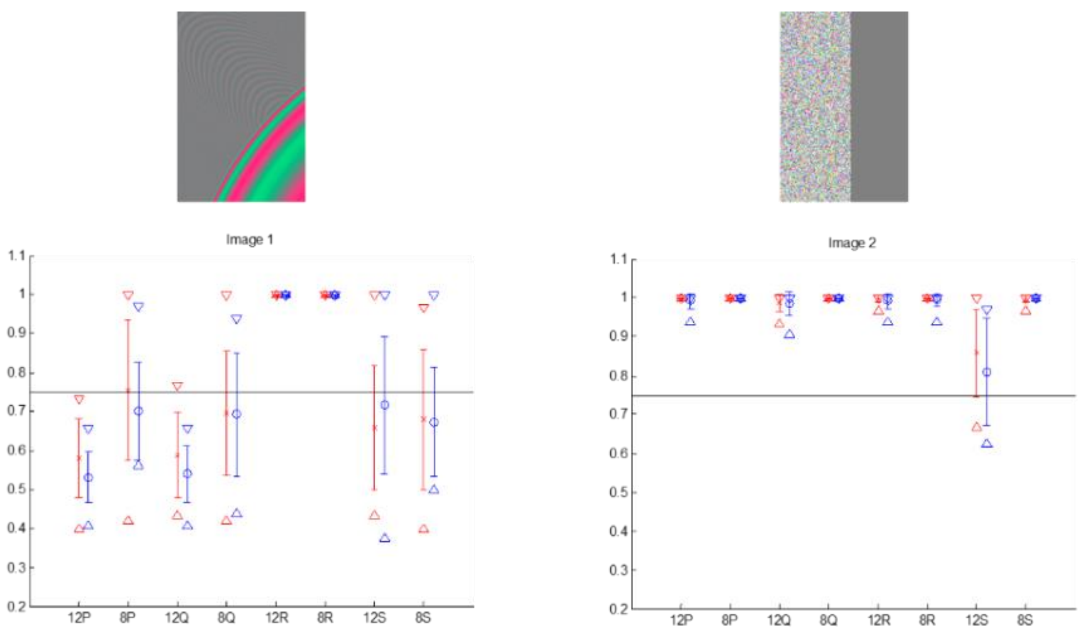


Figure 4 -- Challenging Subjective Trial Material

The selection of test material is the most important factor to ensure accurate results. The images range from artificial test patterns, natural images, faces and a mixture of computer graphics with text and text on natural backgrounds. This wide variety of material is required in many formats (10, 12, 14, and 16 bit; 4:4:4, 4:2:2, and 4:2:0 formats) to cover all requirements of the standard.

The VIME has established a new Flickr group to contain images that can be used for the Adv-DSC standard validation. The following list of HDR images are required to properly test the complete range of use cases. (see table above). VQEG invites all members to submit images from the list to the VIME Flickr group. Qualcomm is also extending an invitation to interested VQEG members who wish to participate in Adv-DSC subjective trials following this protocol.

Please Join the VIME Group and Upload your Photos

The test protocol requires HDR and high-colour 10/12/14 and 16-bit images and video from the following scenarios shot in RAW Format:

- Pane color glass window
- Night campfires and flames
- Outdoor sunset
- Indoor scene with window and sunlight
- Birthday night shots
- Tiki torches
- Sparklers on birthday cakes or very bright candles
- Sunset with car and bright lights

VQEG – Development of Low-Impairment Subjective and Objective Analysis

The following tasks require development and the VQEG is well positioned to help:

1) Collection and Generation of Adv-DSC Test Content
    a) VQEG has the proper technical expertise to create or capture HDR and wide-color gamut test material using the VIME Flickr database structure
2) Subjective trials of Adv-DSC using ISO 29170-2
    a) Full subjective trials for challenging content described in the Adv-DSC call-for-technology
    b) Expansion of ISO 29170-2 to include diagonal motion to stress block based codec design
3) Development of new objective and subjective analysis techniques:
    a) Low-impairment high-dynamic range (HDR) content
    b) Low-impairment wide-color gamut content

VQEG members who are interested in collaboration on this topic should review the standards presented in this article and contact me at jgoel@qti.qualcomm.com.

VIME and Subjective Image Quality Tests

Michele A. Saad and Philip Corriveau

Work is underway within the VIME (Video/Image Models for consumer content Evaluation) workgroup at VQEG to develop tools for no-reference image quality evaluation that cater to consumers' needs. This means that the tools developed are meant to predict quality for images that are representative of real consumer scenarios. Subjective image quality research is at the forefront of this work and serves a number of purposes: 1) to understand the psychophysical response of subjects to images of disparate quality; 2) to understand the dependence of this response on the subjective study design (in other words, to understand how different study designs elicit different responses from subjects); 3) to generate data for use in objective no-reference image quality evaluation model development and testing.

"A subjective study design is described where no images included in the test contain simulated distortions. … The images in the test were scenes captured by a multitude of devices and the quality range in the test is entirely determined by the image quality delivered by the devices used to capture the scenes."

The CCRIQ Database

The recent work in [1] describes a subjective study design that is in line with the consumer-oriented image quality evaluation objective. The outcome of the study is an analysis of a new subjective design approach and a database of images with associated subjective scores which has been dubbed the Consumer Content Resolution and Image Quality (CCRIQ) database.

A subjective study design is described where no images included in the test contained simulated distortions. The images in the test however, were scenes captured by a multitude of devices and the quality range in the test was entirely determined by the image quality delivered by the devices used to capture the scenes. The devices used to capture the scenes included four device categories: two tablets, eleven phones, six compact cameras, and four digital single-lens reflex (DSLR) cameras. The devices were chosen to provide images ranging in resolution from 1 megapixel (MP) to 20 MPs. The devices were also chosen to span a wide range of optics characteristics (lens properties, sensor sizes) and post processing capabilities.

"The significance of an equivalent image set is that no two images in the same set are necessarily exactly the same in content."

Eighteen scenes were captured by each of the 23 devices used in the test. The set of photos pertaining to one image scene was defined as an *equivalent image set*. The significance of an equivalent image set is that no two images in the set are necessarily exactly the same in content, even though they are photo captures of the same scene (i.e., they are the same scene captured by multiple cameras). This is due to inherent differences in the cameras used to capture the photos (such as focal length and aspect ratio) as well as the differences in photo capture angle and fluctuations in the image scene (such as moving clouds) as the photographer changes cameras to reshoot the scene. Fig. 1 shows example scenes from two equivalent image sets. Notice how the scene content within one equivalent image set is not identical to the next. The sample images in Fig.1 also show a variety of camera responses. This approach to building the database of images for subjective testing is in contrast to the more traditional approach of getting a number of high quality reference images and simulating distorted versions of them by introducing artifacts such as blur, noise, or compression and transmission errors. In the more traditional case, the content of each distorted image exactly matches the content of its corresponding reference image.

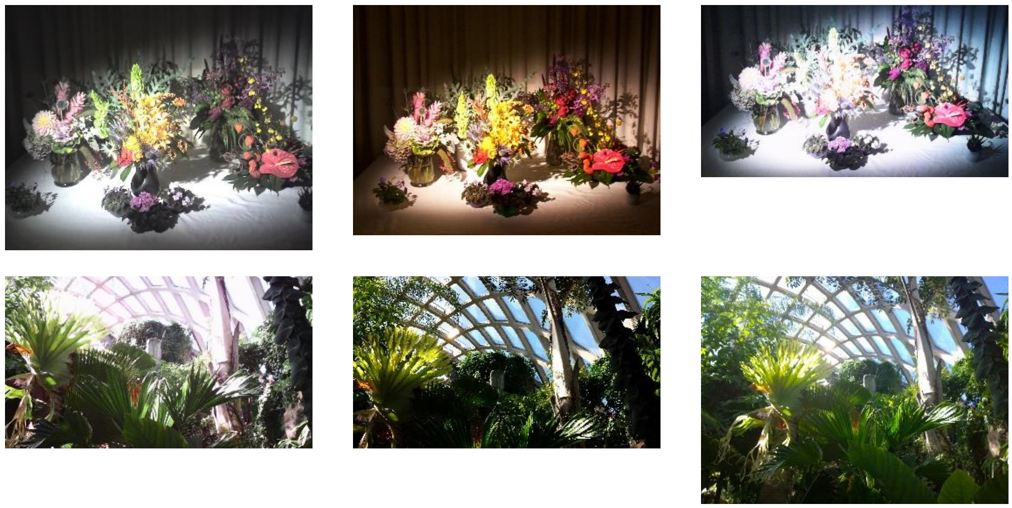


Fig. 1. These sample images show the large variety of camera responses within an equivalent image set. The first row of images beling to one equivalent image set. The second row of images belong to another equivalent image set.

The Subjective Test

The subjective test was performed across three laboratories: 1) NTIA's Institute for Telecommunication Sciences (NTIA/ITS) in Boulder, CO, USA; 2) Ghent University - iMinds, in Ghent, Belgium; and 3) Intel in Santa Clara, CA, USA.

A total of 392 images were rated on two 28'' monitors. One monitor was configured to HD (1920 × 1080) and the other was configured to 4K (3840 × 2160) resolution. The order of image presentation on one monitor or the other was completely randomized. Each scene was rated by either 26 or 27 subjects on each monitor. Table 1 shows the Pearson correlations between the MOS scores obtained from the three labs. The high correlation is an indication of the stability of the experiment design.

	NTIA/ITS	Ghent University	Intel
NTIA/ITS	1	0.952	0.941
Ghent University		1	0.915
Intel			1

Table 1: Pearson correlation between MOS scores of the different labs.

The MOS scores from the HD and 4K monitors were found to be highly linearly correlated. No statistically significant difference was found between the HD and 4K MOS scores on the lower quality images (images that scored a MOS less than 3 on the HD monitor). There was, however, a statistically significant slight difference between the HD and 4K MOS scores on the higher quality images, with the 4K MOSs being on average 0.2 MOS points higher than the HD MOSs.

Several factors impact the final image quality produced by a camera, including the optics and the post processing. Sensor size was found to impact ≈27% to 42% of camera quality on the CCRIQ data set. Among other observations in the study was that the overall quality difference between DSLR cameras and mobile cameras was found to be 0.67 MOS on the CCRIQ data set. Fig. 2 shows the MOS distributions between DSLRs, compact cameras, and mobile devices (phones and tablets). A high degree of overlap in the MOS histograms is observed pointing to the wide range of image quality produced by the different device types.

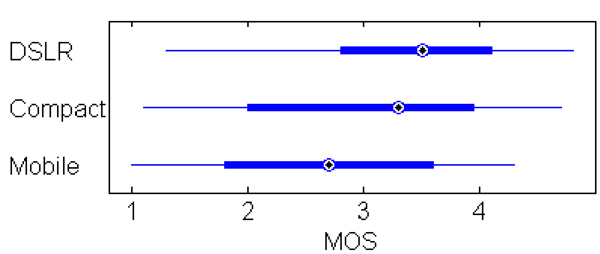


Figure 5: MOS distribution across three device categories.

We refer the reader to [1] for a more complete analysis. The CCRIQ dataset is available to the research community on the Consumer Digital Video Library (CDVL) at www.cdvl.org.

Michele A. Saad is a Senior Engineer and Researcher in perceptual image and video quality assessment at Intel. She received her Ph.D. degree in electrical and computer engineering from the University of Texas at Austin in 2013. Her research interests include statistical modeling of images and videos, motion perception, design of perceptual image and video quality assessment algorithms, and statistical data analysis and mining and machine learning.

Philip J. Corriveau is a Principal Engineer in Experience Development and Assessment in SMG at Intel. He directs a team of human factors engineers conducting user experience research across Intel technologies, platforms and product lines.

References

[1] M.A. Saad, M.H. Pinson, D.G. Nicholas, N. Van Kets, G. Van Wallendael, R. Da Silva, R.V. Jaladi, and P.J. Corriveau, "Impact of Camera Pixel Count and Monitor Resolution on Perceptual Image Quality", *IEEE Color and Visual Computing Symposium (CVCS)*, August 2015.

Combining HSV Color and rootSIFT for Image Retrieval

Ahmad Alzu'bi, Abbes Amira, Naeem Ramzan, and Tareq Jaber

Introduction

Content-based image retrieval (CBIR) [1] is a popular technique that has been widely applied to address the problems of traditional text-based image retrieval systems. CBIR is mainly based on the extraction process of low-level image features.

The proposed model introduces an optimized image descriptor that combines color and local features for image retrieval. Color histograms in HSV space are extracted as global features, while root scale-invariant feature transform (rootSIFT) descriptors are densely extracted as local descriptors. The extracted features are fused and encoded by the visual locally aggregated features (VLAD) approach. The Corel image dataset is used for performance

On one hand, local image descriptors describe local information using key points of some image parts, e.g. regions and corner points. The scale-invariant feature transform (SIFT) [2] is one of the most successful and commonly applied image descriptors. This feature is invariant to image scale and rotation and provides a robust matching across a range of fundamental image variations, e.g. noise addition, affine distortion, and change in illumination and 3D viewpoint. Some efficient variants of SIFT have been proposed such as dense SIFT (D-SIFT) [3] and rootSIFT [4].

On the other hand, color feature is one of the most extensive vision characteristic. In this work, HSV color space is used because the components of hue and saturation are closely related to the pattern of human visual perception. Consequently, combining local descriptors with global features is an essential part of the presented model [5].

Image Retrieval Framework

As shown in Fig. 1, the retrieval model is implemented as correlated blocks that represent images using the extracted descriptors as vectors. Firstly, D-SIFT descriptors are densely extracted from every image in the dataset at 7 scales by a factor $\sqrt{2}$ between successive scales, bin size of 8 pixels wide and a step of 4 pixels. Then, Hellinger kernel is applied to form rootSIFT descriptors and measure the similarity between SIFT descriptors, which yields superior performance in most cases without increasing processing or storage requirements. The 128-*D* rootSIFT descriptor is reduced to 100-*D* vector by the principal component analysis (PCA) projection.

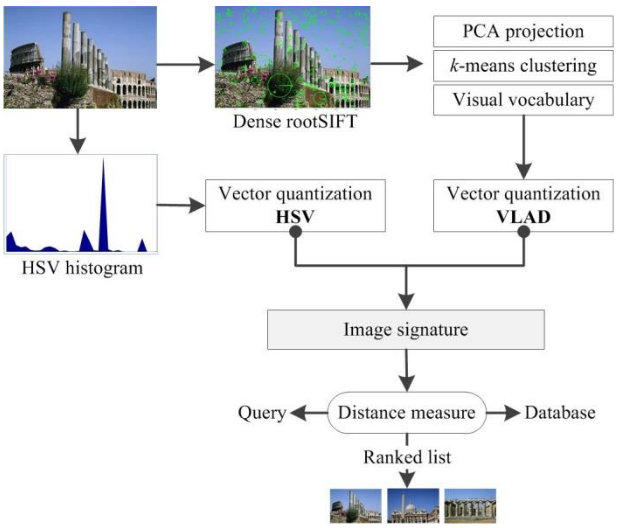


Fig. 6. The general framework of retrieval model.

The vector of locally aggregated descriptors (VLAD) [6] encoding is applied to quantize descriptors into vectors. It is a simplified non-probabilistic version of Fishers kernels, which is trained using *k*-means to accumulate the local descriptors and then normalized by L2 norm. A visual vocabulary of 256 clusters built by *k*-means approach is used in all experiments as a moderate size to keep balance between high discriminative image signature and low computation time.

The color value C is determined in the quantization by the equation $C = 8H + 2S + 2V$, where C is an integer between zero and 31.

Secondly, HSV histograms are extracted and quantized from the whole image as a global color feature. The H component is quantized into eight ranks non-uniformly and the S and V components are quantized into two ranks uniformly. Consequently, H and S are combined into a histogram of only 32 bins which is the representing color feature of each image.

Thirdly, the quantized HSV vector is combined with the VLAD vector to form the final signature of each encoded image in the dataset. Finally, the query image is matched with all dataset images using the Euclidean distance measure and the returned images are ranked based on the similarity scores obtained.

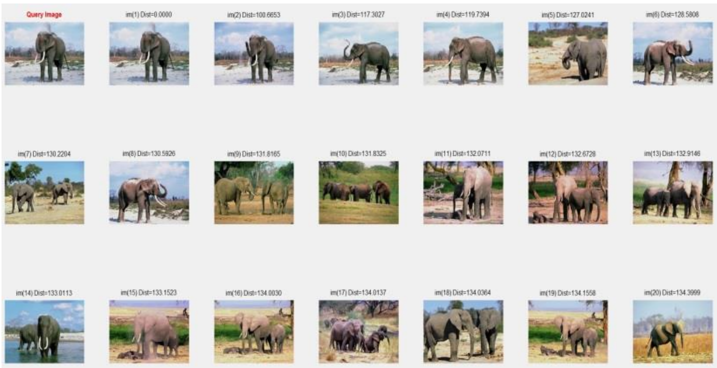


Fig. 2. Sample of query image and top 20 relevant images obtained.

Experiments

A subset of Corel image dataset[2] is used for experiments, which is widely utilized in computer vision applications. It consists of 1000 colorful images with 10 different semantic categories and each category contains 100 different images.

The Euclidean distance measure is computed between features transformed into n-dimensional vector of the query image and each vector in the image dataset in order to retrieve the relevant images. The mean average precision (mAP) is then obtained at different ranking positions to evaluate the retrieval accuracy.

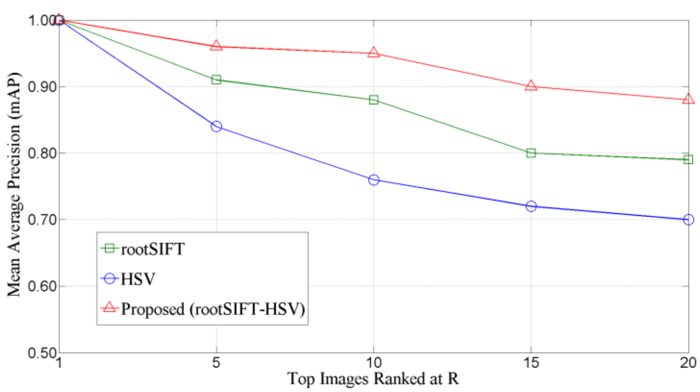


Fig. 3. The retrieval accuracy of three different methods rootSIFT, HSV, and the combined feature (i.e. Global and Local).

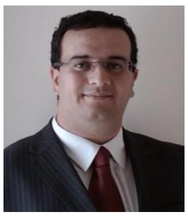
In every testing iteration, five query images are randomly selected from each dataset category and the process is repeated 10 times. Fig. 2 presents a sample of retrieval results. The retrieval accuracy (mAP) is reported over all queries at a range of ranking positions (R), i.e. top R at 1, 5, 10,

[2] http://wang.ist.psu.edu/docs/related/

15, and 20. Fig. 3 shows the best results of retrieval accuracy achieved using different image representations. It is clear that the proposed model outperforms both HSV and rootSIFT at all positions of rank images. At the top 10, 15, and 20 ranked images the retrieval accuracy improved by approximately 20% and 10% over HSV and rootSIFT, respectively.

The retrieval model achieves an efficient performance in terms of retrieval time and memory usage as follows: the actual memory size of each image vector is 100 KB, the average time (AT) elapsed to formulate the image vector is 0.445 seconds, the AT elapsed to search the whole image dataset and show the top 20 relevant images to the submitted query image is 0.795 seconds.

Ahmad Alzu'bi *is a PhD researcher at school of engineering and computing, University of the West of Scotland, Paisley, UK.*



Abbes Amira *is a Professor in visual communications, school of engineering and computing, University of the West of Scotland, Paisley, UK.*

Naeem Ramzan *is a Reader in visual communications, school of engineering and computing, University of the West of Scotland, Paisley, UK.*

Tareq Jaber *is an Assistant Professor, department of information systems, University of Jeddah, Saudi Arabia.*

References

[1] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic Content-based Image Retrieval: A Comprehensive Study," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 20-54, 2015.

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[3] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," In *ACM Proceedings of the International Conference on Multimedia*, pp. 1469-1472, 2010.

[4] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," In *CVPR*, pp. 2911-2918, 2012.

[5] A. Alzu'bi, A. Amira, N. Ramzan, and T. Jaber, "Robust Fusion of Color and Local Descriptors for Image Retrieval and Classification," *In IEEE Proceeding of 22nd conference On Systems, Signals, and Image Proceesing (IWSSIP)*, 2015.

[6] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 34, no. 9, pp. 1704-1716, 2012.

Open Source No-Reference Toolset Moving Image Quality Forward in New Ways!

Philip Corriveau

Image quality is a measure of the perceived goodness (or degradation) within an image. Common measures of image quality have generally been based on the fidelity of a test image to a corresponding original scene or perfect image. In the context of photo capture scenarios, image quality is the stand alone quality of photos taken by a device. Smartphones and tablets are increasingly becoming primary cameras for consumers. Mobile device vendors are striving at differentiating through camera capabilities and delivered quality of captured photos and videos. Camera quality is driving consumer preference in mobile device purchasing and impacts popular consumer usage models like photo and video sharing (Instagram, Facebook, Vine, etc.) and video conferencing. The Video Quality Experts Group has had a long history of driving innovation around algorithm and tool development mapped back to users' expectations. Emerging consumer paradigms are driving the requirement to dynamically adapt and refresh the status quo within VQEG's projects, especially in the imaging space. The changes continue with the offering of an open source no-reference image tool set called VIQET (VQEG Image Quality Evaluation tool). True collaboration can only happen when open source technologies are contributed to a talented pool of innovators. It is for this reason that a preview release of VIQET has been added to www.GitHub.com/VIQET. The current version of the

True collaboration can only happen when open source technologies are contributed to a talented pool of innovators.

Philip J. Corriveau is a Principal Engineer in Experience Development and Assessment in SMG at Intel. He directs a team of human factors engineers conducting user experience research across Intel technologies, platforms and product lines.

algorithm is embodied in a desktop tool and the installer can be found at: https://github.com/VIQET/VIQET-Desktop/releases. The actual open source code for the algorithm can be found at: https://github.com/VIQET/VIQET-Desktop .

The VQEG group is inviting not just members of VQEG, but everyone in the industry and academia to contribute to this new open source image quality toolset.

Meeting & Conference Announcements

The next full VQEG meeting will be Feb. 29 to Mar. 4, 2016, in San Diego, CA, USA. http://www.its.bldrdoc.gov/vqeg/meetings/san-diego-ca-usa-feb-29-to-mar-4-2016.aspx

A special session “Psychophysiological Measures for Visual Quality” in IS&T Human Vision and Electronic Imaging (HVEI) 2016 is organized by VIME co-chairs in Feb. 2016.

The 8th International Workshop on Quality of Multimedia Experience (QoMEX) will be held June 06-08, 2016, in Lisbon, Portugal. The paper submission deadline is March 04, 2016.

We are pleased to announce a new Springer journal: *Quality and User Experience*. The key to excellent research is the community working on particular topic. The Quality of Experience community needs a way to exchange the ideas and new results. The *Quality and User Experience* journal was created to be this knowledge exchange platform. This journal is created jointly by Quality of Experience and User Experience communities, which makes it an exchange platform not only within QoE, but also between these two research communities. We would like to invite you to submit your work to this journal and read articles published in *Quality and User Experience*.

The *IEEE Journal of Selected Topics in Signal Processing* has issued a Call for Papers for a Special Issue on “Measuring Quality of Experience for Advanced Media Technologies and Services.” The submission deadline is March 1, 2016.