2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops

No-reference Harmony-guided Quality Assessment

Christel Chamaret and Fabrice Urban
Technicolor
975, avenue des Champs Blancs ZAC des Champs Blancs CS 17616 35576 Cesson Sevigne
christel.chamaret@technicolor.com, fabrice.urban@technicolor.com

Abstract

Color harmony of simple color patterns has been widely studied for color design. Rules defined then by psychological experiments have been applied to derive image aesthetic scores, or to re-colorize pictures. But what is harmonious or not in an image? What can the human eye perceive disharmonious? Extensive research has been done in the context of quality assessment to define what is visible or not in images and videos. Techniques based on human visual system models use signal masking to define visibility thresholds. Based on results in both fields, we present a harmony quality assessment method to assess what is harmonious or not in an image. Color rules are used to detect what part of images are disharmonious, and visual masking is applied to estimate to what extent an image area can be perceived disharmonious. The output perceptual harmony quality map and scores can be used in a photo editing framework to guide the user getting the best artistic effects. Results show that the harmony maps reflect what a user perceives and that the score is correlated to the artistic intent.

1. Introduction

When manipulating, editing, improving images, the best quality as well as a certain artistic intent are usually the finality. Nevertheless, although the issues related to objective quality assessment have been largely studied in the context of low level artifacts (blur, blockiness, jitter...), the artistic intent is a problem more subjective, leading to strong difficulties in modeling or generalization. As an intermediary indicator, aesthetic quality metrics based on high-level features intuitively related to beauty (colorfulness, line orientation, shape...) and rules of thumb (composition, rules-of-third, skyline...) are showing up recently in the community [11, 4, 23]. Depending on the application context, some approaches take advantage of a reference source or do the best effort without any reference when providing an absolute quality measurement. Color harmony theory is used in [15, 7] as a global image cue for the assessment of aesthetic quality.

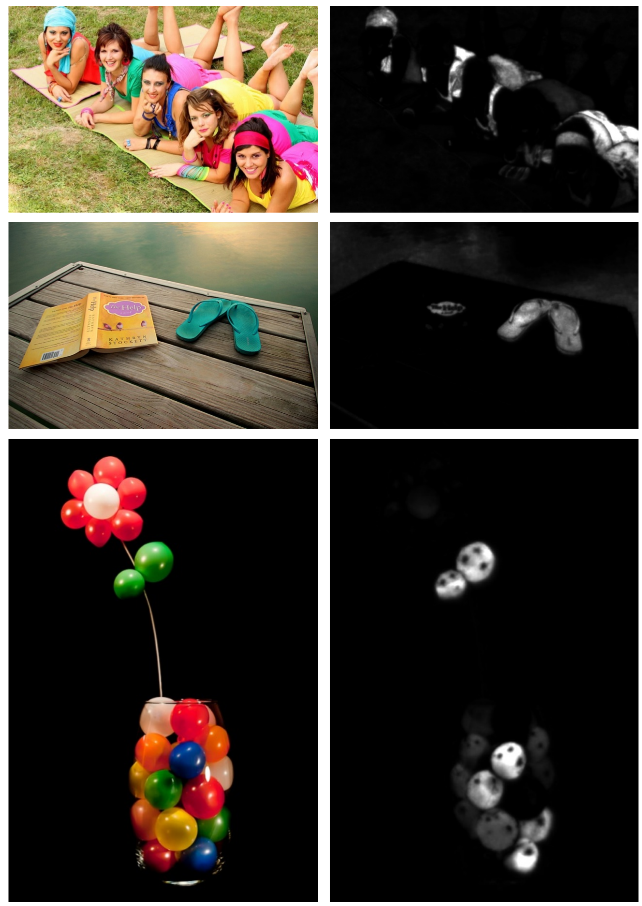

Figure 1. Is it possible to quantify the color harmony of a pixel in a picture? Left column is the original picture, right column is the harmony-guided quality map. The whitter the more disharmonious the pixels relatively to the global picture. On the bottom picture, balls of color can be sorted by disharmony level.

In this paper, we propose a new approach for assessing what the quality of a picture is. As an interesting tool for content creator and targeting a maximization of the artistic effect, the proposed metric provides a no-reference per-

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.161

ceptual harmony-guided quality map as well as a score of disharmony. To our knowledge, the use of color harmony concept in the context of perceptual quality assessment has been limited to the estimation of global image cues.

1.1. Quality Assessment

Perceptual quality metrics [22, 8, 13] take into account the human visual system properties to exploit local and global masking effects in evaluating image or video quality with full reference. They provide perceptual quality maps that mimic the human perception of degradations by highlighting visible degradations. In the same vein, the structural similarity index [19] is largely used due to its fair correlation with subjective judgment and its simplicity of implementation.

Previous work of Wang [19] has been extended to color by introducing hue similarity into the SSIM index [17]. Thakur and Devi [18] proposed a color quality index that performs in the spatial domain and takes advantage of the Human Visual System (HVS) properties to assess color quality with reference. Without any reference, Ouni *et al*. [16] have proposed different color statistics analysis (distribution of hue histogram, proportion and dispersion of the dominant color...) to derive a quality score, but they do not propose any local information such as a color quality map. The field of no-reference color quality assessment has been largely unexplored although the color is a well-known factor in the Human Visual System.

Nishiyama *et al*. [15] focus on color harmony theory to compute an aesthetic estimation of the input image. As it is usually done in aesthetic quality classification, features are extracted and learned from a classification ground truth. Although features are based on color harmony, there is no use of perceptual precept to derive perceptual quality maps.

1.2. Harmony

Harmony is a subjective concept whose difficulties lie on its definition and its measurement. Fedorovskaya *et al*. [5] investigated through a series of experiments the identification of image features involving harmony (dis)comfort. They found out that edge contrast, average lightness, range of lightness, may influence global harmony appreciation. Despite this previous reference, most of harmony investigations in image processing field relates to color combinations and complementaries.

Moon and Spencer [10] have performed psychological experiments to evaluate the relationship between color patterns and affective elicitation of stimulus on different subjects. They have developed a two-color harmony model. Similarly, Matsuda [9] extended Itten contrast measurement [6] where harmonious doublets and triplets of color are defined. Through his experiments, he defined HSV (Hue, Saturation, Value)-based templates that predict sets of harmonious hues (Figure 2). These templates have been used in re-colorization applications (referred to as *color harmonization*) by Cohen-Or *et al*. [2] and recently by Baveye *et al*. [1], where color mapping functions shift pixel hues into harmonious templates after choosing the most appropriate one, i.e. the closest to the original hue distribution.

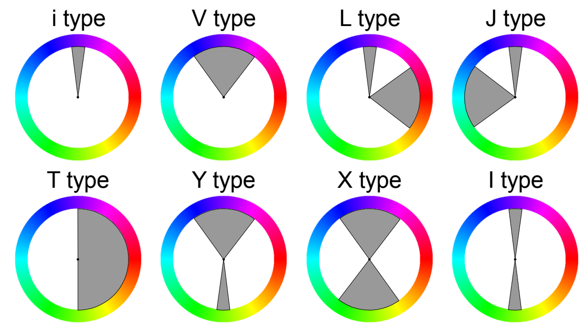

Figure 2. Harmonious hue templates defined by Matsuda experiments [9]. Grey sectors are sets of hues that are harmonious respectively to the complete hue wheel. The sectors width and layout are predefined and the template may be turned in the hue domain.

The main contribution of this paper is to use human visual system models defined for perceptual quality in the context of color harmony assessment.

2. Perceptual harmony quality assessment

The proposed method links two distinct topics: perceptual quality metric and harmonious hue templates. Figure 3 depicts different steps that are detailed in the next sections. The computation of harmony distance provides a disharmony map based on the hue distance from the current pixels to harmonious templates. Afterwards, the perceptual quality map is obtained by applying masking functions that provide information regarding local and global visibility. Those mentioned processing are done at different levels of resolution to mimic human visual system. A pooling step is then required to merge all resolution level maps leading to a perceptual harmony quality map (Figure 1). A final score is then computed from the map.

2.1. Notations

The eight harmonious templates $T_m, m \in \{i, I, L, T, V, X, Y, J\}$ are defined in [2], previously introduced in [9] and depicted in Figure 2. They have different sector layout and size to handle color complementary, color orthogonality and set of close hues. They can be turned around the hue wheel. Template m is defined as:

$$T_m : \{(\alpha_{m,k}, w_{m,k}); k = 1, ...K_m\} \qquad (1)$$

where $K_m \in \{1, 2\}$ is the number of sectors, $\alpha_{m,k}$ is the angle of the k-th sector's center on the hue wheel and $w_{m,k}$

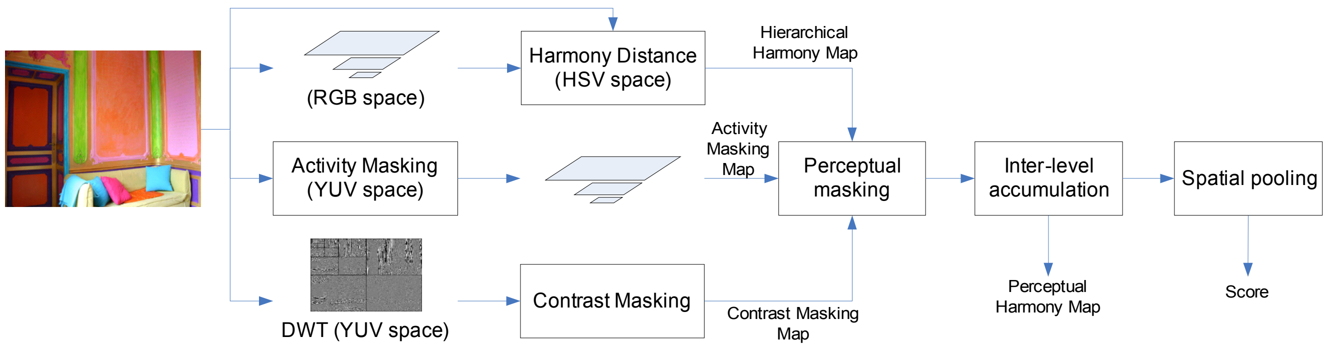

Figure 3. Overview of the complete system.

its width in degrees. For notation simplicity, α_m denotes the angle of the first sector, which is also referred to as the template angle. For a given picture, an appropriate rotation angle $\hat{\alpha}_m$ is computed to align at best T_m with the hue distribution of the image. It is the template angle that minimizes the Kullback-Liebler divergence between the normalized hue distribution $M(h)$ of the picture, typically in the form of an histogram with $L = 360$ bins, and the normalized hue distribution of the template, such as described by Baveye *et al.* [1].

$$\hat{\alpha}_m = \arg\min_{\alpha} \sum_{h} M(h) \ln\left(\frac{M(h)}{P_m(h-\alpha)}\right), \quad (2)$$

where $P_m(h)$ is the hue distribution of the template T_m with angle 0:

$$P_m(h) = \frac{1}{J}\sum_{k=1}^{K_m} P_{m,k}(h), \quad J = \sum_{h}\sum_{k=1}^{K_m} P_{m,k}(h) \quad (3)$$

$$P_{m,k}(h) = 1_{\{|h-\alpha_{m,k}|<\frac{w_{m,k}}{2}\}} e^{\frac{-1}{1-\left(\frac{2|h-\alpha_{m,k}|}{w_{m,k}}\right)^{10}}}. \quad (4)$$

The value $E_m = \sum_{h} M(h) \ln\left(\frac{M(h)}{P_m(h-\hat{\alpha}_m)}\right)$ represents the residual energy of template m for image at hand.

2.2. Harmony distance

The definition of harmonious templates reveals to be convenient information that may be arranged to provide a spatial harmony map. For a given template T_m, a hue h is considered harmonious if it is enclosed by a sector (meaning its harmonious distance is 0), while a hue outside the sector is not harmonious regarding a certain proportion defined by the hue distance $d_m(h)$. It is evaluated by computing the arc-length distance on the hue wheel (measured in degrees) to the closest sector:

$$d_m(h) = \min_{k=1\dots K_m} \left[|h-\hat{\alpha}_{m,k}| - \frac{w_{m,k}}{2}\right]^{+}, \quad (5)$$

where $|.|$ is the arc-length distance and $[.]^{+} = \max(0,.)$. Then, assuming that each template (associated with its optimal angle) provides harmony information about the picture, the d_m maps are computed for all templates and combined at the pixel level. At each pixel $\mathbf{u} = (x, y)$ with associated hue $h(\mathbf{u})$, the harmony distance map $\mathcal{G}(\mathbf{u})$ accumulates the harmony distances $d_m(h(\mathbf{u}))$ as follows. The contribution of each template is weighted according to its respective energy, to give more importance to well suited templates (having low energy);

$$\mathcal{G}(\mathbf{u}) = \left(\sum_{m}\left(1 - \frac{E_m}{\sum_{m'} E_{m'}}\right) d_m\left(h(\mathbf{u})\right)\right) \cdot s(\mathbf{u}) \cdot v(\mathbf{u}) \quad (6)$$

where s and v are saturation and value of the image. Weighting the harmony distance by saturation and value gives a more perceptual result because the more saturated the color or the highest its value, the strongest it is perceived. Some qualitative results are depicted in the second column of figure 4.

2.3. Perceptual masking

Including a perceptual masking that simulates the perception of the human visual system, we transform harmony distances into perceptual harmony maps. Spatial masking refers to the alteration of the perception of a signal by surrounding background, i.e, visibility increase (pedestal effect) or decrease (masking effect) due to the surrounding signal. As recommended by Watson *et al.* [21], both contrast masking and entropy masking are incorporated in the proposed quality metric. Contrast masking models the visibility change of the signal due to contrast values created by edges or color gradation. Entropy masking reflects the uncertainty of the masking signal, due to texture complexity. Entropy masking is also known as activity masking or local texture masking [14]; in the following, activity masking is used. The masking values are computed on the luminance channel of the YUV color space. The multi-channel behavior of the HVS is commonly simulated by multi-resolution analysis [13, 20]. Discrete Wavelet Transform (DWT) has proven to be efficient both in term of prediction and computation performances [13]. A CDF 9/7 (Cohen-Daubechies-Feauveau) kernel is used in our implementation with L decomposition levels. Each decomposition level comprises 3 orientation sub-bands (horizontal, vertical, and oblique frequencies). The spatial frequency range of a decomposition level $l \in [1; L]$ is $\left[2^{-l} \cdot f_{max}; 2^{-l+1} \cdot f_{max}\right]$ where f_{max} is the maximum spatial frequency of the image. The number of decomposition levels is set so that the lowest resolution level L contains the frequency 1 cycle/degree: $2^{-L} \cdot f_{max} < 1c/d < 2^{-L+1} \cdot f_{max}$. Contrast masking $C_{l,o}(\mathbf{u})$ at level l and orientation o is defined as the wavelet transformed value at site $\mathbf{u}$, weighted by the CSF (Contrast Sensitivity Function) that describes the variations in visual sensitivity to the spatial frequency. The CSF value $N_{l,o}$ is the mean value of the 2D CSF [3] over the spatial frequencies covered by the sub-band $w_{l,o}$ at level l and orientation o.

$$C_{l,o}(\mathbf{u}) = w_{l,o}(\mathbf{u}) \cdot N_{l,o}, o \in \{1, 2, 3\}. \tag{7}$$

The activity $A(\mathbf{u})$ is usually evaluated by the computation of entropy on a n-by-n neighborhood. Unfortunately, this tends to give high values on areas with a color step even between uniform areas, leading to overestimate the masking effects. Instead, the spatial gradient of the image luminance Y for different directions (horizontal, vertical, diagonal) are computed for each pixel at the full resolution and the minimum value is retained in the activity map:

$$A(x, y) = \min\left(\frac{\partial Y}{\partial x}, \frac{\partial Y}{\partial y}, \frac{1}{2} \cdot \left(\frac{\partial Y}{\partial x} + \frac{\partial I}{\partial y}\right)\right) \tag{8}$$

This map A is computed at different resolutions, yielding maps A_l's, to match the HVS simulation and the resolution of contrast masking maps.

In quality assessment models, contrast values are used to find a visibility threshold for the contrast difference between two images [14]. Here the studied signal being the color values, and not contrast and structure such as in quality assessment, the masking functions thus apply to the harmony map only. To apply masking functions, a multi-resolution analysis of the harmonious distance map is adopted to match the multi-channel behavior of the HVS. A multi-resolution image pyramid is first computed in the RGB space and converted into an HSV pyramid. Harmonious distance maps G_l, $l = 1...L$ are then computed as explained above, along the pyramid.

Then, perceptual masking is applied to harmonious distance maps at each resolution level. This step consists in applying contrast masking $C_{l,o}$ and activity masking A_l on the harmony distance map at each resolution l. Daly [3] and Nadenau [12] have proposed different complex intra-channel models for integrating the different kinds of masking effects. However this involves the tuning of parameters to adjust masking strength. Instead a simple masking function is used:

$$\mathcal{H}_l^*(\mathbf{u}) = \frac{\mathcal{G}_l(\mathbf{u})}{1 + \frac{1}{2}\left(\sum_o C_{l,o}(\mathbf{u}) + A_l(\mathbf{u})\right)}. \tag{9}$$

The contribution of disharmonious pixels is reduced when the masking effect is high.

2.4. Pooling and rating

This step consists in accumulating the perceptual harmony maps $\mathcal{H}_l^*$ for the different resolutions $l \in [1...L]$ to build the final perceptual harmony map $\mathcal{H}$. Finally, a score or rating is derived from this aggregated map. The perceptual harmony map is obtained with successive upscaling and combination of the map at each resolution level:

$$\mathcal{L}_{l-1}(\mathbf{u}) = \mathcal{L}_l(2^{-1}\mathbf{u}) + \mathcal{H}_l^*(2^{-1}\mathbf{u}), \ \ \mathcal{L}_L(\mathbf{u}) = 0. \tag{10}$$

The perceptual harmony map is the accumulation over all the resolution levels:

$$\mathcal{H}(\mathbf{u}) = \mathcal{L}_0(\mathbf{u}). \tag{11}$$

Note that $\mathcal{H}$ is visually close to the harmony distance map, but integrates masking effects by decreasing impact of colors in textured areas (Figure 4). The final image rating is defined as:

$$\mathcal{R} = \left(\frac{1}{W \cdot H}\sum_{\mathbf{u}} \mathcal{H}(\mathbf{u})^{\beta}\right)^{\frac{1}{\beta}}, \tag{12}$$

where W and H are respectively the width and height of the original picture and β is a parameter empirically set to 2.

3. Validation of method

Due to its novelty, the no-reference harmony-guided quality assessment suffers from a lack of ground truth. Ideally, a dataset with perceptual harmony maps as well as the associated score would be useful for the extensive validation of the proposed algorithm. It is planed in our future work to design and propose such a database. In the meantime, we propose 1) a qualitative assessment of harmony-guided maps and 2) an applicative context of assessment for harmony score. Two questions are thus raised: are the harmony-guided map visually consistent with user perception of disharmony? Is the harmony-guided quality score correlated to harmony improvement performed by a color harmonization algorithm?

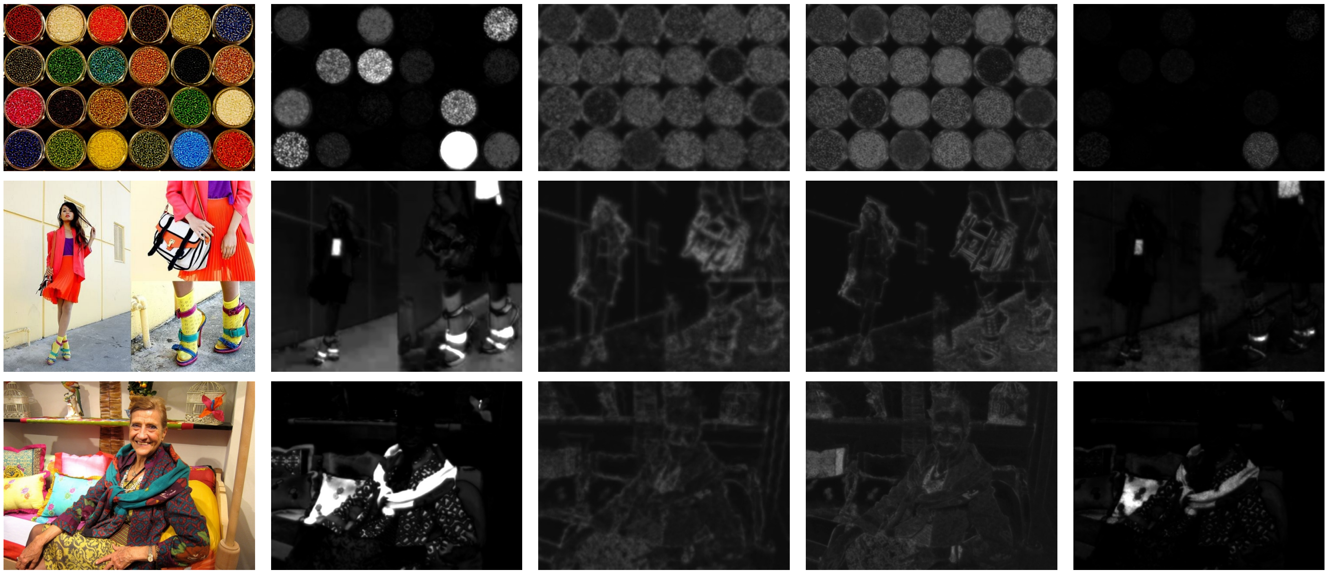

Figure 4. Visual appreciation of perceptual harmony-guided quality maps. The first column is the original picture, the second column is the harmony distance map, the third column is the contrast masking map and the fourth one is the activity map. Final maps are perceptual and harmony-guided in the sense they are close to harmony distance map but with applied masking effects.

3.1. The role of masking maps

As mentioned previously, the masking effect is a key concept in perceptual quality assessment. In this section, we demonstrate the interest of the two masking maps introduced in the context of harmony guidance. First row in figure 4 illustrates main advantages of masking modeling. The harmony map at the first row highlights blue and green areas as being disharmonious with orange/red major hues of the picture. The corresponding masking maps are not related to hues, but to visible spatial frequencies. Regions with complex texture are detected in the two masking maps: typically, black and clear areas with low contrast regarding the background are not masking areas, while red and green areas have complicated texture with high contrast. High masking contrast areas have a reduced contribution on the final perceptual harmony map.

In the second example, the gray/blue hues of the ground are detected as being disharmonious in the harmony distance map. However, the two masking maps clearly distinguished from the two areas of ground. Indeed, the ground of the right hand side of the picture is more textured leading to high contrast and potentially no clear analysis of HVS in this kind of area. The perception of the disharmonious hues in this area is thus masked by the image content. Consequently, the contribution of such area is minimized in the final perceptual map.

Finally, on the third row, the complementary roles of the two masking maps can be appreciated. The blue color is perceived as a disharmonious hue in this picture. Nevertheless, all the blue pixels can not be treated the same way, since masking effects alter the perception of some small group of blue pixels (jacket of the woman). High activity masking effects (4^{th} column) appear in the cushion (above the yellow one) leading to non-detection of non-harmonious blue hues by the HVS. Although green colors (particularly on the shelf) are also mentioned as having an average level of disharmony in the harmony distance map, they are detected in the contrast masking and then masked in the final perceptual map.

The two masking maps are complementary and allow the attenuation of disharmonious regions that are not perceived by the HVS due to their complex neighboring environment.

3.2. Harmony score

Due to the lack of dataset corresponding to such quality assessment, the score related to harmony-guided quality assessment has been benchmarked for different images and their harmonized version computed from [1]. As depicted in figure 5, the original and harmonized pictures are shown as well as their corresponding perceptual harmony-guided maps. As expected the harmonized areas have value of disharmony that have been attenuated in the final perceptual map. Moreover, their disharmony score has decreased as well.

One can notice that the color harmonization method leads to different quantitative improvement in terms of harmony, depending on the image content. The color mapping performs well for the Children and Sofa pictures (figure 5), i.e. the disharmonious areas are re-colorized leading to high decrease for both the perceptual map and the score. Nevertheless, some pictures are more tricky to harmonize. Indeed, their color distribution being more difficult to fit with a template, the color mapping process based on the most appropriate template does not bring high improvement in terms of harmony map or score (Socket picture of figure 5). In this case, the template selection can be questioned for color harmonization but also for the proposed metric. For such large color distributions, new more appropriate templates could be necessary, extrapolated for example from the selection of several regular templates. Computing the harmony distance map from all available templates and weighting by their energy could also be discussed: the weighting function may be based on more perceptual precept such as color appearance model.

4. Conclusion and perspectives

In this paper, a perceptual and no-reference harmony-guided quality metric has been introduced as a new way to assess picture quality. It relies on perceptual masking effects that are important properties of Human Visual System, and on harmony templates that have previously been designed during psychological experiments. The integration of these two concepts leads to the computation of perceptual harmony-guided map and an associated score. Both types of information may be useful in the context of image content creation, edition and retouching for guiding expert and non-expert content creators.

Qualitative results show that the harmony-guided map reflects the perception of color harmony by a human eye. Perceptual maps closely mimic the HVS to take into account masking effects and to discard disharmonious regions that are not perceived. The computation of score is consistent with potential changes of colors that can be done to improve picture harmony.

Our future development in this field relates to the design and creation of a ground truth database reflecting color harmony assessment by users to provide large-scale validation for this new field of computational metric.

References

[1] Y. Baveye, F. Urban, C. Chamaret, V. Demoulin, and P. Hellier. Saliency-guided consistent color harmonization. In *Computational Color Imaging*, volume 7786 of *Lecture Notes in Computer Science*, pages 105–118. Springer Berlin Heidelberg, 2013. 2, 3, 5, 7

[2] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Xu. Color harmonization. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 624–630. ACM, 2006. 2

[3] S. Daly. Digital images and human vision. chapter The visible differences predictor: an algorithm for the assessment of image fidelity, pages 179–206. MIT Press, Cambridge, MA, USA, 1993. 4

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Proceedings of the 9th European conference on Computer Vision - Volume Part III*, ECCV'06, pages 288–301, Berlin, Heidelberg, 2006. Springer-Verlag. 1

[5] E. Fedorovskaya, C. Neustaedter, and W. Hao. Image harmony for consumer images. In *Image Processing, 15th IEEE International Conference on*, pages 121–124, 2008. 2

[6] J. Itten. *The art of color : the subjective experience and objective rationale of color*. Van Nostrand Reinhold, New York, 1973. 2

[7] C. Li and T. Chen. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):236–252, 2009. 1

[8] M. A. Masry and S. S. Hemami. A metric for continuous quality evaluation of compressed video with severe distortions. *Signal Processing: Image Comm*, 19:133–146, 2004. 2

[9] Y. Matsuda. Color design. Asakura Shoten, 1995. 2

[10] P. Moon and D. E. Spencer. Geometric formulation of classical color harmony. *J. Opt. Soc. Am.*, 34(1):46–50, 1944. 2

[11] A. Moorthy, P. Obrador, and N. Oliver. Towards computational models of the visual aesthetic appeal of consumer videos. In *Computer Vision ECCV 2010*, volume 6315 of *Lecture Notes in Computer Science*, pages 1–14. Springer Berlin Heidelberg, 2010. 1

[12] M. Nadenau. *Integration of the Human Color Vision Models into High Quality Image Compression*. Ph.D thesis, Ecole Polytechnique Federale de Lausanne, 2000. 4

[13] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. On the performance of human visual system based image quality assessment metric using wavelet domain. In *SPIE Conference Human Vision and Electronic Imaging XIII*, 2008. 2, 4

[14] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. Which semi-local visual masking model for wavelet based image quality metric? In *Image Processing, 15th IEEE International Conference on*, pages 1180–1183, 2008. 3, 4

[15] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato. Aesthetic quality classification of photographs based on color harmony. In *CVPR'11*, pages 33–40, 2011. 1, 2

[16] S. Ouni, E. Zagrouba, and M. Chambah. Article: A new no-reference method for color image quality assessment. *International Journal of Computer Applications*, 40(17):24–31, February 2012. Published by Foundation of Computer Science, New York, USA. 2

[17] Y. Shi, Y. Ding, R. Zhang, and J. Li. Structure and hue similarity for color image quality assessment. In *Electronic Computer Technology, 2009 International Conference on*, pages 329–333, 2009. 2

[18] N. Thakur and S. Devi. Article: A new method for color image quality assessment. *International Journal of Computer Applications*, 15(2):10–17, February 2011. Published by Foundation of Computer Science. 2

[19] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. 2

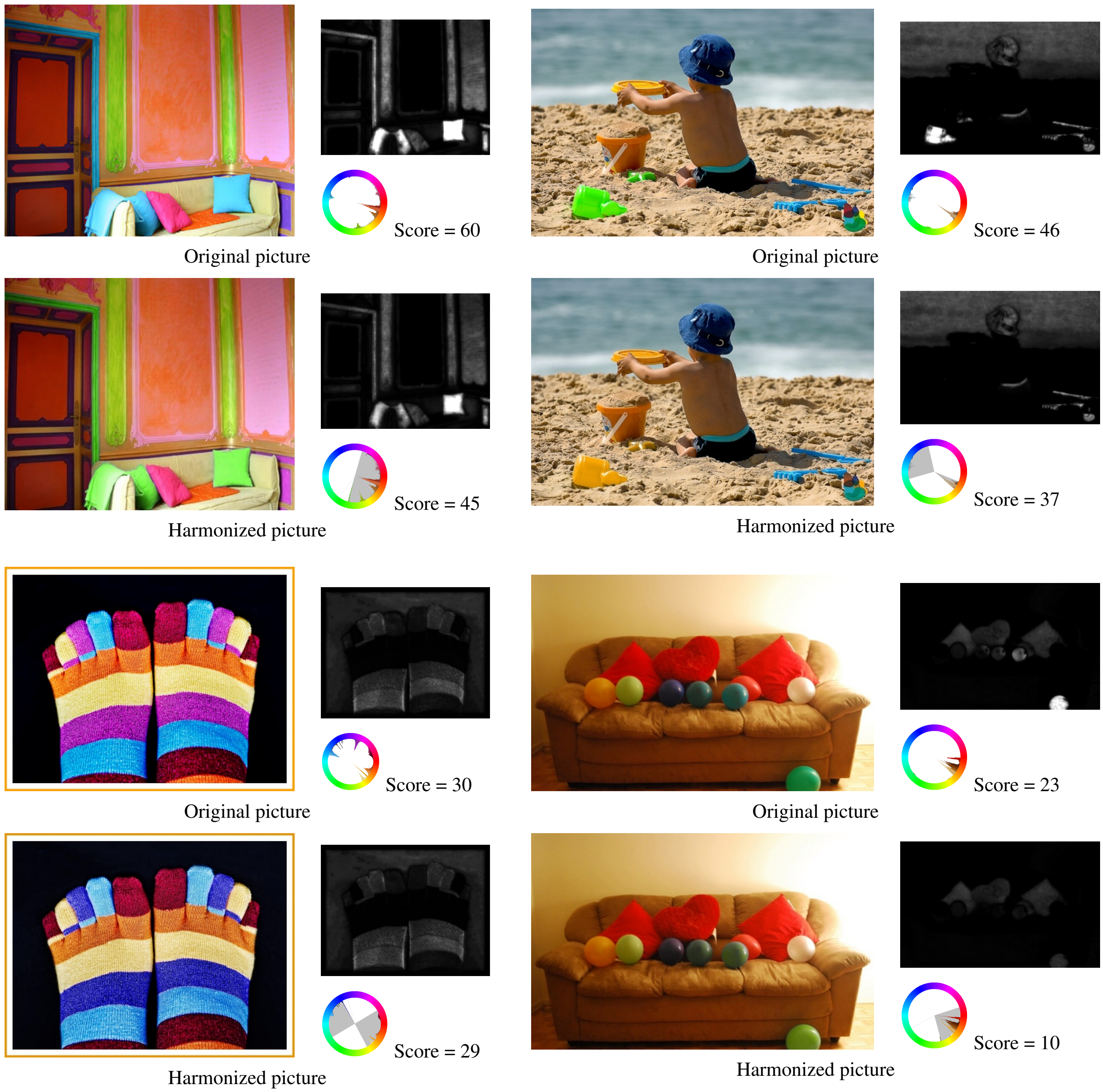

Figure 5. Harmony-guided quality assessment and Color Harmonization method [1]. For each example, original picture, harmonized picture as well as corresponding harmony map are provided. Hue wheel distribution is depicted for original picture and for the harmonized picture with the selected template. Harmonized areas appear less disharmonious on the maps, and scores of harmonized pictures are lower.

[20] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 1398–1402, 2003. 4

[21] A. B. Watson, R. Borthwick, and M. Taylor. Image quality and entropy masking. pages 2–12, 1997. 3

[22] S. Winkler. A perceptual distortion metric for digital color video. In *in Proc. SPIE*, pages 175–184, 1999. 2

[23] C.-Y. Yang, H.-H. Yeh, and C.-S. Chen. Video aesthetic quality assessment by combining semantically independent and dependent features. In *Acoustics, Speech and Signal Processing, IEEE International Conference on*, pages 1165–1168, 2011. 1