Semantic-Driven Selection of Printer Color Rendering Intents

Kristyn Falkenstern[1,2], Albrecht Lindner[3], Sabine Süsstrunk[3], and Nicolas Bonnier[1].
[1] Océ Print Logic Technologies S.A. (France). [2] Institut TELECOM, TELECOM ParisTech, LTCI CNRS (France). [3] School of Computer and Communication Sciences, EPFL, Switzerland.

Abstract

In this paper we introduce a unified framework that automatically selects the optimal color rendering intent for a given print job. We first present how we extract information from both the image features and the semantic information contained in keywords attached to this image. Then we show how our framework unifies the two inputs to select the optimal ICC rendering intent.

The framework is evaluated with a psychophysical experiment on an image data set printed with the ICC media-relative colorimetric and perceptual intents using an Océ large format printer. We find that our method is correctly able to predict the observers preferences in 81% of the images tested when the keyword is included compared to 58% when the keyword is not included.

Introduction

The aim of our research is to automatically generate optimal print reproductions of images using an inkjet printing system. The International Color Consortium (ICC) provides a consistent workflow to manage the color gamut changes between an original image and its reproduction via a given technology (ink or toner based print, silver halide photograph, electronic display). The ICC has defined four different intents to address the different reproduction objectives a user may have [1]. Each one represents a different color reproduction compromise.

In this work we focus on two of these intents, the perceptual and media-relative colorimetric intents. The media-relative colorimetric intent aims for a colorimetric match while the perceptual intent aims for a pleasing reproduction[1, 2]. The visual impact of the media-relative colorimetric intent is likely to cause the reproductions to appear more colorful and with more contrast and the perceptual intent will often prioritize the details over other qualities.

From this observation it appears that the selection, by a user looking for an optimal workflow, of one rendering intent versus the others may also depend on the document content. Perhaps for a very colorful image, the user pays more attention to the color reproduction over the details and chooses the rendering intent accordingly. For images with details, the user may choose another rendering intent that produces a better rendering of details. Our aim is to help the user by building a tool to automatically select the optimal rendering intent for a given print job.

Much effort has been made towards creating an adaptive processing workflow where the final processing is driven by the input document's content [3–5]. These adaptive workflows often use a training set of documents which require two inputs: 1. *features* and 2. *performance input*. The features used in these workflows help to summarize differences between documents or to group documents into categories. The terms statistics, properties, factors, image characteristics and descriptors have also been used to describe a document's features. The performance input is often the result of a psychophysical evaluation.

Our workflow requires the ability to easily change which ICC profile and rendering intent combination (*color workflow*) to apply. This requirement excludes the use of time-consuming psychophysical data as the performance input. Instead, we use a set of performance results derived from metric tests, where each metric compares the color workflow performance of a specific perceptible quality attribute [6].

The psychophysical validation results used to test the quality of our first implementation showed that the observers' preference between color workflows was more significant between rendering intents than between ICC profiles. For most test images, we were able to adaptively select the same rendering intent as the observers when the observers' preference between rendering intents was significant [7]. Our goal is to improve the automated selection of which rendering intent is optimal for documents that embed several conflicting image characteristics by using semantic information. The inclusion of the semantic information will add an understanding of the scene on a higher semantic level which will improve our prioritization of the conflicting characteristics.

The aim of this paper is to introduce a unified framework that automatically selects the optimal rendering intent for a given print job. We first present how we extract information from both the image and the semantic keywords. Then we show how our framework unifies the information to select the rendering intent. We then show some early results obtained with our framework to demonstrate its advantages and compare these framework results with the results of a psychophysical evaluation.

A Unified Framework for Image Features and Semantic Information

This section explains in two separate subsections, how to extract cues from two very different sources: 1) image pixels features and 2) semantic context. After this we complete the framework by uniting the two methods into a single estimation that can be used for the selection of the best rendering intent.

Cues from Numeric Pixel Values

We use eight quality attributes, which are Colorimetric Accuracy (CA), Colorfulness (CO), Gamut Boundary (GB), Smoothness (SM), Details (DE), Shadows (SH), Highlights (HL), and Neutrals (NT), respectively [7]. We denote the set of all quality attributes $\mathcal{Q}$. Each quality attribute is represented by 100 expert-selected example images.

For a new input image we assess its relatedness to a quality attribute by measuring similarity to its set of example images. This is a typical classification task and we implement a standard method with multivariate Gaussians.

Mathematical Background

Given an example set of images, for each quality attribute, we pre-compute a feature vector for all of the images. The feature vector can contain any type of pixel-based image descriptors, such as lightness, color, or texture features. As a first step we whiten the data by subtracting the mean and dividing through the variance in each dimension separately. The mean and variance are computed globally over all quality attributes q and images. For the rest of this paper the whitened features are used without being referred to as whitened. We estimate for each quality attribute the mean (μ_q) and covariance (Σ_q) matrix of its associated point cloud of images in the feature space.

Each quality attribute can then be represented as a multivariate Gaussian density distribution in feature space:

$$g_q(\mathbf{f}) = \frac{1}{(2\pi)^{N/2}|\Sigma_q|^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{f}-\mu_q\right)^T \Sigma_q{}^{-1}\left(\mathbf{f}-\mu_q\right)\right] \quad (1)$$

where $\mathbf{f}$ is a point in the whitened feature space, while $|\cdot|$ and $(\cdot)^{-1}$ are the determinant and inverse of a matrix, respectively.

For a new input image $\mathcal{I}$ we compute its feature vector $\mathbf{f}^{\mathcal{I}}$ and use the outcome of the Gaussian density estimators $g_q(\mathbf{f}^{\mathcal{I}})$ to quantify its relatedness to each quality attribute . Because it is based on numeric pixel values we call this **n**umeric **r**elatedness and denote it as $nr_q(\mathcal{I})$. Since an image relates to the distinct quality attributes in different proportions, we always scale the relatedness values so that they sum to 100%, i.e. $\sum_{q\in\Omega} nr_q(\mathcal{I}) = 100\%$.

Practical Implementation

The features for a classification task have to be chosen with respect to the intended application. In our printing workflow, the key image characteristics are texture and color. Texture is important because it relates to image sharpness that is limited by the print resolution. Color is crucial in two ways: gray levels have to be carefully reproduced since they are sensitive to color casts and very saturated colors might cause difficulties due to a printer's limited color gamut. Consequently we use features that describe the indicated characteristics.

We convert an input image $\mathcal{I}$ to CIELCH color space and compute the mean values of the lightness $\bar{L}$ and chroma values $\bar{C}$ and the entropy of the lightness channel E, respectively. Entropy is a statistical measure of randomness and can be used to characterize image textures. The complete feature vector is $\mathbf{f}^{\mathcal{I}} = \left[\bar{L}, \bar{C}, E\right]^T$.

Figure 1 shows two example images, $N07$ (top) and $N01$ (bottom) [8, 9], and their computed relatedness values nr. It is visible how image characteristics are reflected in the values. The top image $N07$ contains many details and saturated colors and thus the Details (DE) and Colorfulness (CO) quality attributes have the highest scores. The bottom image $N01$ has many neutral colors and details on the bride's dress and bouquet as expressed by high scores for Neutrals (NT) and Details (DE).

Cues from Semantic Image Context

The semantic context of an image can be given by associated keywords in the EXIF file header. Example annotations of the images in Figure 1 are *colors* and *blackwhite*, respectively. Adding this information to the decision process can add an understanding of the scene on a higher semantic level. For the $N07$ image, a keyword *colors* indicates that the colors are important to the user and thus the relatedness to Colorfulness (CO) has to be higher at the expense of Details (DE).

Figure 1. *Two input images, with associated numeric relatedness values nr for the eight quality attributes (horizontal axis). Top: the content of the image $N07$ relates to Details (DE) and Colors (CO). Bottom: the image $N01$ contains details (DE) and many neutral colors (NT).*

The goal is to estimate for a given keyword its relatedness to the eight quality attributes, as we did for a given image. The challenge is that keywords come from an uncontrolled vocabulary and we do not want to limit it, because this has considerable drawbacks. One drawback is the difficulty to choose a subset that is small enough to handle, but large enough to cover the semantic space. Our statistical approach handles vocabulary sizes in the order of thousands so that we can cover a significant part of the semantic space.

We use a statistical significance test to estimate the relations between the three image features and an arbitrary keyword [10]. The test is based on the MIR Flickr database, which contains one million high quality images from Flickr along with their annotations given by the internet community [11].

Theoretical Background

We start by defining the set of all feature values (for each feature) from images with a specific keyword w in the annotation string and designate a second set for the remaining images, the disjoint set. In order to quantify how a keyword w influences the given image feature, the feature distributions of the two disjoint sets are compared. This is accomplished with the Mann-Whitney-Wilcoxon ranksum test that assess whether the medians of two populations differ [12, 13].

We have a means of measuring the significance a given keyword has on the feature vectors by determining the significance value (z) of the normalized ranksum statistic (T), with its expected mean (μ_T) and variance (σ_T^2), [10].

The used MIR Flickr database contains one million annotated images [11], which makes it possible to compute significance values for a large number of keywords. There are 2858 keywords in the database that occur in at least 500 image annotations. Based on our experience, this is a scale where statistical estimations are robust enough to be exploited in this framework.

The significance value z_w indicates whether the underlying feature has the tendency to be larger ($z_w > 0$), smaller ($z_w < 0$) or unaffected ($z_w \approx 0$) for images annotated with keyword w. To give a better intuition about the z values we show some examples that appear in the database. For the lightness mean we obtain for example $z^{\bar{L}}_{shadows}= -21$ and $z^{\bar{L}}_{design}= 25$, for the chroma mean $z^{\bar{C}}_{blackwhite}= -86$ and $z^{\bar{C}}_{colors}= 63$, and for the entropy $z^{E}_{red}= -28$ and $z^{E}_{trees}= 27$. Keywords with z values close to zero are $z^{\bar{L}}_{water}= 0.41$, $z^{\bar{C}}_{place}= 0.1$, and $z^{E}_{libraryofcongress}= 0.04$, respectively. The example z values clearly show that the significance measure is a solid estimate for what a keyword implies for an image feature – or what it does not imply.

Practical Implementation

We propose a simple yet effective way to compute the relatedness of a keyword w to the eight quality attributes using significance values. The example image set of the quality attribute q is centered at $\mu_q = \left[\mu_q^{\bar{L}},\ \mu_q^{\bar{C}},\ \mu_q^{E}\right]^T$ in the training feature space. We then define for all features $f \in \mathcal{F} = \{\bar{L}, \bar{C}, E\}$ the minimum and maximum values:

$$\begin{aligned} m^f &= \min_{q\in\mathcal{Q}} \mu_q^f \\ M^f &= \max_{q\in\mathcal{Q}} \mu_q^f \end{aligned} \qquad \forall f \in \mathcal{F} \tag{2}$$

If a z_w^f value for keyword w and feature f is positive, the quality attribute q that has the maximum mean value M_q^f is most related to the keyword and the quality attribute with the minimum mean value m_q^f is least related. For a negative z_w^f value, it is the opposite. We thus estimate a keyword w's **s**emantic **r**elatedness $sr_q(w)$ to a quality attribute q as:

$$sr_q(w) = \sum_f sr_q^f(w), \qquad \forall q \in \mathcal{Q} \tag{3}$$

$$sr_q^f(w) = \begin{cases} \frac{\mu_q^f - m_q^f}{M_q^f - m_q^f} z_w^f & z_w^f \geq 0 \\ \frac{\mu_q^f - M_q^f}{M_q^f - m_q^f} z_w^f & z_w^f < 0 \end{cases} \qquad \forall q \in\ Q, f \in \mathcal{F} \tag{4}$$

Similar to the nr values we scale also the sr values so that they sum to $\sum_{q\in\mathcal{Q}} sr_q(w) = 100\%$.

Figure 2 shows the semantic relatedness values for four keywords *colors*, *nature*, *fabric*, and *blackwhite*, respectively. The bar plots show that *colors* and *nature* are both related to Colorimetric Accuracy (CA), Colors (CO), Gamut Boundary (GB) and Smoothness (SM) whereas *fabric* and *blackwhite* relate stronger to Details (DE), Neutrals (NT), Highlights (HL) and Shadows (SH). The keyword *blackwhite* is only very weakly related to Colors (CO) and not at all to Gamut Boundary (GB).

The plot also shows that two different words can stand for similar image characteristics in terms of ICC rendering intent. The keywords *colors* and *nature* both imply colorful images. As the statistical framework detects these relationships automatically on a large scale, it is not necessary to define them as synonyms.

Unification of Numeric and Semantic Cues

In the two previous subsections we have presented methods to estimate the relatednesses of quality attributes to numeric image pixels based features and to semantic image context, respectively. The numeric relatedness $nr_q(\mathcal{I})$ takes as input an image $\mathcal{I}$, and the semantic relatedness $sr_q(w)$ a keyword w.

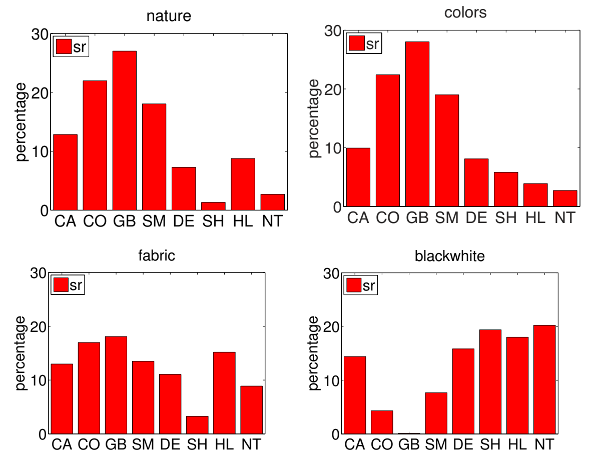


Figure 2. Semantic relatedness values sr for four keywords as indicated in the titles for the eight quality attributes . The keywords colors and nature both imply the importance of Colorfulness (CO) and Gamut Boundary (GB), whereas fabric and blackwhite relate stronger to Neutrals (NT) and Details (DE). Note that colors and nature are similar in terms of their meanings for ICC rendering intents.

In order to estimate how an image in a specific context relates to the quality attributes, the values have to be united in a single relatedness measure $r_q(\mathcal{I}, w)$. Because the nr and sr values express probabilities, it is reasonable to use a multiplication:

$$r_q(\mathcal{I}, w) = nr_q(\mathcal{I}) \cdot sr_q(w), \qquad \forall q \in \mathcal{Q} \tag{5}$$

In this early work, multiplication has been the only way used to combine the relatedness measures, other methods should be investigated for any future work. Figure 3 reproduces the two example images from Figure 1, but with an associated semantic context given by a keyword. The bar graphs show the numeric $nr_q(\mathcal{I})$ and semantic relatedness $sr_q(w)$ as well as the united relatedness $r_q(\mathcal{I}, w)$. Note how the relatedness values adapt to both the image and its context, enabling an automatic semantic-driven selection of the printer ICC rendering intent.

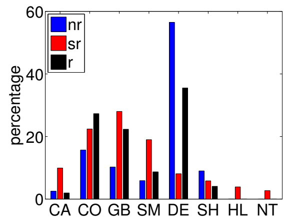
From Relatedness Measures to Rendering Intents

We are interested in understanding how the keywords impact the choice of rendering intent, which includes the numerical relatedness (nr) and the unified relatedness (r), and no longer the semantic relatedness (sr) results. We combine the metric ranked results (rRI) from Table 1 first using the numeric relatedness (nr) results and then the unified relatedness (r) with the following:

$$\chi_{k,j} = \sum_{q\in\mathcal{Q}} (k \cdot rRI)_{q,j}\,, \qquad k \in \{nr, r\} \tag{6}$$

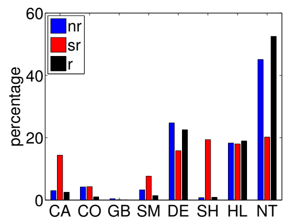
where rRI is the ranked metric score of the two rendering intents, q is the quality attribute index, j is the rendering intent index 1:m, and $m = 2$ the number of rendering intents in this experiment. The final selection (S_k) is determined by: $S_k = \underset{j}{\operatorname{argmin}}(\chi_{k,j})$. Our method of combining the metric results and the relatedness scores

(a) semantic context: *colors*



(b) relatedness values

(c) semantic context: *blackwhite*



(d) relatedness values

Figure 3. *Same images as in Figure 1, but with a semantic context indicated by the keyword in the caption. The bar graphs show the numeric* $nr_q(\mathfrak{I})$, *semantic* $sr_q(w)$ *and united relatednesses* $r_q(\mathfrak{I},w)$, *respectively. Top: the semantic context* colors *reduces the relatedness to Details (DE) and increases it for Color (CO) and Gamut Boundary (GB). Bottom: the semantic context* blackwhite *makes Details (DE), Highlights (HL) and Neutrals (NT) the dominant quality attributes.*

has not yet been proven as the most optimal. Other methods should be considered and compared for future work.

Since nr and r have been scaled so that they sum to 100% then the following holds true, $\sum_{j=1}^{m} \chi_{k,j} = m\frac{(m+1)}{2}$ and the scores from 6 can be scaled so the sum of χ_k is equal to 100%. Let the index $j=1$ correspond to the perceptual intent and $j=2$ correspond to the media-relative colorimetric intent, so that if $m=2$ then if $\chi_{k,1} < 50\%$ the perceptual intent is selected and if $\chi_{k,1} > 50\%$ the media-relative colorimetric intent is chosen. Additionally, when $m=2$ then $\chi_{k,1}$ can be inferred from $\chi_{k,2}$ and vice versa for the same k. For the rest of this paper we refer to only the perceptual intent $\chi_{k,1}$ when discussing the rendering intent scores. To simplify we set $\chi_{nr} = \chi_{nr,1}$ and $\chi_r = \chi_{r,1}$.

We express the keyword's impact on our selection as, $d = \chi_r - \chi_{nr}$. When $d > 0$ it is illustrated with $\uparrow$, indicating a move towards the media-relative colorimetric intent and when $d < 0$ a $\downarrow$ is used to illustrate a move towards the perceptual intent.

Experimental Validation

For our first implementation we chose to compare the perceptual and media-relative colorimetric rendering intents for a custom ICC v2 output printer profile. The ranked metric results used in Equation 6 are summarized in Table 1, along with a description of the metrics used and their results. As is expected, the Perceptual (Per) intent performs better with the more details related attributes, while the Media-Relative intent performs better with Color Accuracy and Colorfulness.

Demonstration of the Framework

We first consider the images in Figure 3 to compare the numeric framework that uses only the image features to the unified framework that uses the semantic information along with the image information. As shown in Table 2, the addition of *colors*

Table 1: The evaluation results of each quality attribute. Each quality attribute is listed with the metric used to assess it, the type of document used in the assessment, the results and the ranking of the Perceptual (Per) and Media-Relative colorimetric (MR) rendering intent, [7].

Quality Attribute	Metric	Per(rRI)	MR(rRI)
Color Accuracy	CIE ΔE^*_{94} (in gamut target)	3.79(2)	2.63(1)
Colorfulness	ΔCui [14] (targets)	1.81(2)	0.24(1)
Gamut Boundary	CIE ΔLCH^*_{ab}[6] (out of gamut gradients)	0.93(2)	0.95(1)
Smoothness	2^{nd} Derivative[15] (gradients)	2.73(1)	3.00(2)
Details	Visual Information Fidelity [16] (images)	0.21(1)	0.26(2)
Shadow Details	CIE ΔL^*_{STDV}[7] ($\leq L^*30$)	0.15(1)	0.48(2)
Highlight Details	CIE ΔL^*_{STDV}[7] ($\geq L^*75$)	0.02(1)	0.03(2)
Neutrals	CIE $C \times \Delta H_{ab}$[6] (ΔL^* target)	0.71(2)	0.66(1)

changes the selection for image $N07$ from the perceptual intent ($< 50\%$) to the media-relative intent ($> 50\%$). However, when the keyword is *shop* the selection does not change, it moves further towards the perceptual intent. For the $N01$ image, if the framework uses *blackwhite* the selection does not change, the inclusion of *fabric* causes a change in selection to the perceptual intent.

Table 2: A comparison of the scaled results obtained with our framework, first the selection with the numeric solution and then the unified framework. The Perceptual (Per) intent is chosen for percentages below 50% and the Media-Relative colorimetric (MR) intent when above.

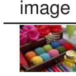
image	numeric	*keyword*	unified
	39% Per	*colors*	52% MR $\uparrow$
		shop	22% Per $\downarrow$
	54% MR	*blackwhite*	58% MR $\uparrow$
		fabric	49% Per $\downarrow$

Image & Keyword Selection

The results from our past psychophysical experiments have shown that for many images the observers and the automatic selection resulted in the same rendering intent [7]. However, images which embed several conflicting image characteristics resulted in the observers having a less significant choice of which rendering intent to choose and reduced the correlation between the automatic selection and observer results. For this experiment, we intentionally chose these more challenging images, which exhibited conflicting characteristics, since these are the images which are most likely to be impacted by the addition of the keywords. Of the 26 test images, 16 of them were used in our past experiments. 10 additional images were included, which also have conflicting characteristics.

From the 2858 keywords which appeared in the MIR Flickr database at least 500 times, we chose 26 of them for the demonstration of the framework. The number of times the words appeared within the set ranged from 551 (*shapes*) to 33339 (*sky*). We chose 14 keywords which changed the automatic selection, 7 changed the selection from media-relative to perceptual and 7 from perceptual to media-relative. The remaining 12 keywords either increased the certainty of the rendering intent choice or moved the choice closer to the opposite selection without causing a change of selection. Additionally, words were subjectively paired with each test image, which were considered to be relevant to the content of the given image.

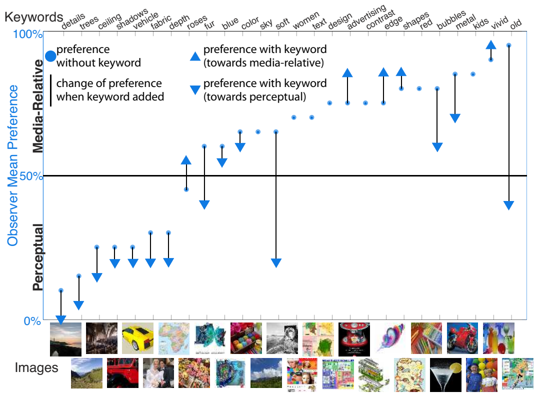


Figure 4. Plotted are the average observer preferences for each image both without and with the keyword. An arrow is plotted when the average preference changed with the inclusion of the keyword. If the arrow head points down, the preference moved towards the perceptual intent. If the tip of the arrow head is below the line the observers' favored the perceptual intent and if its above the line the observers' prefer the media-relative intent. Thumbnail examples of the images and the keywords given with each sample can also be found on the plot.

Psychophysical Experiment

To validate our automatic smart selection of rendering intent, we asked a set of observers to give their preference of over image quality between an image printed with the perceptual intent and the media-relative colorimetric intent. The reproductions for the test were printed on the Océ ColorWave 600 wide format printer with the LFM090 uncoated Océ Top Color Paper.

It was of interest to know if the preference changed when the image pair was given with the selected keyword versus without. To accomplish this, we divided the observer pool of 20 Océ employees into two sets of 10. A survey was conducted to help divide the pool of observers into two roughly equivalent sets, based on: age, occupation, nationality, country of origin, native language, fluency in any additional languages, experience assessing image quality, interest and experience in imaging and photography. The experiment was conducted entirely in English and all keywords were English.

The observers were presented with 54 pairs of documents and asked to choose which one of the pair they most preferred for overall image quality. Image quality was defined for them as: "the impression of the overall merit or excellence of an image, as perceived by you" [17]. The observers judged each pair of images twice, once without a keyword and once with a keyword. When the keyword was included, the observers were told that "The word was given to the sample by the creator. Please first read the word and then assess the pair of samples and choose which one of the pair you most prefer for overall image quality".

The pairs of samples were viewed in the Judge II viewing booth at a color temperature of 5015 Kelvins and illumination level of 1150 $\pm$ 75 lux and a distance of approximately 60 cm. The observers were allowed to move the samples around freely. The evaluation was conducted in a room without windows and the lights were turned off, the room is designed and operated for visual testing with neutral gray surround, walls and counter-tops.

Psychophysical Results

The results of the psychophysical evaluation are summarized in Figure 4. The perceptual intent ($< 50\%$) was favored for 8 of the 26 test images when a keyword was not given and 10 when the keyword was given. The average preference changed rendering intents for 4 images, 3 to perceptual and 1 to media-relative colorimetric. The test was a forced choice, so we are unable to determine if the points close to the rendering intent threshold (50%) are there because the observers' had conflicting preferences or their preferences' were not significant. This question will need to be investigated in the future work.

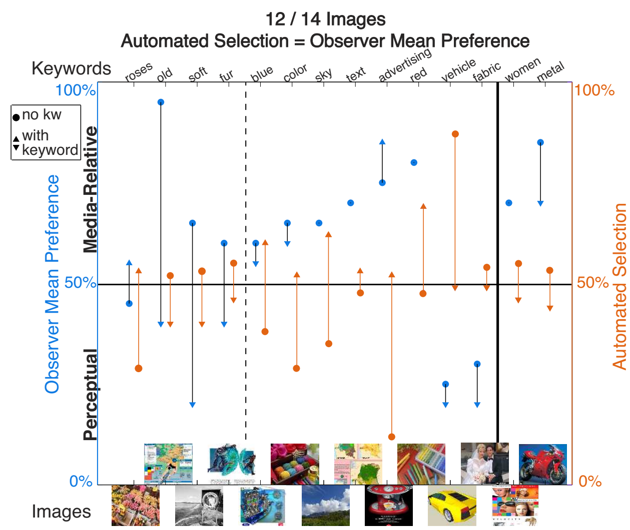


Figure 5. In this figure we focus on the 14 images where the observers or the automated selection changed which rendering intent to select with the inclusion of the keyword, a subset of the 26 images used. The observer choices are in blue and to the left of the automated selection which is in orange. The first 4 images (before the dotted line) are the instances where the rendering intent selection changed with both the observers and the automatic selection. For all 4 images, the automated selection chose the same as the observers. The next 8 images (before the solid line) are the instances where the automated selection changed to the same rendering option as the observers when the keyword was added. The furthest 2 images to the right are the only instances where the addition of the keyword negatively impacted the rendering intent selection.

Assessing the Framework Results

In total, our framework predicted the same rendering intent as the observers for 81% of the images when the keyword was included, an improvement compared to 58% without. In Figure 5 we focus on the 14 images where either the observers or the automated selection changed with the inclusion of the keywords, 12 of which were in agreement with the observers. There are 4 instances where the observers' mean choice changed rendering intents, for these images the framework selected the same rendering intent both with and without the keyword. There were only two instances where the inclusion of the keyword had a negative impact on the automated selection. In general, the keywords had a greater impact on the automated selection than with the observers' preferences. Future work may include investigating which key-

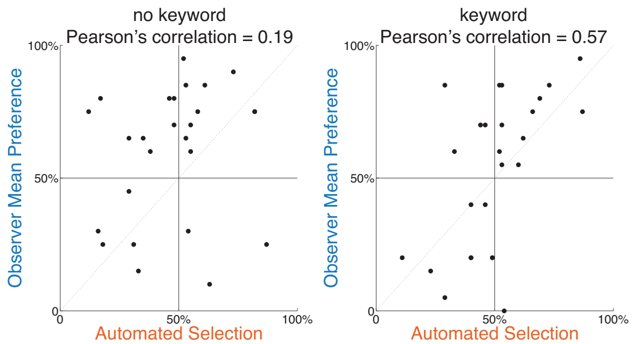


Figure 6. Plotted are the results of the automated selection versus the psychophysical (observer mean preference) results, both without the keywords and also with the keywords. The keywords, for most images, improved the results by moving the points closer to the diagonal line, which is the ideal.

word and image combinations impact the automated tool versus combinations that impact the observers' preferences.

Pearson's correlation coefficient was used to describe the relationship between the framework results and the observer results. The coefficient indicates the linear relationship between two variables on a scale of +/- 1, the closer the values are to +1 the better the correlation is. The correlation without the keywords was 0.19. When the keywords were added the correlation increased to 0.57 and the directional correlation was 0.46. Again, images which were expected to be difficult for both the observers and our framework were intentionally chosen. It is expected that images without the conflicting characteristics would result in higher correlations. We have illustrated the improvements in Figure 6, which shows the improved correlation of the automated selection when the keywords were included.

Conclusion

In this paper we introduced a unified framework that automatically selects the optimal color management settings for a given print job. The framework takes cues from both the image pixels and the semantic context. The cues from the image pixels are represented in the form of a feature vector and we use a multivariate Gaussian classification engine to compute the numeric relatedness to eight quality attributes. The semantic cues are given by a keyword associated with an image. We use a statistical framework to relate the keyword with the same eight quality attributes. Finally, we unify the relatedness values from the numeric and the semantic cues. The unified relatedness values represent the image together with its semantic context.

Our unified framework has improved the results of our automated selection. We were able to correctly predict the observers' preference for 81% of the images tested when the keyword was included, compared to 58% when the keywords were not included. The correlation of the automated selection to the observer choice improved to 0.57 from 0.19.

Future work includes investigating other ways of unifying the two methods. We will also consider other types of image features. Another area for future work is to investigate what types of keywords have the most impact on either the observer choice or the automated selection.

Author Biography

Kristyn Falkenstern is a PhD student under the direction of Hans Brettel at Télécom ParisTech, sponsored by Océ Printing. In 2009, she completed her MSc at the London College of Communication, where she studied various methods of finding spectral reflectance estimates. She worked from 2002 to 2008 in an Image Science team at the Vancouver facility of Hewlett-Packard. She also has a BS in Imaging from Rochester Institute of Technology.

References

[1] International Color Consortium. White paper 9: Common color management workflows and rendering intent usage. www.color.org, 2005.
[2] International Color Consortium. White paper 2: Perceptual rendering intent use case issues. www.color.org, 2005.
[3] P. Sun and Z. Zheng. Selecting appropriate gamut mapping algorithms based on a combination of image statistics. In *SPIE/IS&T Electronic Imaging*, volume 5667, 2005.
[4] T. Newman, T. Kohler, and J. Haikin. Dynamic gamut mapping selection, 2005. US Patent 6,947,589.
[5] A. Golan and H. Hel-or. Novel workflow for image-guided gamut mapping. *Journal of Electronic Imaging*, 17:033004–1–11, 2008.
[6] K. Falkenstern, N. Bonnier, H. Brettel, M. Pederen, and F. Viénot. Using metrics to assess the ICC perceptual rendering intent. In *SPIE/IS&T Electronic Imaging*, volume 7867, 2011.
[7] K. Falkenstern, N. Bonnier, H. Brettel, M. Felhi, and F. Viénot. Adaptively selecting a printer color workflow. In *IS&T CIC*, 2011.
[8] ISO 12640-2. *Graphic technology – Prepress digital data exchange – Part 2: XYZ/sRGB encoded standard colour image data (XYZ/SCID)*. ISO, 2004.
[9] ISO 12640-3. *Graphic technology – Prepress digital data exchange – Part 3: CIELAB standard colour image data (CIELAB/SCID)*. ISO, 2007.
[10] A. Lindner, N. Bonnier, and S. Süsstrunk. What is the color of chocolate? - extracting color values of semantic expressions. In *IS&T CGIV*, 2012.
[11] M. Huiskes and M. Lew. The MIR flickr retrieval evaluation. In *ACM MM*, 2008.
[12] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.
[13] A. Lindner, A. Shaji, N. Bonnier, and S. Süsstrunk. Joint statistical analysis of images and keywords with applications in semantic image enhancement. In *ACM MM*, 2012.
[14] C. Cui and S. Weed. Colorfulness? in search of a better metric for gamut limit colors. In *IS&T PICS*, 2000.
[15] P. Green. A smoothness metric for colour transforms. In *SPIE/IS&T Electronic Imaging*, volume 6807, 2008.
[16] H. Sheikh and A. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15:430–444, 2006.
[17] ISO 20462-1. *Photography - psychophysical experimental methods to estimate image quality - part 1: Overview of psychophysical elements*. ISO, 2004.