



Classifying images by light aesthetics

Julien Sicre

► To cite this version:

| Julien Sicre. Classifying images by light aesthetics. Machine Learning [cs.LG]. 2013. dumas-00854874

HAL Id: dumas-00854874

<https://dumas.ccsd.cnrs.fr/dumas-00854874>

Submitted on 28 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INTERNSHIP REPORT



Classifying images by light aesthetics

Author:
Julien SICRE

Supervisors:
Olivier LE MEUR
Rémi COZOT
FRVSense Team

Abstract

With the growing availability of storage space for data, such as images, comes the need for research tools that reach beyond usual criteria. Such tools include, but are not limited to, the automatic prediction of aesthetic value for indexing purpose. For this reason the field of supervised learning for aesthetic prediction has been widely active for more than a decade.

We propose a new method aimed at classifying photos based on the aesthetic of their lightning. It comes as an introduction to providing more accurate specialized supervised learning solutions for inferring photo aesthetic.

In the field of Aesthetic prediction classifier are broadly used, combining them with expert knowledge of the medium allow predictions with a score up to 70% matching the aesthetic judgement of humans. However Supervised Learning solutions have a hard time dealing with the notion of different style in the datasets. While some aspects of photography, such a spatial composition, are considered as general and always applicable, the notion of style defines specific aesthetics choice that change drastically the way the lightning (for example) is done. For this reason the classifiers performances are limited.

We choose to focus on the subject of light aesthetic in photography because we think it is relevant to be able to assess if a photo belongs to High-Key, Medium-Key or Low-key style prior to trying to infer it's aesthetic value. We show that it is possible to achieve such classification with over 80% success rate while only studying pixel populations of input images.

Keywords: Aesthetic prediction, Lighting styles, Supervised Learning

Contents

| | |
|--|-----------|
| Introduction | 1 |
| Introduction | 1 |
| Subject and problematic | 1 |
| Report structure | 1 |
| 1 Related Works | 2 |
| 1.1 Main feature classes | 2 |
| 1.1.1 Spatial layout | 2 |
| 1.1.2 Sharpness | 2 |
| 1.1.3 Lighting | 3 |
| 1.1.4 Colorimetry | 3 |
| 1.2 Feature extraction | 4 |
| 1.2.1 Saliency map | 4 |
| 1.2.2 Blur determination | 4 |
| 1.3 Illumination and Colorimetry | 5 |
| 2 Overview of the problem | 7 |
| 2.1 High Key | 7 |
| 2.2 Low Key | 9 |
| 2.3 Medium Key | 10 |
| 2.4 Goal | 11 |
| 2.5 Justifications on the approach | 11 |
| 3 Light aesthetics classification | 12 |
| 3.1 Signature building | 12 |
| 3.2 Classifying inputs | 15 |
| 4 Implementation | 15 |
| 4.1 Building the dataset | 15 |
| 4.2 Functions and methods | 17 |
| 4.2.1 GeneConfus | 17 |
| 4.2.2 GeneSign | 18 |
| 4.2.3 DiffSign | 19 |
| 5 Results and criticism | 20 |
| 5.1 Results | 20 |
| 5.2 Criticism | 22 |
| Conclusion | 23 |

Introduction

Context

This report relates the internship that I carried out during the second part of a Master's degree formation at ISTIC. It took place during a period of 5 month from February to June 2013. The master's degree is untitled "Master Recherche en Informatique" (Computer Science Research) and I attended the specialized course P5 "images and interactions". As such I applied for this internship in late October 2012.

My supervisors are Olivier Le Meur and Rémi Cozot both researchers at the IRISA research center as well as teachers at the ESIR computer engineering school. Prior to the beginning of the internship I studied the domain of the subject and wrote a bibliographic report of related papers provided by my supervisors.

Subject and problematic

The constant drop of price in data space hardware leads consumers and professionals alike to gather data in quantities that cannot be indexed and browsed with usual means. As a new way to provide automatic indexing and browsing for those large databases the field of supervised learning for aesthetic prediction is active for more than a decade.

Even though the topic of the bibliographic report I wrote is large and revolves around aesthetic classification for all photos, the subject that Rémi and Olivier proposed me is more focused. Indeed they showed me by examples that, if classifiers from the domain could achieve overall good scores, they usually miss the high aesthetic value of stylized professional photographs. More precisely, the lighting style can change drastically from one image to another while both images have a high aesthetic value. Classifiers are currently unable to take this fact into account. Examples that highlight this issue are images with uncommon lighting aesthetic. The subject of this internship is therefore to propose a new way to classify photos on their aesthetic value without ruling out highly-aesthetic light-stylized images.

Report structure

After this introduction, the part 1 of the report will present the related works in the domain. We will then define terms related to the subject of light aesthetic in photography as well as explicit our goal in the part 2. Part 3 will be dedicated to the introduction of our approach and the proposed method. Part 4 will describe our implementation of that method whereas part 5 will display some of our results and formulate criticism toward our work. We will finally conclude this report in part.

1 Related Works

1.1 Main feature classes

A vast majority of the publications in the domain consist in Supervised learning methods applied to image data. In order to be able to perform a step of learning it is necessary to decide which attributes this learning will be based upon. Such a choice is important because those attributes need to express the scattering of the classes one wants to classify as close as possible to reality. Yiwen and al. [7] explain that relevant attributes relate to classical rules of professional photography (rules meant to enhance aesthetic features). We will introduce those features regarding their relation to composition, sharpness, lighting and colorimetry of images.

1.1.1 Spatial layout

In photography, spatial composition refers to the intentional spatial layout of the elements in an image. It means choosing and placing those elements in harmony in order to communicate more efficiently ideas and intents to the viewer.

Determining the number of subjects in the picture is a logical first step. The most used method consist in computing a saliency map of the picture ([1], [5], [3], [6] and [4]). The saliency map is a tool used to locate which parts of the image attract the attention of viewers. It enables us to know if the composition carries a few clear subjects or if the image is encumbered with details.

A complimentary approach to spatial composition consist in quantifying the photographic "rule-of-third" : by aligning the subjects along the axis dividing the image in thirds of its size it is possible to guarantee a good spatial composition. This method also makes use of saliency maps but only focus on the position of clusters of interest ([4] and [7]).

1.1.2 Sharpness

The sharpness refers to the presence of blurry zones as well as the intensity of the blur. It is a major factor when assessing the aesthetic quality of a photo. By studying the blur of an image we can determine if the camera was correctly focused or if the image was shot in bad conditions.

It is known ([3] and [6]) that a consistent blur on a photo shows either a movement during the shot or a poor tweaking of the focus length. Such blur is directly linked to an inferior aesthetic quality.

Aside from homogeneous blur, authors ([1] and [6]) show that a low depth of field , a setting that aims at only focusing on the subject of the image so that everything else in the photo is blurred , is always used as a

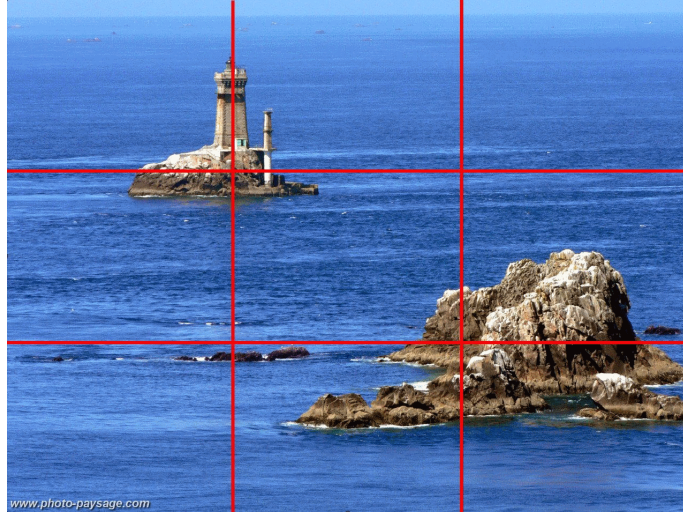


Figure 1: Example of the effect a good composition can have on a photo : following the rule of thirds allows the viewer to easily identify the subjects as well as easily roam from one to another.

mean to ensure high aesthetic results. A determination of heterogeneous blur is therefore relevant if we aim to assess the aesthetic quality of images.

1.1.3 Lighting

In the same fashion as with spatial composition a poorly lit scene impairs the aesthetic quality of a photo ([7] and [1]). Light is a key factor to the perception one has of a scene since it is the vector of it's mood.

Lighting must clearly show the subject but also enhance it's features and volumes. On that matter authors([4], [6], [3] and [7]) point that a good lighting enhance contrast between subject and background.

1.1.4 Colorimetry

Colorimetry in photography designate the arrangement of colors. Just as for spatial composition it is used to highlight the subjects and its relationship with the background using the unequal perception by human of hues and intensities along the light spectrum.

Related works ([6], [1] and [3]) show complementary colors templates can be used to balance or unbalance an image. As such it is relevant to measure the color harmony of a photo in order to assess its aesthetic quality.

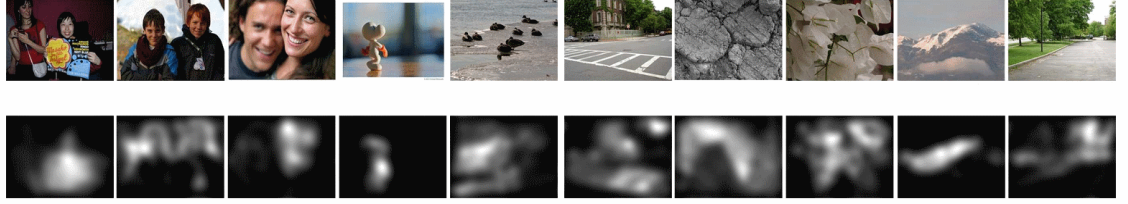


Figure 2: Saliency maps (bottom) computed from their respective source images (top).

1.2 Feature extraction

We will now introduce the computing of some of the main features used in the domain.

1.2.1 Saliency map

The saliency map is a tool that provides an information on which part of the image are more likely to draw a viewer's attention. Several factor must be taken into account : color and brightness variation, shape and the appearance of human faces. Both Xiashuai and al. and as Sagnik and al. ([5], [1]) take into account the human face factor by using facial recognition algorithms. Regardless of that, as explained by Lai-Kuan and al. ([4]), as of now no method can perform well for all types of images. The most efficient ones take party of the knowledge on the human visual system to evaluate the rank of each factor introduced previously. As a result we obtain weighted sum that can be illustrated as a 2D grey scale image. Figure 3 displays examples of saliency maps.

1.2.2 Blur determination

In order to identify the subject of a photo Yiwen et al. ([7]) suggest a method of local blurs computing. At first the image is blurred by convolving with variable size core matrices. By comparing the results on the original image it is possible to determine where and how much is blurred the original image. The blur passes will not change already blurred part as much as originally sharp zones. The accuracy of the blur intensity information is linked to the number of different convolution pass used on the original image. Figure 4 illustrate such process. In the end we produce a map of the blurred zones. This can be used to compute a center of mass for the subjects using for example surrounding volumes.

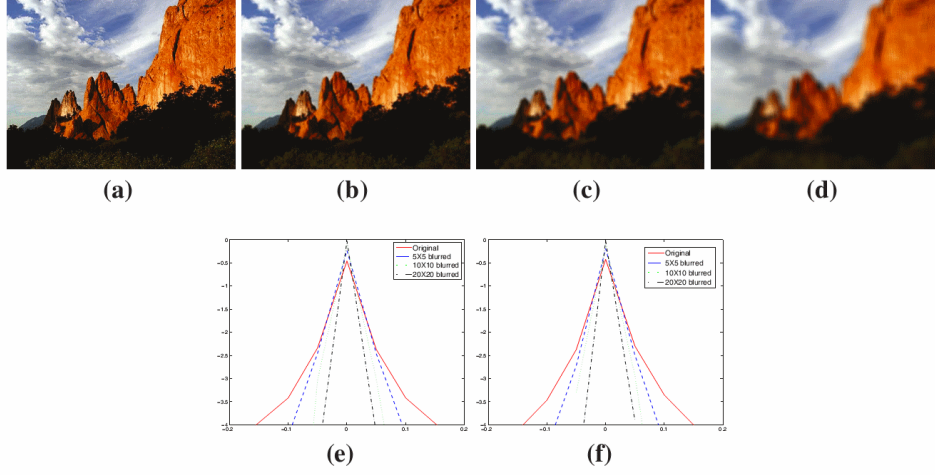


Figure 3: Image blurred with various intensity.(a) Original image.(b) 5×5 matrix.(c) 10×10 matrix. (d) 20×20 matrix. (e) Horizontal gradients histogram on a log scale for pictures (a), (b), (c) et (d). (f) Vertical gradients histogram on a log scale for pictures (a), (b), (c) et (d). Illustrations from the work of Yiwen et al. ([7]).

1.3 Illumination and Colorimetry

In order to obtain a metric on the illumination of the scene Yiwen et al. ([7]) propose, having previously identified subjects in the image, a simple comparison of mean brightness between them and the background such as :

$$f_f = |\log(B_s \setminus B_b)|$$

With B_s, B_b mean brightness of respectively the subjects and the background.

Furthermore a vast majority of authors ([3], [6], [1]) agree on using a more precise method taking into account colorimetry informations. A first step consists in changing the colorspace in order to be able to read euclidean distances as perception differences from a human point of view. The HSV (Hue Saturation brightness) colorspace is chosen for this reason. Then we compute a 3D histogram for each image, the number of beans is linked to the precision of the information obtained, however it can overburdened the classifier during the following learning task.

Although it is possible to provide histograms to the classifiers, Yan et al. ([6]) propose two different interpretations of a higher level that allow us to get close to the notion of color harmony. First we compare distances between the most populated bins of the histogram to find out if the subject and the background are in a very strong contrast. Then we count the number of different hues present in the image : the most aesthetic photos generally

hold much less hues than more common quality images. Those criterion, although simple, are rich information-wise on the aesthetic quality of the image. They are also lightweight and don't clog the classifier inputs during the learning phase.

As a conclusion we propose a table to illustrates main features in the domain (Fig5).

| Criterion | Meaning | Tools |
|---------------------|---|---------------------------------|
| Spatial composition | Distinct subjects and respect of the third rule | Saliency maps |
| Sharpness/blur | lack of homogeneous blur and heterogeneous blur highlighting subjects | Blur determination and analysis |
| Illumination | Clearly distinct subjects | Brightness evaluation |
| Colorimetry | Color Harmony and color contrast between subjects and background | HSB color space and histograms |

Figure 4: Synthesis for the role and nature of features used for classifying images regarding their aesthetic value.

2 Overview of the problem

In order to be able to successfully sort highly lightning-stylized images we will introduce 3 classes : High Key, Medium Key and Low Key lighting aesthetics. Even though those classes are soft-bordered due to the subjective part of photo appreciation, they are meant to be mutually exclusive. Those classes are defined by professional photographers and cinematographers but their definition isn't strict. We will present our understanding of those lighting aesthetic.

2.1 High Key

Images including a High Key lighting consist primarily of light tones. Although dark hues can be present, the majority of a High key image is bright. Most of High Key images are made within studios using controlled three-point-lighting however nowadays this light aesthetic uses also post process work. As such it is not uncommon to have outdoor High key images.



Figure 5: High Key image by Olivier Chauvignat, a leading artist in the domain. Dark tones and color are present even though the image is primarily bright.



Figure 6: Low Key image, despite the overall dark tones colors and bright light patch are present.

2.2 Low Key

Low Key refers to a type of lighting that aims at maximizing contrast between subject and background by using primarily dark tones. Usually Low Key images use pitch black backgrounds and only one, very controlled, light. Unlike High Key light aesthetic, it is rare to produce a Low Key image somewhere else than in a studio.



Figure 7: Medium Key image

2.3 Medium Key

Medium Key lighting refers to images in which background and subjects are evenly lit. This type of lightning is optimal to enhance colors. Although the adjective is rarely used, most photos produced by professionals and amateurs are Medium Key. Considering our approach, this class will be used as a filler class that gathers photos that are neither Low Key nor High Key. A quick research shows that most images taken and stored by professional and amateurs alike are using Medium Key lighting.

2.4 Goal

Current Classifiers from the state of the art successfully attain a rate of 70% of good predictions. However those classifiers are trained on Medium Key image set and behave poorly when fed High Key or Low Key images. Studying the related works tells us it is inevitable since all the solutions rely on one or more features extracted from lighting informations. As High Key and Low Key light aesthetics differ a lot from Medium Key illumination-wise it is only normal that classifiers fail to assess those images.

Based on those observation we think that, in order to keep steady prediction score, we should not try to build one single all-purpose classifier. If we do we might have to stop using features regarding light and illumination, thus becoming unable to make aesthetic judgement on those parameters. While this may be good to allow the classifier to take different lighting aesthetic into account during the learning phase, it will also prevent it from learning on each one of those features as well.

At this point it seems unavoidable to use different learner instances for each of our classes. For the prediction task to be automatic we then have to provide a way to route each input image to the right classifier, the one trained on the light aesthetic the image belongs to.

Building this light aesthetic classification method is our main objective.

2.5 Justifications on the approach

Even though we aim at classifying images between three classes we don't necessarily need to use a Supervised Learning method. As Low Key, Medium Key and High Key images are easily distinguished by a Human viewer we want to find a way to express with data what is obvious to the eye. Discussions with Olivier and Rémi as well as studying a few sample images showed a few guidelines.

The approach must be luminance-based since neither color composition nor spatial composition are discriminant features for High Key and Low Key images.

A global approach is desirable since local parts of certain image can be mistaken for the wrong aesthetic. For example, in Fig 1 the hair of the model are dark and a local study of that part of the image would not allow us to state the class it belongs to. This would not be a helpful information and could lead to some errors. We decide to consider each image globally.

Our goal is to be able to automatically classify images among three classes. We need a dataset of images sampling those three classes in order to test and validate our method.

3 Light aesthetics classification

As justified formerly we set up to work on image-wise lightning information. For this reason we decide to study exclusively pixel population of luminance images. We work with the luminance channel Y of the (XYZ) color space and extract an histogram out of every image considered.

Although not namely a learning algorithm, our method relies on "learning phase" including the building of a signature for each class. Once the signatures are produced we proceed to rate input images regarding the divergence of their luminance histogram to each class's signature.

The initial results were not good as shown in the Fig 8 where luminance histograms of two High Key images don't show much similarities.

therefore Olivier and Rémi proposed to use cumulated histograms. As those histograms are constantly ascending their shape reveals the location of the majority of the pixel population along the luminance axis. A quick study shows a stronger correlation between cumulated histograms of same-class images than with simple histograms, as seen in Fig 8. We also normalize all our histograms to make up for the diverse sizes of the images in our dataset.

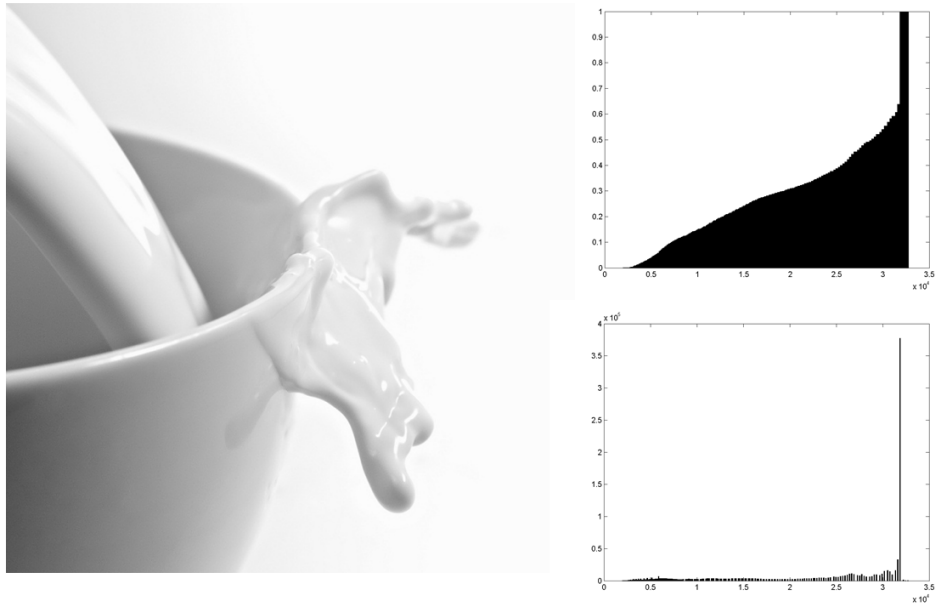
3.1 Signature building

The signatures are meant to be typical examples of a cumulated histogram for each class. We build them as histograms in which each bin is a mean of corresponding bins in a number of images belonging to one of the classes.

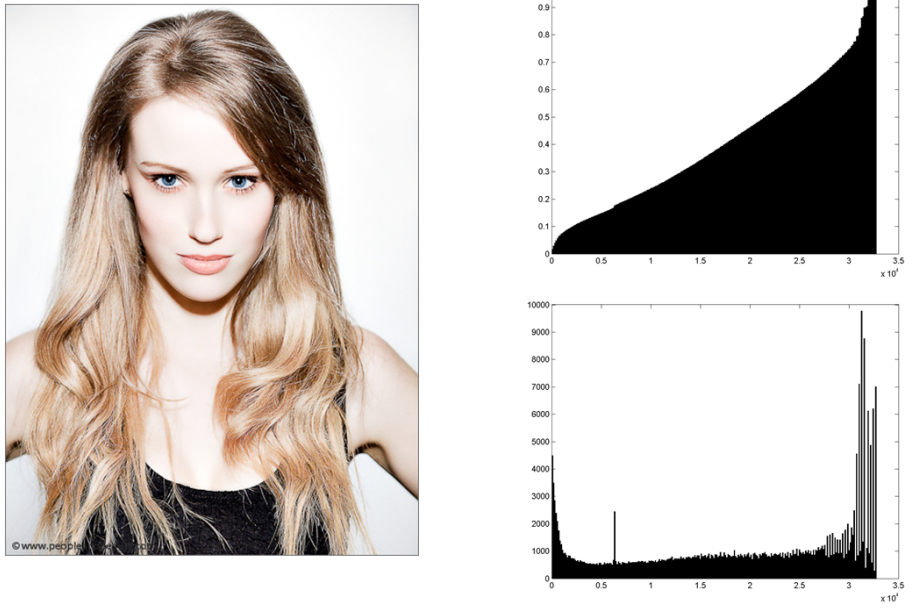
Given H_s the computed signature histogram, $nbImage$ the number of images constituting the signature and H_{img} the histogram of each images constituting the signature we compute signature as :

$$H_s = (\sum_{i=1}^{nbImage} H_i) / nbImage$$

Examples of signature for each class are shown in Fig 9. They have remarkably different shapes which shows that we successfully highlighted a way to differentiate High Key, Medium Key and Low Key lighting aesthetics.

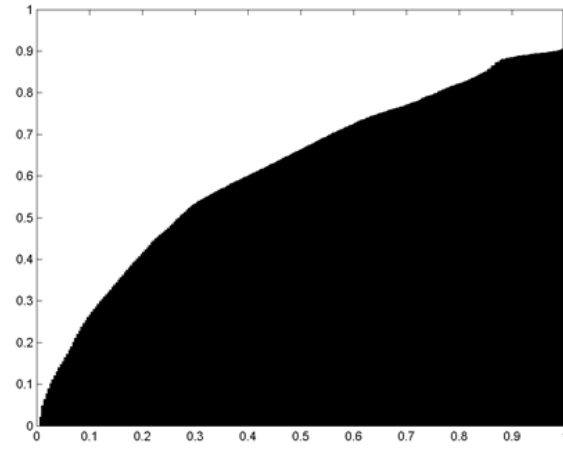


(a)

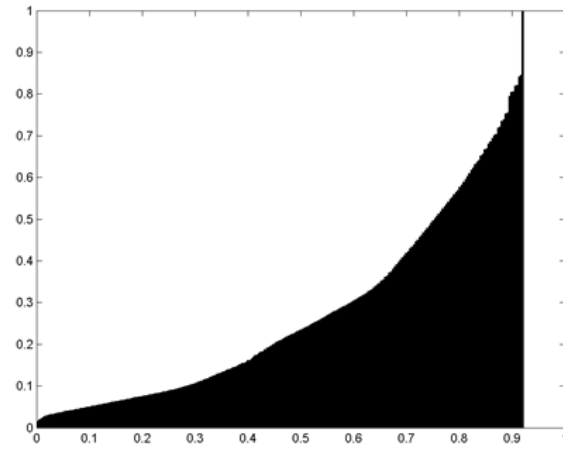


(b)

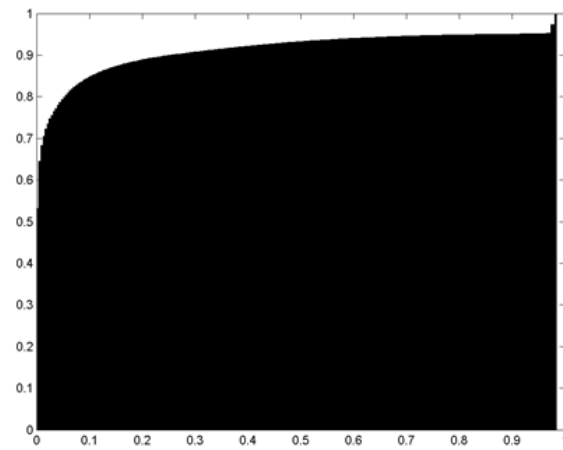
Figure 8: Examples of histograms (bottom right) and cumulated histograms (top right) for different High Key image. As we aim at regrouping them the cumulated histogram seems to hold more relevant information.



(a) Medium Key



(b) Low Key



14
(c) High Key

Figure 9: Examples of computed Signatures in the form of cumulated histograms for our three classes.

3.2 Classifying inputs

Once our signatures are ready we need to decide how will be assessed the divergence between an input cumulated histogram and each signature.

As these histograms and cumulated histograms can be considered as probability distributions, Olivier proposed that we use the Kullback-Leibler divergence. The KL divergence is a non-symmetric measure of the difference between two probability distributions P and Q . Specifically it is a measure of the information lost when Q is used to approximate P . Even though it is not a true metric we will use it to find out which signature an image is closer to.

We express our calculus as follow.

$$c* = \operatorname{argmin}_{c \in \mathcal{C}} KL(P_I, P_c)$$

With $c*$ being the most probable class for input image I which probability distribution is P_I , P_c being the probability distribution (signature) of the class C .

$$\mathcal{C} = LK, MK, HK$$
$$KL = \sum_{i=0}^{nbBin-1} P_I(i) \log(P_I(i)/P_c(i))$$

With nbBin being the number of plot in our histograms, equal to 256 since they are luminance levels in the XYZ color space. The results of our method take the form of a confusion matrix.

While this algorithm works, initial tests show that some of our specific medium key images are wrongly classified as High Key. An example of such images is Fig 10. As such images are not uncommon Rémi insisted that we find a way to rule them out.

We do so by computing two-level Gaussian pyramids of input images and then choose which signature is the closest to a majority of the input's histograms among the three (original image, first level, second level). As the KL Divergence require normalization for entry histograms, the difference in the size of the images in the pyramid will not be an issue. With this technique we use the averaging factor of Gaussian pyramids to spread the colors, of lower intensity than the white background, on the image in order to modify enough the pixel population so the image is considered Medium Key. Unitary tests show that this new version of our method successfully sorts images such as Fig 10 in the Medium Key class.

4 Implementation

4.1 Building the dataset

As stated previously our method requires a dataset in order to build signatures and evaluate performances. Images were gathered by Rémi as he is



Figure 10: Even though the image is overall bright, the highly saturated colors on the subject prevents it from being considered High Key. The early version of the method wrongly tagged it as High Key.

a photographer sensitive to the domain of lighting aesthetic in professional photography. This way we ensured that no major mistakes were made regarding the compositions of the sets. Even though most of our images can be considered as highly aesthetic, this was not our main concern. We wanted to gather as much different looking images as possible in order to create heterogeneous sets for each of our classes.

| Classes | Images count | Avg. Size (pixels) | Tot. size (MO) |
|------------|--------------|--------------------|----------------|
| High Key | 57 | 533030 | 10.3 |
| Medium Key | 321 | 733831 | 35.2 |
| Low Key | 40 | 371267 | 3.44 |

Figure 11: Dataset composition

The resulting set contains a total of 418 images distributed unevenly between the three classes. The Low Key and High key aesthetic lighting are professional styles that require equipment as well as knowledge in order to be used. For this reason it is harder to find images of those classes and the Medium Key class is a lot more crowded than the others. The Medium Key subset was created from a set of low aesthetic standard photos. We then proceeded to add images that we knew would risk to be miss-classified : Medium Key images that share some of the features of High Key or Low Key images. Constructing the subset that way would allow us to test the behaviour of the method on insightful samples, as well as check that it does not miss-classify common medium key images.

4.2 Functions and methods

Matlab was chosen as a workframe since it hosts complete image processing toolboxes as well as easy to use input and output functions. The code was split between several functions to allow easier modifications and a better intelligibility. We describe here the main functions : GeneConfus, GeneSign and diffSign.

4.2.1 GeneConfus

GeneConfus is the launcher, it contains all the calls needed to run the method. It takes the location of the database as an entry and forward a more dedicated path to each sub function. Its algorithm is as follows :

```
function [confus]=GeneConfus(inpath)

    signHK=gene_sign(strcat(inpath,'/high_key_pos/signature'));
```

```

signMK=gene_sign(strcat(inpath,'/medium_key/signature'));
signLK=gene_sign(strcat(inpath,'/low_key_pos/signature'));

lineHK=diff_sign(signHK,signMK,signLK,strcat(inpath,'/high_key_pos/test'));
lineMK=diff_sign(signHK,signMK,signLK,strcat(inpath,'/medium_key/test'));
lineLK=diff_sign(signHK,signMK,signLK,strcat(inpath,'/low_key_pos/test'));

confus=[lineHK;lineMK;lineLK];

end

```

4.2.2 GeneSign

This function builds a signature from the images present in the input subfolder. It will create any signature, it is aimed at the right directory by GeneConfus. The algorithm consists in an iteration on all the files detected to build a sum of histograms.

```

function [sign]=gene_sign(inpath)
file = dir(inpath);
NF = length(file);
sign=zeros(1,255);

for k = 3 : NF

    img = imread(fullfile(inpath, file(k).name));
    [hIMG wIMG pIMG] = size(img);
    sizeIMG=hIMG*wIMG;
    C = makecform('srgb2xyz'); % creating conversion matrix to XYZ
    xyz = applycform(img,C); % converting
    xyz = xyz2double(xyz); % conversion to double
    lum = xyz(:,:,2); % separating luminance
    lum=reshape(lum,hIMG*wIMG,1);
    %histogram
    [lumHist,lumx] = hist(double(lum),255);
    %cumulated histogram
    clumHist = cumsum(lumHist); % edit this to toggle cumulative sum
    %normalized cumulated histogram
    nclumHist = clumHist/(hIMG*wIMG); % and this too
    sign=sign()+nclumHist();

end
end

```

```
sign=sign/(NF-2);
```

Even though our KL divergence function normalizes it's entries, we apply a normalization step ourselves to be able to compare the signatures as they are generated by GeneSign.

4.2.3 DiffSign

The last function's goal is to evaluate images and generate results. It creates one line of the final confusion matrix. As for GeneSign it is aimed at the right sub-folder by GeneConfus. As it is sizeable we will only show code specific to this function.

```
function [line_confus]=diff_sign(signHK,signMK,signLK,inpath)

for k = 3 : NF

    ... computing cumulated histogram for image k ...

    % closest signature to Hist of original image
    xBaseHist = closest_sign(signHK,signMK,signLK,clumHist);

    %generating 2 level of gaussian blur image
    lum_pyr=Gscale(lum,2,[5,5],1.6);

    [lumHist1,lumx] = hist(double(lum_pyr(1).img),255);
    clumHist1 = cumsum(lumHist);
    % closest signature to Hist of first lvl blur image
    xBlur1Hist = closest_sign(signHK,signMK,signLK,clumHist1);

    [lumHist2,lumx] = hist(double(lum_pyr(2).img),255);
    clumHist2 = cumsum(lumHist);
    % closest signature to Hist of second lvl blurr image
    xBlur2Hist = closest_sign(signHK,signMK,signLK,clumHist2);

    dist_results=[xBaseHist, xBlur1Hist, xBlur2Hist];

    [X F]=mode(dist_results);

    line_confus(1,X)=line_confus(1,X)+1; % increase the winner's counter

end
end
```


As explained in the description of our method, we use the original image as well as two blurred copies and take a decision based on a majority vote. The function `ClosestSignature` is a simple code snippet that uses the `KLDiv` function implementation we chose ([2]) to ascertain which signature is the closest to the input image’s histograms.

5 Results and criticism

In this section we will present our results and develop criticism and limits of our approach.

5.1 Results

The method we described in part 3 of this report requires the building of signatures. This phase is not meant to be automatic as we feel it is possible to create signatures general enough. If our datasets contains a few hundreds of images, the subsets of High key and Low Key images are still limited in numbers. Therefore we decided to present several results based on signatures set generated from randomly-picked pictures among the subsets. To make sure signatures are balanced we also decide to craft each of them from the same number of images. The images used to created signatures are temporarily plucked from the dataset to provide a sounder result. Our validation dataset is populated as follow :

| Subsets | Images count |
|------------|-----------------|
| High Key | 47 |
| Medium Key | 311 |
| Low Key | 30 |

Our results (Fig 12) prove to be consistently higher than 80% valid classifications. We can see that a lot of errors are made with Medium Key images wrongly classified as Low Key. We also never mistake a High Key image for a Low Key or the opposite. However those good results must be put into perspective.

| class | \tilde{hk} | \tilde{mk} | \tilde{lk} |
|------------|--------------|--------------|--------------|
| HK | 44 | 3 | 0 |
| MK | 9 | 242 | 60 |
| LK | 0 | 1 | 29 |
| Error rate | 0.812 | | |

(a) Signature 1

| class | \tilde{HK} | \tilde{MK} | \tilde{LK} |
|------------|--------------|--------------|--------------|
| HK | 42 | 5 | 0 |
| MK | 7 | 266 | 38 |
| LK | 0 | 0 | 30 |
| Error rate | 0.871 | | |

(b) Signature 2

| class | \tilde{HK} | \tilde{MK} | \tilde{LK} |
|------------|--------------|--------------|--------------|
| HK | 43 | 4 | 0 |
| MK | 8 | 266 | 37 |
| LK | 0 | 1 | 29 |
| Error rate | 0.871 | | |

(c) Signature 3

Figure 12: Confusion matrices for three different sets of signatures. Rows are true classes and columns are predictions.



Figure 13: High Key, Low Key or Medium Key ?

5.2 Criticism

As stated in part 4.1, building the dataset for this precise problematic proved to be hard since professional images are less available than consumer photos. Our High Key and Low Key images subsets don't contain enough images. This introduces an unbalance in the dataset. As our pools of images are small we cannot ensure that our resulting classification is able to generalize enough for full scale use. The unbalanced dataset could introduce a bias in our results : if we managed a perfect classification of Medium Key images but only average scores for High Key and Low Key, the populations would have made it so our score would still be very high.

If the study of the state of the art made it clear that a major part of an aesthetic judgement is objective, it is not so sure for light aesthetic judgements. As shown in Fig 13, the borders between the different classes defined in part 2 always provide images for which stating a true class isn't an objective task.

Conclusion

This report relates an internship I undertook as a Master's Degree student. It took place at the IRISA research center, where my two supervisors Olivier and Rémi work, during a period of 5 months in spring 2013. The subject of this internship revolved around the field of machine learning applied to aesthetic prediction for photography.

Relying on a bibliographic report written beforehand, my supervisors proposed me to explore ways of enhancing existing learning solutions regarding some precise weaknesses. More precisely my work has been focused on addressing the case of highly lighting-stylized photography, which are currently rejected by classifiers.

The fourth part of this report describes a two-steps solution that opens new ways to approach the problem of aesthetic prediction. By distinguishing different aesthetics and describing them with a typical signature, we can automatically route images toward different lighting-aesthetic classes. We obtain promising results given the inherent subjectivity of the task.

This method allows us to reflect on two roads for future developments. First it is now possible to dedicate supervised learning aesthetic-prediction solutions to specialized datasets while still being able to automatically classify a random set of images. This should help the classifier deal with the task of aesthetic assessment since they will require less effort doing generalization. Secondly, the subject was focused on the aesthetic of lightings in a photography. As explained in the related works, aesthetic prediction relies on other features such as color harmony, spatial composition and sharpness determination. A closer study of the field of professional photography could allow us to extract other aesthetic styles that would benefit the overall efficiency of aesthetic prediction by being treated separately, using specialized classifiers.

Images used in this document were found using Google Image Search and are the properties of their respective owners.

References

- [1] Sagnik Dhar, Vincente Ordonez, and Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1657–1664. IEEE Computer Society, 2011.
- [2] David Fass. Kldiv function code at mathworks, December 2007.
- [3] Tong Hanghang, Li Mingjing, Zhang Hong-jiang, He Jingrui, and Zhang Changshui. Classification of digital photos taken by photographers or home users. In *Proceedings of Pacific Rim Conference on Multimedia*, pages 198–205. Springer, 2004.
- [4] Wong Lai-Kuan and Low Kok-Lim. Saliency-enhanced image aesthetics class prediction. In *Image Processing (ICIP), 2009 16th IEEE International Conference*, pages 997–1000, 2009.
- [5] Xiaoshuai Sun, Hongxun Yao, Rongrong Ji, and Shaohui Liu. Photo assessment based on computational visual attention model. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 541–544. ACM, 2009.
- [6] Ke Yan, Tang Xiaoou, and Jing Feng. The design of high-level features for photo quality assessment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 419–426, 2006.
- [7] Luo Yiwen and Tang Xiaoou. Photo and video quality evaluation: Focusing on the subject. In *Computer Vision ECCV 2008*.