



HAL
open science

Content-based image retrieval: Colorfulness and Depth visual perception quantification

Solene Vilfroy, Thierry Urruty, Philippe Carre, Lionel Bombrun, Arnaud Bour

► To cite this version:

Solene Vilfroy, Thierry Urruty, Philippe Carre, Lionel Bombrun, Arnaud Bour. Content-based image retrieval: Colorfulness and Depth visual perception quantification. 2021 International Conference on Content-Based Multimedia Indexing (CBMI), Jun 2021, Lille, France. pp.1-6, 10.1109/CBMI50038.2021.9461892 . hal-04263877

HAL Id: hal-04263877

<https://hal.science/hal-04263877v1>

Submitted on 23 Jan 2024

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.

Content-based image retrieval: Colorfulness and Depth visual perception quantification

Solène Vilfroy ^{*}, Thierry Urruty [†], Philippe Carré [†], Lionel Bombrun [‡] and Arnaud Bour ^{*}

^{*}Einden Company, Poitiers, France, Email: about@einden.com

[†]XLIM Laboratory, University of Poitiers, CNRS, UMR 7252, Poitiers, France, Email: first.last@univ-poitiers.fr

[‡]IMS Laboratory, University of Bordeaux, CNRS, UMR 5218, Talence, France, Email: first.last@u-bordeaux.fr

Abstract—This study focuses specifically on content-based image retrieval. The objective being to characterize the images according to their content, in the form of an index, a score which will be used for the manipulation of the images.

We propose two visual perception quantification methods, one for colorfulness and one for depth. In terms of quantifying an impression, it was necessary to set up objective and rigorous criteria reflecting a sometimes subjective concept.

Based on saturation, lightness and color disparity, we propose a framework that assigns to image an index between 0 and 100 to convey the impression that the image is very colorful or not. While in the case of depth, the estimate of a depth map will serve as a starting point. We estimate that the impression of depth is linked to the spread of the elements that make up the image from the farthest to the closest.

Index Terms—Content-based image retrieval (CBIR), Visual perception evaluation, Colorfulness impression, Depth impression.

I. INTRODUCTION

Image processing is a very broad field with many challenges that require solutions. One of them is the management of databases in order to make the best use of them.

We are particularly interested in the context of Content-based image retrieval (CBIR). Image processing involves the use of very large and diverse databases. It is then necessary to manage and process many images. Thus, indexing tools are essential, especially for the Web or for the creation of digital archives. CBIR, as we see it today, is any technology that in principle helps organizing digital picture archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. The principle is to recognize the discriminative specificities of each image. This so-called visual signature of an image can be adapted to a given context or exploitation case. The images are described as reflecting their texture, color, shape, subject, etc. Several examples of image retrieval strategy can be cited, among which [1], or even [19]. Our applicative context includes the use a faceted based search engine where the needs of expert photographers evolve rapidly ¹.

In this context, we are interested in visual impression evaluation. If many retrieval strategies translate our understanding of an image, this paper focus on the quantification of a visual impression, in this case, two: color and depth. The challenge of

our issue is the fact that interpretation of what we see is hard to characterize, and the assess of the image can be considered as a subjective task. Our contribution consists in the proposal of a method of evaluation and quantification of visual criteria such as the color or the depth visible on the image. Our approach is based on the exploitation and the combination of several tools of the literature, thus allowing the exploitation of this index to manage a database, for the search of images by similarity or even the classification of images according to these criteria.

What makes one image seem more colorful to us than another? Where does the impression of depth given by an image come from? This paper discusses numerical criteria for quantifying the degree of color and depth of an image. A large part of the work lies in the development of a process based on evaluable and objective criteria. It was also necessary to select a means of validation, in order to judge the effectiveness and robustness of the processes. In our case, we decide to sort the images according to their color or depth indexes and visually validate over a random set of images the performance of our proposal.

The paper is structured as follows. Section II focuses on the color index, while section III focuses on the depth index. Finally, Section IV concludes this paper and provides some perspectives of this work.

II. COLOR INDEX

Color is one of the most widely used features for CBIR. Indeed, many tools translating visual queries and retrieval strategies rely on color. Among them, E. Binaghi et al. [4] proposed to find images containing specified colors in assigned proportions, B. Barolo et al. [2] or J.Hafner et al. [12] focused on the search for similarity between images, C. Colomboa et al. [5] proposed a method based on the color distribution for visual representation and retrieval.

A. Colour perceived

In *Color Appearance Models* [8], Mark D. Fairchild et al. made a very complete analysis of numerical descriptions of color and its manipulation using mathematical models. S. Sural et al. [21] focused on the analysis of these different characteristics, in order to exploit them for segmentation purposes. I. Lissner et al.[16] also interested in different types of perception-based image processing problems.

Characteristic of a visual perception can be described by attributes of hue, brightness (or luminance), and saturation (or

¹www.ephoto.fr; Einden Company

chroma). It was acknowledged that each point is only defined by these three unique color attributes. A. Hanbury et al. [13] described these three elements and their interpretation in the RGB color space.

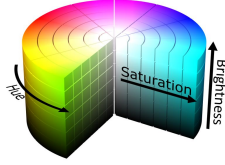


Fig. 1. **Visual perception criteria** - three attributes are defined to characterize a color : hue, saturation (chroma), and brightness.

The definition we choose to establish that an image is colored is based on the presence of many strong and vivid colors. On the other hand, we consider that a dull or dark image is perceived as less colorful. Among the numerous criteria that can be used to characterize an image, we choose to focus on those three : Saturation, Brightness and Color Diversity.

B. Saturation

Definition: *Colorfulness of an area is judged in proportion to its brightness.* [8]

Visually, a dull image appears light-colored to us, when strong and vivid colors make an image colorful. That is why the main criterion we considered is saturation.

For each pixel composed of 3 values (RGB), we can evaluate the saturation. The saturation for a pixel, is an index between 0 and 1, defined as follow:

$$S = 1 - \frac{\min(R, G, B)}{\max(R, G, B)} \quad (1)$$

The saturation value for an image is defined as the average of all pixel saturation values. The image classification according to this criterion makes it possible to highlight a weakness: the white color has a zero value saturation.



Fig. 2. **Example images with a large portion of white** - The white part lowers the overall saturation index of the image. Example: the saturation indexes for these images are 15 (left) and 25 (right).

In order to prevent the white from weighing down the index of certain images which are nevertheless colored, the white pixels are not taken into account in the averaging. By sorting the images according to the saturation index, we obtain the following ranking:

This ranking makes it possible to validate the saturation of the image as a basis of our algorithm. This is because grayscale images are given a low score, while vivid color images score among the highest.



Fig. 3. **Images ranking according to saturation index** - Saturation Index is defined as the average of each pixels saturation, except the white ones

If we consider the impression of color, we can see that some images seem misclassified. Some dark images that appear light-colored, despite a high saturation rate (some of these images are circled in red in the figure 3). This observation pushes us to take into account the criterion of brightness.

C. Brightness

Definition: *Attribute of a visual perception according to which an area appears to emit, or reflect, more or less light.* [8]

We made the observation that the dark images, despite a high saturation rate, appear to us not very colorful, which is why we have integrated a brightness criterion.



Fig. 4. **Dark image saturation** - The image above has a saturation score of 52, however, visually this image seems less colorful than those of Figure 2 for example, due to the darkness of the image.

We will therefore determine a brightness index using the relative luminance formula [13]:

$$L = 0.2126R + 0.7152G + 0.0722B \quad (2)$$



Fig. 5. **Images ranking according to brightness index**

At this stage of our algorithm, an index based on a combination of a saturation score and a luminosity score gives a coherent result, visually verifiable by a ranking as in Figures 3 and 5.

As explained in the *Saturation* part, we have chosen that the saturation criterion would be the basis of our algorithm, the brightness criterion being there only to try to correct the weaknesses induced by the use of saturation alone. In this case, integrating the brightness makes it possible to favor clear images, with white, and penalize dark images.



Fig. 6. **Ranking Example** - Ranking example for coefficients parameters : saturation 45%, brightness 25%, color diversity 30%

D. Color diversity

In a ranking based on saturation and brightness, we notice a certain type of outlier: more or less monochrome images, bright and with a not so low saturation, like images on Figure 7.



	Image 1	Image 2	Image 3
Saturation	66	24	35
Brightness	41	71	78

Fig. 7. **High score images examples** - These images have a saturation and luminosity score which ranks them above certain images which nevertheless appear more colorful.

These observations led us to take an interest in the diversity of colors present in order to penalize these images.

We use an RGB cube as a 3D histogram in order to estimate the relative presence of each color. Let N be the number of sub-cubes that make up the RGB cube, and h_n the number of pixels of the image included in the sub-cube n . The C score assigned to the image for color disparity is calculated as follows:

$$C = \frac{1}{N} \times \sum_{n < N} \frac{h_n}{\max_{n < N}(h_n)} \quad (3)$$

We choose not to count the black and the white so that the grayscale images keep the lowest scores: a black and white image will have a score of 0, and an image where the color disparity is evenly distributed, will have a score of 1. Solid background images are also at a disadvantage, As can be seen in the examples in the figure 8.



Fig. 8. **Color disparity criterion examples** - For example, these images have a score of 5, 16 and 33 respectively.

Starting from a criterion based on saturation, the observation of inconsistency in the classification led us to be interested in other criteria such as the brightness of the image or the diversity of colors present. In the end we have identified three

algebraic criteria, interpreting the visual impression as well as possible.

The final score assigned to the image will therefore be a linear combination of these index. Several tests and classification simulations enabled us to establish the following order of importance: saturation, brightness, color diversity. The coefficients assigned to brightness and color diversity being defined so that images that are too dark or too uniform are penalized, while ensuring that grayscale images remain with the lowest scores (see Figure 6). In the proposed final method, the coefficients of these three criteria are variable parameters accessible to the user via a faceted search engine.

III. DEPTH INDEX

Always with a view to defining new criteria for the characterization of an image, we are now interested in the visual impression of depth. Retrieval image strategies based on a depth characteristic are rarer. A. Jain et al[15] offered a tool for retrieval of objects using a similarity measure that combines shape and the depth information, The general idea is to rely on the geometry of the object. Regarding depth-based strategies, the state-of-the-art is limited. If there are many algorithms for estimating the depth of a scene, in particular via a depth map, this is only very rarely used as a CBIR criterion. We propose in this paper a method allowing to quantify the visual impression of depth.

A. Depth perception

As A. Saxena et al. [20] pointed out, there are many factors contributing to our depth perception such as texture variations and gradients, defocus, color, shadows, occlusion, etc. These elements greatly contribute to our understanding of an image.

Following the same process as for color, we started by asking ourselves: what are the elements that make an image give us an impression of depth? If the overall impression of depth is related to the depth of the elements in the image, is it the farthest observable object or the spread of those elements on different depth planes?

Several methods have been considered and tested. Since our image database included a large number of images with focus, we first tried to exploit the blur present in the images. However, it is difficult to rank the images only according to the blur, especially since some images are either totally clear or totally blurry. This criterion is therefore not usable, or at least insufficient. Several uses of the depth map were also considered: an estimate of the different planes that make up

the image and their spread in space or a study of the gradient of the depth map.



Fig. 9. **Depth Perception example**

If we consider the images given in Figure 9, even though, factually, the landscape of the left image is further away, the image on the right gives a greater impression of depth. These observations led us to base our algorithm on the difference between the closest and the farthest observable elements.

B. Depth map

First, we need to acquire the estimated depth for each point in the image. For this we will focus on the different algorithms allowing the estimation of a depth map from a single image. Having only one image, we must concentrate on monocular estimation methods. At the present time, the most widely used algorithms for estimating a reliable depth map from a single image are deep learning algorithms. Several neural networks have been considered; A. Kumar [6], J. Hu [14], D. Eigen [7], F. Lui [17], and C. Godard [10, 11].

The process is often similar; the network used is in most cases, an auto-encoder, with an encoder responsible for learning the characteristics of the image and a decoder responsible for reconstructing the depth map. Regarding learning, most of these algorithms use supervised learning. The database used for learning is sometimes taken from the context of autonomous car (Lidar captures) (KITTI [9]) or some interior scenes (see Figure 10) (Nyu Depth [18]).

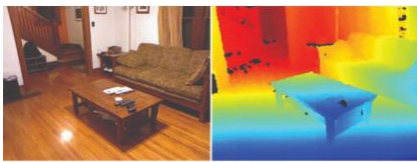


Fig. 10. **Nyu Depth Dataset** - The NYU-Depth V2 data set is composed of video sequences from a variety of indoor scenes as recorded by both the RGB and Depth cameras from the Microsoft Kinect.

After having tested several neural networks, we favored the monodepth 2 network [11], because of its efficiency, its speed of execution, and its simplicity of implementation. This algorithm is described as a self-supervised monocular depth estimation model, different from standard auto-encoder type models by the addition of tools such as self-masking loss to ignore confusing stationary pixels and a full-resolution multi-scale sampling method, give a simple and efficient model for depth estimation.

This depth estimation network is based on the general U-Net architecture, i.e. an encoder-decoder network, with

skip connections, enabling us to represent both deep abstract features as well as local information. ResNet18 is used as encoder, and the decoder is the same as [10]

So we use in our process the monodepth 2 algorithm. [11]. For this, we will use transfer learning, the network having been trained on the KITTI database. A result example is displayed in Figure 11

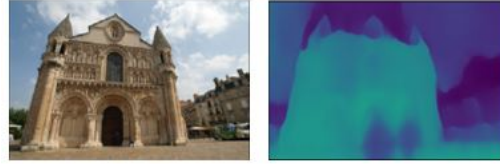


Fig. 11. **Depth Map** - Provided by the network monodepth2, model: mono_640 × 192

We start from the postulate that the impression of depth is linked to the distance between the closest element and the furthest element. Relying on the difference in depth between the closest and the furthest "pixel" seems risky, an isolated pixel can very well turn out to be an outlier. We will therefore proceed with an averaging, minimizing this risk, we therefore focused on the segmentation of the image into regions.

C. Segmentation

At this stage, the strategy of averaging by zone (grid of the image) was considered. However, such a division is not necessarily suitable for certain images with small or thin elements, which is why we are interested in segmentation. We therefore chose to insert a segmentation step, in order to identify the different elements of the scene, and to rely on the overall depth (average) of each.

We selected the watershed algorithm [3] for its efficiency and ease of use. It is a classical segmentation algorithm used for demarcating different objects in an image. Starting from user-defined markers (the connected components labeled image of boolean image: It labels background of the image with 0, then other objects are labelled with integers starting from 1), the watershed algorithm count pixels values as a topography, relief. The algorithm expands basins from the markers until basins assigned to different markers meet.

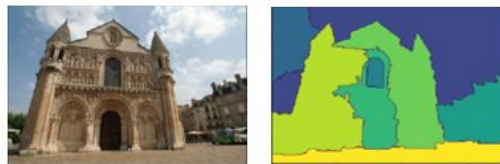


Fig. 12. **Image Segmentation** - Image segmented by watershed algorithm. The parameters have been defined so as to privilege an over-segmentation to an under-segmentation.

In our application context, a rigorous segmentation at too small a scale is not necessary and may even affect the efficiency of the method. The watershed algorithm ensures a sufficiently flexible and adaptable segmentation to allow a

rather crude detection of the elements of the image (see figure 12). The segmentation step also has the advantage of limiting the score of the “solid” images to 0, the image not being segmented or very little.

D. Depth Visual Impression Estimation

Once the depth map has been estimated and the segmentation done, we can proceed to a depth averaging by segmented regions, in order to attribute to each one an average depth. Each segment is then assigned an overall depth value, which corresponds to the average of each pixel that makes up the element. This method is used to estimate the overall depth of each element in the image.

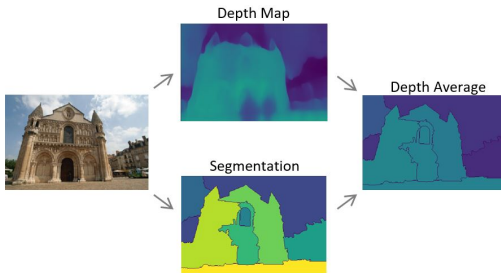


Fig. 13. **Algorithm principle** - An averaging is carried out by zone.

As defined in the part III.A, after observing the images we associated the visual impression of depth with the spreading of the elements in space and more precisely with the distance between the closest object and the farthest one. Once the depth depth averaging estimated (see Figure 13), we will assign a score to the image relative to the difference between the minimum and maximum depths.

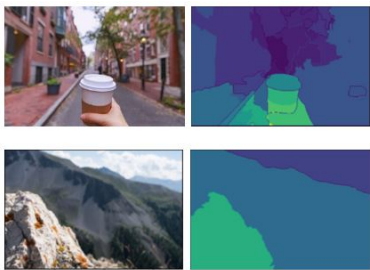


Fig. 14. **Image examples with their depth scores** - the first image has a score of 92 and the second, a score of 47.

Figure 14 examples of depth indexing using our method and Figure 16 example images sorted according to our method show its effectiveness. We can observe a classification with a certain consistency : as can be seen, the "plain" texture-type images are found at the bottom of the ranking, strongly penalized by the segmentation step, while the images illustrating an object in the foreground, in front of a distant background, obtain the highest depth scores.

The processes aiming to quantify the visual impression of depth, visual satisfaction constitute a first validation.

Let us now focus at a specific case of a series of photographs of a same scene. We will take for example the series of images represented in figure 15.



Fig. 15. A different point of view gives each of these images, a different sense of depth, and therefore, it is an appropriate sample set to evaluate our depth index. The angle of the photo gives an impression of increasing depth of the image from left to right.

The proposed process allows each of these images to be assigned the following scores, from left to right : 17.9488, 35.6251, 39.7234, 53.774. These results allow to rank the images within this series. We observe here that the scores attributed are consistent in this sense with the series of captures. It is also possible to consider the classification of such a batch of samples as a means of validation for the algorithm. Our depth index clearly procures the expert photographer large categories of image: from close up or portrait to large depth of field photographs. It also insures for a series of photographs from a similar scene a ranking order. These contributions are important to our expert partner.

IV. CONCLUSIONS

This paper proposed a process to assess a visual impression of color and depth in the idea of defining new elements used to characterize an image, in a retrieval image context.

While there are already many color-based image recovery strategies, we are offering a new one based on our perception of colors: how colorful is this image? The observation of example images as well as the comparison of these images made it possible to define the criteria of saturation, luminosity and diversity of colors as essential tools to quantify our perception of the colors of an image. The study of a ranking based on the assigned index (and therefore the comparison of several images) made it possible to highlight the weaknesses of the algorithm as it was developed, making it more and more robust. The addition of criteria and their degree of importance to each other has been defined as saturation, lightness and color diversity in that order of importance. By gauging the influence of these criteria, we have ranked and combined them in order to build a score that tends to be as universal as possible.

Regarding depth, there are very few image recovery strategies that exploit a depth characteristic and we have proposed a method to quantify the perception of depth in an image. After some reflection on the test images, the criterion that we have defined as the origin of the impression of depth is the distance between the nearest element and the farthest (usually the background). For this, a depth map, calculated by a neural network model of the state of the art, allowed us to estimate the depth of all the elements present in the image, and a segmentation followed by an averaging by depth area allowed us a reliable estimate of the overall depth of an image



Fig. 16. **Depth Ranking Example** - Examples of 10 images ranked via the index assigned by our method.

element. The assigned index will then be associated with the difference between the maximum and minimum distance. We have thus established an arithmetic method to best assess a visual impression via an index assigned to each image. Whether for color or depth, a comparison, in addition to being the reflection at the basis of the method, also constitutes validation.

In the case of visual impressions, visual satisfaction constitutes a first validation. Indeed our expert is visually satisfied, his search can use a new color and depth criteria. He may search from dull to colorful, or close up to large depth of field photographs. In the near future, the evaluation by a panel of users (experts or not), is envisaged as a validation protocol.

REFERENCES

- [1] Mutasem K Alsmadi. "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features". In: *Arabian Journal for Science and Engineering* (2020), pp. 1–14.
- [2] B Barolo, I Gagliardi, and R Schettini. "An effective strategy for querying image databases by color distribution". In: *COMPUTERS AND THE HISTORY OF ART 7* (1997), pp. 3–14.
- [3] Serge Beucher and Fernand Meyer. "The morphological approach to segmentation: the watershed transformation". In: *Mathematical morphology in image processing* 34 (1993), pp. 433–481.
- [4] Elisabetta Binaghi, Isabella Gagliardi, and Raimondo Schettini. "Image retrieval using fuzzy evaluation of color similarity". In: *International Journal of Pattern Recognition and Artificial Intelligence* 8.04 (1994), pp. 945–968.
- [5] Carlo Colombo and Alberto Del Bimbo. "Color-induced image representation and retrieval". In: *Pattern Recognition* 32.10 (1999), pp. 1685–1695.
- [6] Arun CS Kumar, Suchendra M Bhandarkar, and Mukta Prasad. "Depthnet: A recurrent neural network architecture for monocular depth prediction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, pp. 283–291.
- [7] David Eigen, Christian Puhrsch, and Rob Fergus. "Depth map prediction from a single image using a multi-scale deep network". In: *Advances in neural information processing systems*. 2014, pp. 2366–2374.
- [8] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.
- [9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 3354–3361.
- [10] Clément Godard, Oisín Mac Aodha, and Gabriel J. Brostow. "Unsupervised Monocular Depth Estimation with Left-Right Consistency". In: *CVPR*. 2017.
- [11] Clément Godard et al. "Digging into Self-Supervised Monocular Depth Prediction". In: (Oct. 2019).
- [12] James Hafner et al. "Efficient color histogram indexing for quadratic form distance functions". In: *IEEE transactions on pattern analysis and machine intelligence* 17.7 (1995), pp. 729–736.
- [13] Allan Hanbury. "The taming of the hue, saturation and brightness colour space". In: *Proceedings of the Seventh Computer Vision Winter Workshop, Bad Aussee, Austria*. Citeseer. 2002.
- [14] Junjie Hu et al. "Revisiting Single Image Depth Estimation: Toward Higher Resolution Maps With Accurate Object Boundaries". In: 2019.
- [15] Amit Jain, Ramanathan Muthuganapathy, and Karthik Ramani. "Content-based image retrieval using shape and depth from an engineering database". In: *International Symposium on Visual Computing*. Springer. 2007, pp. 255–264.
- [16] Ingmar Lissner and Philipp Urban. "Toward a unified color space for perception-based image processing". In: *IEEE Transactions on Image Processing* 21.3 (2011), pp. 1153–1168.
- [17] Fayao Liu et al. "Learning depth from single monocular images using deep convolutional neural fields". In: *IEEE transactions on pattern analysis and machine intelligence* 38.10 (2015), pp. 2024–2039.
- [18] Pushmeet Kohli, Nathan Silberman, Derek Hoiem, and Rob Fergus. "Indoor Segmentation and Support Inference from RGBD Images". In: *ECCV*. 2012.
- [19] Abdolreza Rashno and Elyas Rashno. "Content-based image retrieval system with most relevant features among wavelet and color features". In: *arXiv preprint arXiv:1902.02059* (2019).
- [20] Ashutosh Saxena, Min Sun, and Andrew Y Ng. "Make3D: Depth Perception from a Single Still Image." In: *AAAI*. Vol. 3. 2008, pp. 1571–1576.
- [21] Shamik Sural, Gang Qian, and Sakti Pramanik. "Segmentation and histogram generation using the HSV color space for image retrieval". In: *Proceedings. ICIP*. Vol. 2. IEEE. 2002, pp. II–II.