Content-Aware Dominant Color Extraction and Its Application to Mltiple-key-Color Image Retrieval

Mei Hashimoto* and Michiharu Niimi†
* Graduate School of Computer Science and Systems Engineering Kyushu Institute of Technology, Iizuka, Japan
E-mail: hashimoto.mei582@mail.kyutech.jp
† Faculty of Computer Science and Systems Engineering Kyushu Institute of Technology, Iizuka, Japan
E-mail: niimi@ai.kyutech.ac.jp

***Abstract*—This paper proposes a novel dominant color extraction method that incorporates salient region analysis to better reflect human visual impressions. Traditional dominant color extraction methods treat the entire image uniformly, often resulting in dominant colors that do not correspond to visually important regions. In this method, at first stage, images are classified into "landscape images" and "non-landscape images," and then perform appropriate processing for each. For landscape images, dominant colors are extracted using conventional clustering over the entire image. For non-landscape images, we use saliency maps to detect regions of interest, and extract the most frequent color within these regions as the dominant color. Additionally, in order to evaluate the quality of dominant color, we construct color feature vectors based on predefined reference colors, enabling image retrieval tasks by multiple key-colors. Experimental results demonstrate that our method provides dominant colors that better match human perception, and improves the accuracy of image retrieval using color-based emotional representation. This approach shows promise for applications in affective computing, multimedia retrieval, and content recommendation.**

I. Introduction

In the field of modern image processing, quantitatively understanding the impression and characteristics of images is an important challenge. Among various elements in an image, "color" is one of the most influential factors affecting its visual impression and atmosphere, and is widely used in applications such as image retrieval, classification, and recommendation.

Dominant Color (DC) is a widely used technique that captures visual impressions by representing a few visually impactful colors in an image. Traditional DC extraction methods typically apply clustering or histogram analysis over the entire image to identify the most frequent colors[1][2]. However, these methods assume that all regions of the image contribute equally, and do not consider the human visual tendency to focus on certain areas more than others. There is a researche[3] that pays attention to a specisic objects in an image. However this method also uses entire image to extract dominat colors.

In our previous research, we developed an image display system synchronized with music lyrics based on the semantic similarity between lyrics and images[4]. We attempted to extract color features from images using their most frequent colors as a shared concept between lyrics and images. However, dominant colors extracted from the background failed to reflect the actual objects that humans pay attention to, such as strawberries or fireworks. This also revealed a limitation in representing visual impressions using only a single dominant color when multiple subjects are present in an image.

These issues stem from the fact that conventional DC extraction methods process all regions uniformly and fail to consider human gaze patterns and attention biases. This discrepancy leads to a gap between the extracted image features and the human impression.

In this study, we propose a novel dominant color extraction method that incorporates human visual attention during image viewing. We first classify natural images into two types: "landscape" and "non-landscape," and apply different DC extraction strategies depending on the classification. For landscape images, we apply conventional DC extraction across the entire image. For non-landscape images, we use saliency maps to detect salient regions and extract the most frequent color within these regions as the dominant color. This enables us to emphasize the colors of visually important objects.

Therefore, our method can extract dominant colors that are more aligned with human impression, which is expected to improve performance in image retrieval and impression-based classification tasks.

This study has a clear application: image retrieval based on multiple key-colors extracted from song lyrics[4]. Specifically, multiple keywords are extracted from lyrics and mapped to a color space using the Color Image Scale[5]. Following these steps, multiple key-colors that adequately represent the impression of the lyrics are generated. One of the goal of the our previous study is to find the best match image to the multiple key-colors. In order to realize it, we generate color feature vectors for both multiple key-colors and the dominant colors extracted from images. Image retrieval is then performed by comparing the distance between these vectors.

The rest of this paper is organized as follows. Section II summarizes related work and clarifies the position of this research. Section III describes the proposed method for dominant color extraction. Section IV explains how color feature vectors are generated. Section V presents the experimental setup and results. Finally, we conclude the paper and discuss future directions.

979-8-3315-7206-8/25/$31.00

II. Related Work

A. Research on Dominant Color Extraction

Techniques for extracting dominant colors, which represent the characteristic colors in an image, have been widely used as a fundamental technology for understanding the impression and semantics of images. Applications include image summarization, compression, retrieval, and classification. A representative approach involves extracting a small number of dominant colors from the overall color distribution of an image using clustering or histogram analysis in RGB color space.

Zhang et al.[1] proposed a method for extracting dominant colors from natural images that incorporates not only color frequency but also visual elements such as hue, saturation, and brightness. Their method considers the characteristics of natural images and improves the accuracy of extracting visually impressive colors by carefully analyzing the distribution of foreground and background colors in color space. Chang et al.[2] also proposed a technique to extract image composition colors based on color features, further refining the process of DC extraction.

These studies contribute to improving the accuracy of DC extraction over the entire image, but they do not fully consider visual perception aspects such as local saliency and attention to the main object. Thus, the extracted dominant color may not match the impression perceived by humans.

Although some studies have attempted to combine DC extraction with techniques such as edge detection or region segmentation[6], it remains difficult to appropriately extract meaningful regions depending on the image content. There is a research to extract DC using salient objects from an image[3]. However we have confirmed through preliminary experiments that this method fails when a specific object doesn't appear the image, and this method also performs semantic segmentation on the entire image and extracts dominant colors based on the segmented regions, rather than extracting only the colors from the region of interest. Therefore, it does not achieve dominant color extraction specifically focused on the region of interest.

B. Saliency Maps and Estimation of Attention Regions

The human visual system does not treat all areas of an image equally, but tends to focus on regions that are visually "salient," which are areas with distinctive features such as color, brightness, shape, or motion. Saliency maps are computational models designed to mimic this behavior by predicting the areas in an image that are likely to attract human gaze.

Itti et al.[7] proposed an algorithm that models visual saliency using low-level features such as color, intensity, and orientation to generate a saliency map that approximates human eye movement. This work had a significant impact on subsequent research, and various extensions have since been developed. In recent years, deep learning-based saliency estimation methods using network architectures such as U-Net and VGG have also been proposed[8].

C. Contribution of this study

In this study, we propose a dominant color extraction method that focuses on the gaze area within an image to more accurately capture the visual impression of the image. When extracting the gaze area using a saliency map, the gaze area is extracted for all images in order to extract the gaze area for each image. For example, in a photo with a farm in the background and a hand holding strawberries, good results were obtained with the strawberries and the hand holding the strawberries as the gaze area. On the other hand, in an image of a riverbank with trees, grass, stones, and a river, only the trees were obtained as the gaze area, resulting in undesirable results where the color information of the grass, stones, and river was lost.

Therefore, while extracting the gaze area is effective for images with a single object, such as a hand holding strawberries, it is considered ineffective for images with multiple objects or landscape images where objects are spread across the entire image.

Therefore, if landscape images and non-landscape images can be distinguished, it is possible to extract DC that better represents the characteristics of the image by dividing the processing into "extracting color information from important areas by extracting the area of interest in non-landscape images" and "extracting color information from the entire image in landscape images." Specifically, while the conventional method processed the entire image uniformly, in this method, the image is first classified into landscape images and non-landscape images, and processing is performed according to the classification results.

For landscape images, we apply conventional global clustering, and for non-landscape images, we apply gaze region extraction using saliency maps to extract the most frequent color locally as the representative color. By flexibly switching the DC extraction strategy based on both image type and gaze region, we can obtain representative colors closer to human visual perception, which is expected to improve the accuracy of image search and classification.

III. Dominant Color Extraction Considering Salient Regions

To effectively extract visually significant colors, we classify images into landscape images and non-landscape images, and apply different extraction methods based on this classification.

A. Image Classification

First, we describe the method for distinguishing between landscape and non-landscape images. Based on the method proposed in [9], we perform the classification as follows:

1) **Color Space Conversion**: The target image is assumed to be a full-color RGB image. We convert it into the Lab and HSV color spaces to obtain additional feature representations.
2) **Superpixel Segmentation**: We perform superpixel segmentation (SPS) in the Lab color space to divide the

image into meaningful small local regions. Unlike [9], which used fixed blocks, our method leverages SPS to achieve more flexible region extraction that conforms to image structure.

3) **Attention Region Detection**: Based on the features calculated for each region, we evaluate the importance of each region using fuzzy logic. Fuzzy membership functions determine the degree to which each region is "important" based on features such as high saturation, strong contrast, and large region size. Regions that meet these visual criteria are given high fuzzy scores. Using these scores, we determine whether a region qualifies as a Region of Interest (ROI) as in [9].
4) **Image Classification**: If an ROI is detected in the image, it is classified as a non-landscape image; otherwise, it is considered a landscape image.

B. Dominant Colors in Landscape Images

Following the approach of Zhang et al., we extract dominant colors (DCs) from landscape images. Hereafter, the dominant colors are denoted as $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_D$, where $\boldsymbol{d}_i = (d_{i1}, d_{i2}, d_{i3})$.

While Zhang et al. extract DCs based on 11 predefined basic colors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_{11}$, where $\boldsymbol{b}_i = (b_{i1}, b_{i2}, b_{i3})$, their method does not sufficiently represent the desired image information in our case. Therefore, we modify the method to allow a wider range of colors to be selected.

Let us consider a full-color image of size $w \times h$. Pixels are denoted by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_L}$, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3})$ and $N_L = w \times h$. These pixel values are represented in a 3D color space (x_1, x_2, x_3).

We apply clustering in color space with K clusters. Each cluster center is represented as $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_K$, where $\boldsymbol{c}_i = (c_{i1}, c_{i2}, c_{i3})$.

Each dominant color $\boldsymbol{d}_i$ is determined by the following rule:

$$\boldsymbol{d}_i = \begin{cases} \boldsymbol{b}_m & \text{if } \|\boldsymbol{b}_m - \boldsymbol{c}_i\| < T_1, \\ \boldsymbol{c}_i & \text{if } \|\boldsymbol{b}_m - \boldsymbol{c}_i\| \geq T_1 \end{cases} \tag{1}$$

where

$$m = \arg\min_j \|\boldsymbol{b}_j - \boldsymbol{c}_i\| \tag{2}$$

T_1 is a predefined threshold. This approach assigns the cluster center to the nearest basic color if it is within a threshold distance, or keeps the cluster center as is otherwise. The number of pixels assigned to each dominant color is denoted as $n(\boldsymbol{d}_i)$.

C. Dominant Colors in Non-Landscape Images

For images classified as non-landscape, an ROI is extracted using a saliency map, separate from the classification stage. The saliency map used in this study is based on Global Contrast based Salient Region Detection by Achanta et al.[10][1].

The computed saliency map is converted to a grayscale image with 256 levels (0 to 255). A bounding box enclosing the top 20 pixels with the highest saliency values, which we found to work well through some preliminary experiments, is used as the ROI. Let N_R denote the number of pixels in the ROI. Then, clustering is applied to the ROI in color space to extract the most frequent color as the dominant color. Since only the most frequent color is used, $\boldsymbol{d}_1$ is set with $D = 1$. The number of pixels with this dominant color in the ROI is denoted as $n(\boldsymbol{d}_1)$.

IV. Color Feature Vectors Based on Dominant Colors

Our goal is to search for the most suitable image when multiple key-colors are extracted from lyrics. Here, multiple key-colors are colors set based on the Color Image Scale for multiple keywords obtained from lyrics. By using the Color Image Scale, the impression of lyrics can be expressed in multiple colors. Therefore, to consider the correlation between lyrics and images, which are different media, it is necessary to represent both as feature vectors in the same color space (shared conceptual space: color space) and calculate their similarity. In this study, the colors extracted from the color image scale, which is called "reference colors", serve as the search criteria in this color space. In order to achieve the search we aim for, we introduce color feature vectors. Color feature vectors utilize the proportion of pixels belonging to the dominant color in the target image, based on the proximity between the dominant color and the reference color, to express how much the reference color is contained in the dominant color. The specific settings are as:

These are denoted as $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_R$, where $\boldsymbol{r}_i = (r_{i1}, r_{i2}, r_{i3})$. Based on these, a color feature vector $\boldsymbol{f} = (f_1, f_2, \ldots, f_R)$ is created. Each component f_i represents the weight associated with reference color $\boldsymbol{r}_i$.

A. From Landscape Image DCs to Feature Vectors

We initialize the feature vector $\boldsymbol{f}$ with zeros. For each dominant color $\boldsymbol{d}_i$, we compute

$$min_i = \arg\min_j \|\boldsymbol{d}_i - \boldsymbol{r}_j\| \tag{3}$$

This means that we have found the $\boldsymbol{r}_i$ closest to $\boldsymbol{d}_i$, and its index is min_i. Therefore, for the element f_{min_i} of $\boldsymbol{f}$,

$$f_{min_i} = f_{min_i} + n(\boldsymbol{d}_i)/N_L \tag{4}$$

Set the weight for the most recently referenced color of $\boldsymbol{d}_i$. If there are multiple $\boldsymbol{d}_i$, values are set for multiple elements of $\boldsymbol{f}$.

In other words, $\boldsymbol{f}$ expresses how many reference colors are contained in the entire landscape image being considered, and how large each reference color is.

B. From Non-Landscape Image DCs to Feature Vectors

Again, we initialize $\boldsymbol{f}$ to zero. Assuming the most frequent color in the ROI represents the object of interest, the occupancy rate o is defined as:

$$o = n(\boldsymbol{d}_1)/N_R \tag{5}$$

[1]Implementation based on PAIR code on GitHub[11]

We compute distances between $\boldsymbol{d}_1$ and each reference color $\boldsymbol{r}_i$, sorting them in ascending order. The indices are $j_1, j_2, \ldots, j_R$ and the distances are $k_1, k_2, \ldots, k_R$.

For indices satisfying $(k_m/k_1) < T_2$, we update the feature vector as:

$$f_m = \frac{(1/k_m)}{V} \times o \tag{6}$$

where $V = \sum_m (1/k_m)$. This distributes the occupancy rate o to reference colors based on their proximity to $\boldsymbol{d}_1$. T_2 is a predefined threshold.

There may be multiple m values, in which case multiple elements of $\boldsymbol{f}$ will have values. In other words, $\boldsymbol{f}$ expresses which reference color the most frequent color extracted from the ROI of the non-landscape image is closest to, and how similar each reference color is to the most frequent color.

C. Image Retrieval Based on Color Feature Vectors

The method to compute the color feature vector for each image has been described. In our system, we compare color features from images and textual inputs. Keywords are extracted from the lyrics text and mapped to color values using the color image scale, which functions like a dictionary with a defined correspondence between words and colors. Each such color is assumed to correspond to a reference color.

For each reference color, we set the corresponding $f_i = 1$ to construct a one-hot vector $\boldsymbol{f}$. Multiple keywords yield vectors $\boldsymbol{q}_i$, and their sum produces the final query vector $\boldsymbol{q} \in [0, 1]$. For example, three keywords result in a $\boldsymbol{q}$ with three elements set to 1.

Both $\boldsymbol{f}i$ and $\boldsymbol{q}$ are R-dimensional vector. To find the best-matching image i is equal to find the largest dot product.

V. Experiment

A. Experimental Method

This experiment utilized 666 natural images taken by laboratory students. The dataset includes a diverse range of categories such as landscapes, people, animals, buildings, and food. All images were resized to 640×480 pixels. to ensure uniform processing time and consistent conditions. Using the proposed method, DC extraction was performed on these images, followed by image-lyrics matching based on the extracted DCs.

B. Results

First, we describe a preliminary experiment for classifying images into landscape and non-landscape categories. Two images were selected as landscape examples and two as non-landscape examples For each of the four images, a histogram of fuzzy scores was generated. Figure1 shows these histograms, where the x-axis represents the fuzzy scores and the y-axis indicates the frequency (number of regions), the symbols I1 and I3 represent landscape images, and the symbols I2 and I4 represent non-landscape images.

The results indicate that non-landscape images contain a significant number of local regions with fuzzy scores above 3.8. Based on this observation, we adopted a threshold of 3.8 for distinguishing between landscape and non-landscape images in this experiment.

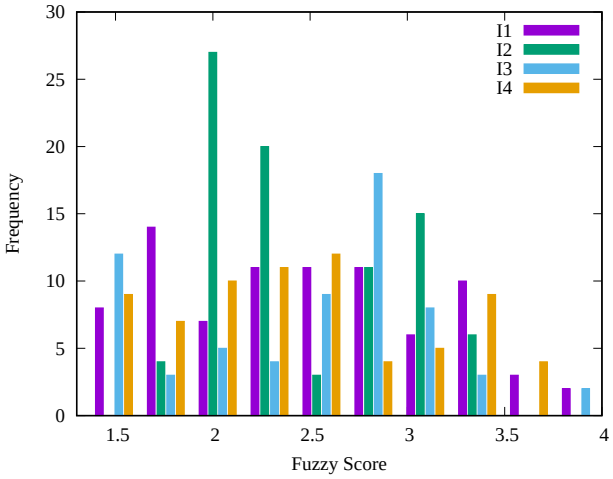


Fig. 1. Histogram of fuzzy score distribution

Based on the above criteria, we classified each image and extracted its dominant colors. For landscape images, the number of cluster was set to $K = 25$ following [1] and the threshold for extracting dominant colors was set to $T_1 = 20$, which was determined through several experiments with ranges varying from 10 to 50. In the proposed method, dominant colors are extracted by clustering in the color space for both landscape and non-landscape images. Through preliminary experiments, we subjectively found that the Lab color space was more suitable for landscape images, while the HSV color space produced more appropriate results for non-landscape images. Therefore, in this experiment, we present the results obtained in those respective color spaces. An example of the results is summarized in Table I.

In the input images for non-landscape classification, red rectangles indicate the regions identified as ROI. Multiple dominant colors were extracted for landscape images, while a single dominant color was obtained for non-landscape images. Notably, the dominant color of non-landscape images corresponds to the most frequent color within the ROI, which also aligns with subjective evaluations.

Next, we conducted an image retrieval experiment using color feature vectors. The number of reference colors was set to $R = 27$, and the threshold used in generating feature vectors for non-landscape images was set to $T_2 = 10$. Specific keywords, their corresponding colors (key-colors), and the retrieval results are summarized in Table II. The retrieved images shown are the top three results of the inner product. The searching results using only conventional method also are displayed in it.

C. Discussion

In the proposed method, image classification is based on fuzzy logic scoring of features extracted from segmented local regions, and categorization into landscape or non-landscape is

TABLE I
DOMINANT COLORS IN LANDSCAPE/NON-LANDSCAPE IMAGES

Image	Decision	Dominant colors
	non-landscape	
	landscape	
	non-landscape	
	landscape	
	non-landscape	
	landscape	
	non-landscape	
	non-landscape	

determined by the presence or absence of ROI.

Subjectively, images that were expected to be classified as non-landscape were indeed categorized correctly. Notably, an image expected to be a landscape image due to the presence of flower beds was classified as non-landscape because an ROI was detected on the flowers, demonstrating the effectiveness of our approach. However, in cases where objects are naturally embedded within the landscape or spread across the entire image, ROI detection becomes difficult, leading to ambiguous classifications.

For landscape images, 25 colors were extracted by clustering in the Lab color space. By mapping the clustered colors to predefined basic colors based on distance, visually redundant colors were eliminated, enabling accurate representation of the overall image impression. For non-landscape images, salient regions were extracted using saliency maps, and clustering was applied only within those regions. This allowed dominant color extraction to focus on object regions, minimizing background influence.

Our method does not rely on human annotations. Instead, it integrates colors automatically by evaluating the distance between cluster centers and 11 predefined basic colors, achieving automatic and flexible color integration. This is a notable feature of our approach and contributes to better reproducibility and general applicability.

To examine the effectiveness of dominant color extraction considering salient regions, we introduced color feature vectors and conducted image retrieval using multiple key-colors. Key-colors were set based on the color image scale, each associated with a corresponding keyword. From the experimental results shown in TableII, the retrieved images appear to match the meaning of the lyrics used to derive the dominant colors.

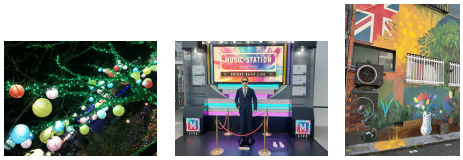
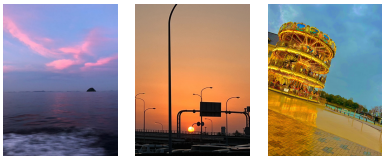
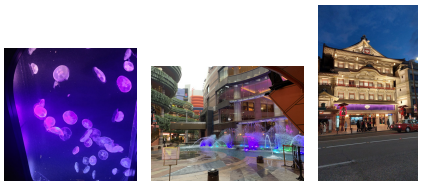
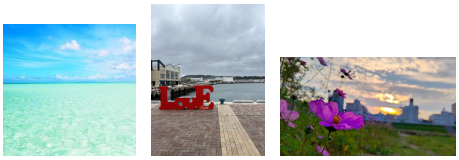
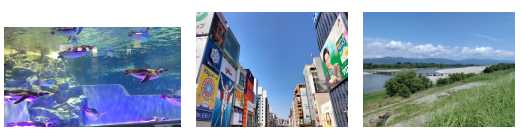
The results using salient region detection are the followings. For “Smile, Cute, You,” images containing all three key-colors were obtained. For “Blue, Sky, Spread, Sea,” images of blue skies and seas were selected that not only matched the key colors but also covered all the keywords contained in the lyrics. In the case of “Tears, Wipe, Strong,” the selected image of large fireworks was one I took at my first school festival, and it reminded me of the memories of my hard work and the moments that moved me at that time. Not only does it match the keyword in terms of the memory of wiping away tears, but it also appears to be effective as one of the new media viewing methods we are aiming for. Therefore, it can be said that images reflecting the meaning of the lyrics and forming the dominant color were selected.

When no area of interest was set, images containing many colors (landscape images) tended to be selected. On the other hand, in the image of a lotus flower with the keywords “Flower, Bloom, Future,” even though the area of the flower color in the image as a whole was small, by setting an important area (area of interest) in the image, the area of the flower color within that area became larger. “Blue, Sky, Spread, Sea” and “Flower, Bloom, Future” have similar key-color compositions. They share the same purple and different shades of blue, with the main difference being the presence or absence of light pink. Therefore, when no focus area is set, similar images are selected. However, by setting a focus area to represent the characteristics of the image rather than the entire image, it is possible to select images that match the subtle differences.

VI. CONCLUSION

In this study, we proposed a novel method for extracting visually impressive representative colors (dominant colors) from natural images. By classifying input images into “landscape” and “non-landscape” images and applying different processing accordingly, our method incorporates salient region analysis into dominant color extraction. This enables the extraction of

TABLE II
COMPARISON OF SEARCH RESULTS WITH AND WITHOUT SETTING A FOCUS AREA

Keywords	Key-Colors	Without Salient Region Detection	With Salient Region Detection
Smile, Cute, You			
blue, sky, spread, sea			
flower, bloom, future			
Tears, Wipe, Strong			

representative colors that emphasize the characteristic colors of objects within the image.

In particular, for non-landscape images, the introduction of salient region analysis enhanced the semantic validity of the extracted representative colors, demonstrating potential for applications in image retrieval and impression summarization. Experimental results confirmed the effectiveness of the classification and color extraction processes, indicating the utility of our method as a concise way to summarize the visual impression of images.

Future work will focus on improving the accuracy of saliency maps, optimizing color integration and the number of clusters, and conducting further evaluations based on subjective assessments to enhance the generality and precision of the proposed method.

REFERENCES

[1] Y. Chang, T. Iida, and N. Mukai, "Dominant color extraction method from natural images," *The Journal of the Institute of Image Electronics Engineers of Japan*, vol. 44, no. 4, pp. 637–643, 2015.

[2] Y. Chang and N. Mukai, "Color feature based dominant color extraction," *IEEE Access*, vol. 10, pp. 93 055–93 061, 2022.

[3] A. B. Gunduz, B. Taskin, A. G. Yavuz, and M. E. Karsligil, "A better way of extracting dominant colors using salient objects with semantic segmentation," *Engineering Applications of Artificial Intelligence*, vol. 100, p. 104 204, 2021.

[4] M. Hashimoto and M. Niimi, "Generation of photo slideshow with song based on closeness between concept of lyrics and that of images," in *APSIPA ASC 2024*, 2024, pp. 1–6.

[5] S. Kobayashi, *Color image scale*. 1991.

[6] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[8] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1531–1544, 2019.

[9] T. Yahara and J. Maeda, "Automatic detection of perceptually lmportant regions in a color image," *The Institute of Image Information and Television Engineers*, vol. 27.9, pp. 71–75, 2003.

[10] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[11] *PAIR-code/saliency*, https://github.com/PAIR-code/saliency.