Contextual Dominant Color Name Extraction for Web Image Search

Peng Wang[1], Dongqing Zhang[2], Gang Zeng[1] and Jingdong Wang[3]
[1]*Key Laboratory on Machine Perception, Peking University, Beijing, China*
[2]*Embedded and Pervasive Computing Center, Shanghai Jiao Tong University, Shanghai, China*
[3]*Microsoft Research Asia, Beijing, China*

***Abstract*— This paper addresses the problem of extracting perceptually dominant color names (DCN) of images. Our approach is motivated by the principle that the pixels corresponding to one dominant color name identified by human are often context dependent, spatially connected and form a perceptually meaningful region. Our algorithm first learns the probabilistic mapping from a RGB color to a color name. Then, a double-threshold approach is utilized to determine the color name of a RGB pixel in a specific image by considering its neighboring pixels. This scheme effectively deals with the pixels ambiguously belonging to several dominant color names. Last, the saliency information is combined to extract perceptually dominant colors. Experiments on our labeled image data set and the Ebay image set demonstrate the effectiveness of our approach.**

***Keywords*-Color Naming, Image Dominant Color Descriptor, Contextual Propagation.**

I. Introduction

Dominant color name (DCN) proposed in this work includes two concepts: dominant color descriptor and color name. Dominant color descriptor [1], [2] is a set of main colors perceived in an image. It is an efficient descriptor which has been widely applied to many applications such as content-based image retrieval [3], [4], [5]. Color name [6] is the linguistic label to a color vector given by human beings other than general representations such as RGB, HSV, Lab color space. It is firstly studied in visual psychology [7] and later induced in multimedia by Conway [8]. In sum, the DCN of an image is to extract linguistic color names that dominate the image color.

In this paper, we propose to extract perceptually dominant color names based on contextual information (Fig. 1(b)) which provides both high level and efficient descriptions of an image. Besides its usage on the retrieval tasks, such a scheme could be used for web image color filter to exclude the images that do not contain the user's perceptually dominant color, as illustrated in Fig. 1(a).

The contribution of this work includes three aspects: 1) A color naming scheme is proposed. This step is used to estimate the probability distribution of the color belonging to the given color names; 2) A context-based scheme for extracting dominant color regions with double thresholds is proposed. A strict threshold is first used to identify the regions with high confidence belonging to some color names, leaving the remaining as ambiguous regions. A loose threshold is then used for propagating the confident regions to ambiguous regions. 3) Furthermore, the saliency maps of the image are exploited to localize the salient object for extracting perceptually meaningful dominant colors.

This work was done when Peng Wang and Dongqing Zhang were interns at Microsoft Research Asia.

(a) An application of dominant color naming

(b) A sample image

Figure 1. (a) A real application as search engine's color filter, images including the query object dominated by a demanding color name (a color block is clicked on left side) are returned. (b) From a sample image, we notice that the color naming of pixels in an image cannot be solved locally, but requires the interpretation of contextual information. In the image containing a car (middle), we recognize the dominant color in the rectangles as "red", while in isolation, the color name may be confusing with "pink", "brown" or "orange".

A. Related Work

In recent years, many works studied on the color naming [9], [10], which construct a mapping from color space to color name space. The most recent work learns the mapping probability through PLSA [11]. However, they identify the color name of each pixel by its maximum posterior probability without considering the context information.

For extracting the dominant color of an image, to perform k-means of pixels in color space is efficient, however, such a scheme brings noisy regions and is lack of considering the spatial connectivity of pixels on the image, i.e. the color regions on image, which is important for color name identification. One recent work [12] exploits region growing to extract the dominant homogeneous color regions. Nevertheless, directly growing on the color image may make

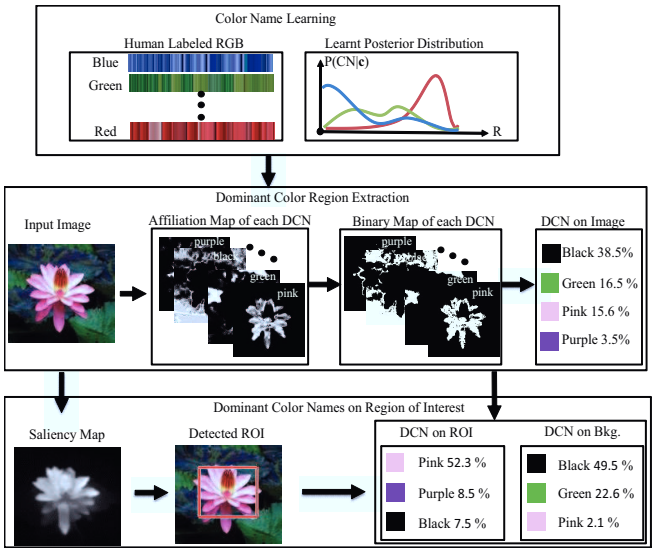


Figure 2. The overview of our dominant color name extraction system.

one dominant color regions encircle neighbor pixels in other color names due to the inconsistency between the distance in color space and human vision on color names. Our propagation is conducted on the affiliation maps in which the value of a pixel describes the membership degree of the pixel's color belonging to each color name (detail in Sec. II-B).

In addition, to extract the perceptually dominant color, we extract humans' region of interest (ROI) on images. Because in many searched web images, the ROI is the salient object inside, we exploits the object saliency maps from the recent works [13], [14] to highlight the salient object regions of an image. To localize the salient object, recent work [15] uses random forest to generate one globally best bounding box, in this work, we proposed a simple but practical strategy to localize the object much more efficiently.

B. Overview

Fig. 2 shows an overview of our system. In the color name learning procedure, we utilize the human labeled color chips to learn a probability $P(CN|\mathbf{c})$ through Gaussian mixture models (GMM), where CN is a color name and $\mathbf{c}$ is a color vector.

In the DCN region extracting procedure, given an input image, we first map the image to 12 affiliation maps based on $P(CN|\mathbf{c})$. Then for each affiliation map, the pixels that have high confidence, i.e. the probability of these pixels belonging to the color name is greater than a strict threshold, are reserved and merged as initial seeds regions for later propagation, but the regions with few pixels are filtered. After that, the contextual propagation is conducted from the seeds to include the neighboring ambiguous pixels that have lower confidence but contextually perceptible by human, resulting in a binary map for each DCN.

Figure 3. (a) A challenging synthetic image. The color name of each pixel obtained through the PLSA-bg of [9] and through the GMM of our approach are showed in (b) and (c) respectively.

As people are more likely to pay attention to the salient object in web images, in the last step, we estimate the salient region by a bounding box on the image getting from an object saliency map. Then the dominant colors over the salient object and background are re-estimated respecting ROI and Background (Bkg.), yielding perceptually dominant colors.

II. Contextual Dominant Color Name Extraction

Given an input image $I(\mathbf{x})$, where $\mathbf{x} = (x, y)^T$ indicates a pixel and each pixel has a color $\mathbf{c} = (r, g, b)^T$, our goal is to extract a set of dominant color names $\mathcal{C}_I = \{(CN_i, P_i), i = 1, \ldots, K_I\}$, where K_I denotes the number of color names in the image, CN_i is the name of the dominant color and P_i is the percentage of CN_i's region area comparing with that of image.

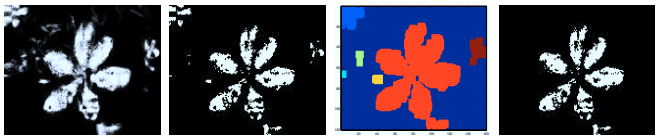
A. Color Names Learning

For each color name, we have a set of color vector: $\{\mathbf{c}_1, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_N\} \subset \mathcal{C}$ and respective labeled color names $\{CN_1, \ldots, CN_j, \ldots, CN_N\} \subseteq \mathcal{CN}$, where N is the labeled color chips number. We model the mapping by calculating a posterior probability for each color name $P(CN_i|\mathbf{c})$. In our case, we apply 12 color names in Russian based on [16], and the color name space is $\mathcal{CN} \equiv \{CN_i | CN_i \in \{$ black, blue, brown, grey, green, orange, pink, purple,red, white, yellow, teal $\}, i = 1, \ldots, K\}$, in which $K = 12$.

To get the training set, we take use of both the ebay images with masks on a single color name (see Fig. 7) and our enumerated RGB color. The color in the mask and color chips are manually labeled by multiple users into the 12 color names.

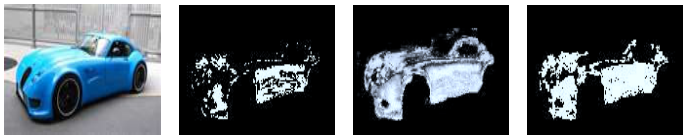
Another problem is which color space should be selected for learning. We considered RGB, HSL and Lab-space. As indicated in [9], we select Lab as it is perceptually linear, which means that similar differences between Lab values are considered about equally important color changes to humans. We assume a $D65$ white light source to compute the Lab value following [9].

We apply a generative approach, Gaussian Mixture Model (GMM) in our case, for the likelihood distribution estimation. For each color name CN_i, GMM could provide a



(a) (b) (c) (d)

Figure 4. Getting the seeds for propagation. (a) An affiliation map and (b) the binary map after thresholding. (c) A dilation operation is conducted to merge the spatially close pixels to regions. (d) We filter the small regions, resulting in the propagation seeds.



(a) (b) (c) (d)

Figure 5. Contextual propagation. (a) A searched web image and (b) the propagation seeds. (c) The affiliation map on which we perform propagation and (d) the result dominant color regions.

likelihood model $P(\mathbf{c}|CN_i)$. We further assume that the priori distribution $P(CN_i)$ of the color names is uniform. The posterior distribution could be written as:

$$P(CN_i|\mathbf{c}) = \frac{P(\mathbf{c}|CN_i)P(CN_i)}{\sum_{i=1}^{K} P(\mathbf{c}|CN_i)P(CN_i)}. \quad (1)$$

We tested the learned result by using a challenging synthetic image in Fig. 3, in which we assign each pixel the color name with the maximum posterior probability. Compared with the previous 11 color names learned by PLSA in [9], we also well segment the *teal* out.

In detail, to speed up the algorithm, this mapping is computed in preprocessing. We quantize the RGB-space equally into $32\times32\times32$ bins and pre-calculated the posterior probability distribution of each bin's center color.

B. Context-aware Dominant Color Region Extraction

With the posterior probability distribution, for the given image $I(\mathbf{x})$, each color of pixel $\mathbf{x}$ could be mapped into K probability values. Thus, the image $I(\mathbf{x})$ produces K affiliation maps $\mathcal{AM} = \{AM_i(\mathbf{x})\}$ as showed in Fig. 2.

For each affiliation map $AM_i(\mathbf{x})$, we aim at extracting the regions which could be assigned with the color name CN_i. We set those regions to 1 and the rest to 0 yielding a binary map $B_i(\mathbf{x})$. The output P_i is then computed by: $\frac{\sum_{\mathbf{x}} B_i(\mathbf{x})}{NX}$, where NX is the number of pixels of the image.

To generate the regions, we apply contextual propagation embedding the double thresholding method applied in Canny Edge Detection [17]. The scheme propagates from some seed regions aware of context information. In the following session, the details of propagation seeds and criteria for CN_i are presented. Other color names are similarly computed.

Propagation seeds. For a robust color name region, here we argue that the initial seeds $\{\mathbf{x}_{seed}\}$ for later propagation should contain at least three distinctive characteristics: (a) have high probability belonging to CN_i; (b) have a relatively large amount regarding the image pixel number; (c)should assemble closely or are spatially connected.

Formally, in our approach, a seed pixel $\mathbf{x}_{seed}$ for CN_i satisfies the following criteria sequentially:

$$\begin{aligned} C_{conf} &= AM_i(\mathbf{x}_{seed}) > T_{hi}, \\ C_{region} &= A(R_{i\mathbf{x}_{seed}}) > T_a, \end{aligned} \quad (2)$$

where $R_{i\mathbf{x}_{seed}}$ means the region containing $\mathbf{x}_{seed}$ and $A(R_{i\mathbf{x}_{seed}})$ is the area of $R_{i\mathbf{x}_{seed}}$. T_{hi} is the strict threshold for selecting the pixels with high confidence and T_a limits the minimum area of $R_{i\mathbf{x}_{seed}}$. To get $R_{i\mathbf{x}_{seed}}$, a thresholding is first conducted using the first criterion in Eqn.(2), then we merge the spatially close pixels into regions through dilation operation. We also showed the procedure of producing propagation seeds in Fig. 4. T_a in our experiments is set to be 3% of the image area. In Eqn.(2), the first one represents the characteristic (a) and the second one denotes (b) (c).

Propagation criteria. In Fig. 4(d), it seems that the seed regions $\{\mathbf{x}_{seed}\}$ represent the regions of CN_i well. In Fig. 5, however, the seeds with high confidence on the blue car only cover a small portion of human detected dominant color regions due to the light intensity's variation. For such a reason, a contextual propagation is applied to capture the neighbor regions that also potentially belong to CN_i.

In detail, we label the seeds $\{\mathbf{x}_{seed}\}$ as $+$, and the contextual propagation is performed through the following algorithm:

1) Append the labeled pixels with unlabeled neighborhood (we use 4 neighbors) into one sequence, saying $\{\mathbf{x}_s\} \subseteq \{\mathbf{x}_{seed}\}$.
2) Pop out a pixel $\mathbf{x}_s$ of the sequence. For each of its unlabeled neighborhood pixel $\mathbf{x}_{ns}$, append it into $\{\mathbf{x}_s\}$ and label it as $+$ if the pixel meets the following conditions:

$$\begin{aligned} C_{lowconf} &= AM_i(\mathbf{x}_{ns}) > T_{li}, \\ C_{connect} &= |AM_i(\mathbf{x}_{ns}) - AM_i(\mathbf{x}_s)| < T_s. \end{aligned} \quad (3)$$

3) Repeat step 2 until there is no pixel that can be further appended.

Here a loose threshold T_{li} is applied to search for the pixels potentially belonging to CN_i with weaker confidence, and T_s limits the maximum changing between the adjacent pixels' membership degree to CN_i. We set $T_s = 0.07$ in the experiments.

By taking the operation, as showed in Fig. 5, the pixels contextually belonging to CN_i are successfully detected. Our dominant color region extraction algorithm is also summarized in Alg. 1.

III. Perceptual Dominant Color Names

Motivated by human perception principles, under the application as show in Fig. 1(a), through clicking on the "red" color block after searching a query such as "Audi car", users intend to see the images with a red car from

Algorithm 1: $\mathcal{C}_I$ = DominantColorName(I)

Input: Image I with NX pixels

Output: Dominant Colors

$$\mathcal{C}_I = \{(CN_i, P_i), i = 1, \ldots, K_I\}$$

initialization();

calculate the affiliation maps $\{AM_i(\mathbf{x})\}$ of each color name CN_i by the $P(CN_i|\mathbf{x})$ in Eqn.(1);

foreach $AM_i(\mathbf{x})$, $i \leftarrow 1$ **to** K **do**

  calculate initial seeds $\{\mathbf{x}_{seed}\}$ by Eqn.(2);

  calculate $B_i(\mathbf{x})$ by contextual propagation based-on the criteria of Eqn.(3);

  set $P_i \leftarrow \frac{\sum_{\mathbf{x}} B_i(\mathbf{x})}{NX}$

Select the CN_i with $P_i \geq T_R$

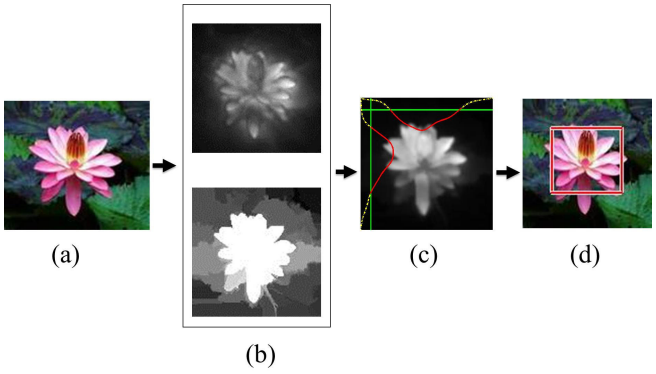


Figure 6. ROI detection. (a) A color Image. (b) Computed saliency maps from pixel level [13] and region level [14]. (c) We combine the two maps into one, then accumulate the saliency values on coordinate x and y separately and threshold the profiles. (d) The final bounding box.

the searching results, but the images containing a car with other colors in red background. This asks the detection of ROI on images.

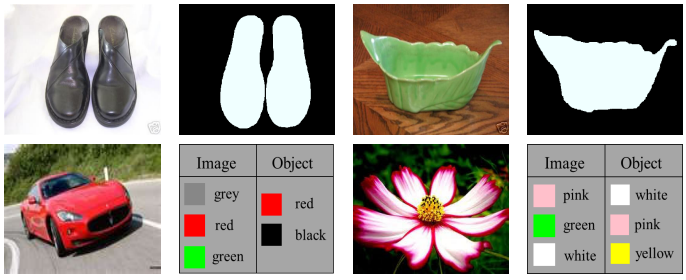
Furthermore, noticing that the web images always include a single large object region, we extract a single bounding box $\mathbf{w} = (t, l, b, r)^T$ to locate the ROI, where t, l, b and r denote the top, left, bottom and right edge position respectively. Within the $\mathbf{w}$, we could additionally calculate DCNs on ROI, $\mathcal{C}_{ROI} = \{(CN_i, P_{iROI}), i = 1, \ldots, K_{ROI}\}$, by computing the percentage of each color name's regions inside $\mathbf{w}$.

To get the $\mathbf{w}$, the object saliency maps from multiple cues are integrated as illustrated in Fig. 6:

1) Calculate the pixel level multi-scale contrast map $SM_{mc}(\mathbf{x})$ of [13] and region level contrast based map $SM_{cb}(\mathbf{x})$ of [14].
2) Combine the two maps into one single map by a non-linear scheme:

$$SM(\mathbf{x}) = (SM_{mc}(\mathbf{x}) + \alpha SM_{cb}(\mathbf{x}))^2, \qquad (4)$$

where α is a balancing parameter. The map is then normalized into the range $[0, 1]$.



(a) (b) (c) (d)

Figure 7. Human labeled image examples in Ebay database (the first row) and our database (the second row). (a)&(c) Sample images. (b)&(d) Corresponding labeled DCN ground truth.

3) Project the map on coordinate x, and the projected profile is later smoothed by a Gaussian filter. Formally, the profile $P(x) = G_\sigma * (\sum_y SM(\mathbf{x}))$. We later find the x which meets $P(x) \geq T_P$, resulting in several sections $\{S_j, j = 1 \ldots M\}$, where M is the section number. We chose the longest section S_m and set the boundary points of S_m to be the (l, r) of $\mathbf{w}$. The same operation is conducted on coordinate y to compute (t, b).

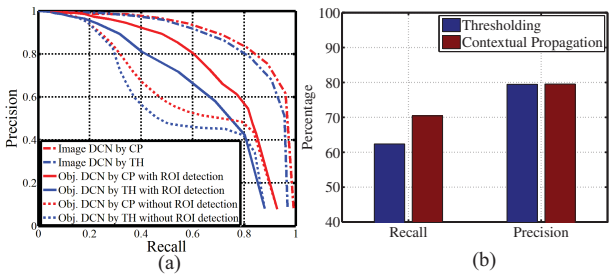
In our experiments, the $\alpha = 1$, and the $T_P = \beta\overline{P(x)}$, where $\overline{P(x)}$ is the mean value of $P(x)$. We set $\sigma = 5$ and the $\beta = 0.7$ by validating on a subset of the MSRA $\mathcal{B}$ dataset with a bounding box ground truth for each image provided by [13].

IV. Experiments and Evaluation

Database. We tested the algorithm on two dataset. The first one is Ebay image set provided by [9] which contains four kinds of objects, and each object has 132 images with an mask on the DCN regions of the object inside, as show in Fig. 7. For accurate evaluation, we pruned the images containing same color in object and background.

Furthermore, targeting DCNs' detection on web images, we constructed a database with images collected from searched object queries on bing's image search engine. In detail, we searched 15 object queries. For each query, as to collect enough images of each color name, we searched the queries by adding a color name term, e.g. "blue+car". Then, images collected from all color names of a single query are merged into one folder with random order. In total, 3600 images are collected for each query. We further pruned the images with no object and randomly select 200 images from each query folder. We later asked different users to label the 3000 images. Each image is labeled at most three DCNs both in object region and the whole image region. Fig. 7 shows two labeled examples.

Comparison. To our best knowledge, we are the first aiming at proposing DCN combining both low level dominant color descriptor and high level color naming. Thus, for evaluation, we just compared our method with the simple but broadly applied thresholding method which is resulted



(a) (b)

Figure 8. (a) The recall-precision curve under different testing conditions on our labeled database (CP indicates the Contextual Propagation, TH indicates the Thresholding and GT indicates ground truth). (b) The recall and precision on Ebay database.

from thresholding the affinity maps by the first criterion in Eqn.(2).

Thresholds. In our approach, the most influential parameters are the strict thresholds $\{T_{hi}\}_{i=1}^{K}$ for initial seeds and loose thresholds $\{T_{li}\}_{i=1}^{K}$ for contextual propagation. Over strict thresholds would exclude many correctly detected pixels and vice versa. Moreover, the probability distributions of color names learned through GMM (Sec.II) are various in maximum value and standard variation, which indicates that using mono threshold for all color name generates undesired results. We propose to adopt a validation set to find the thresholds T_{hi} and T_{li} for each color name CN_i. In practice, we enumerate the thresholds in the range $[0.2, 0.9]$ with step 0.1. Finally, our thresholds learned are $\{0.5, 0.8, 0.4, 0.5, 0.7, 0.8, 0.7, 0.9, 0.5, 0.5, 0.9, 0.8\}$ and $\{0.3, 0.4, 0.2, 0.3, 0.3, 0.5, 0.5, 0.6, 0.3, 0.3, 0.6, 0.4\}$ for the 12 color names respectively.

A. Quantitative Evaluation

To show the performance of our method, the recall-precision curves under different testing situations are calculated on our database by changing the threshold T_R in Alg. 1. We compared the detected DCNs and the human labeled DCNs in our database. The results are showed in Fig. 8(a). Comparing the DCNs detected using contextual propagation and thresholding (the red line and blue line), we can see that in all cases, the contextual propagation method out-performs thresholding as expected. This is because in most images, though the thresholding can accurately capture a portion of the dominant color regions by our validated thresholds, the correct color names are more likely to be reserved after propagation with increasing of T_R.

Comparing the result of detected image DCNs evaluated using image DCN ground truth (GT) and detected object DCNs evaluated using object DCN GT (the dash-dot line and solid line), the precision-recall of the object DCNs is negatively impacted by the object localization accuracy. But we still see a significant improvement on the objects' DCN detection accuracy using the DCNs within the detected ROI rather than using the DCNs of the whole image (the solid line and the dot line).

Figure 9. Color name region extraction results. Sample images (the first row) with their respective extracted color name regions (the second row).

For further testing the effectiveness and region accuracy by contextual propagation, we utilize the color mask of Ebay images set by computing the pixel level recall and precision of detected DCN regions. Mathematically, for each DCN, given a GT mask $B_{gt}(\mathbf{x})$ and a detected mask $B_o(\mathbf{x})$ of one test image, the recall and precision are computed by:

$$Recall = \frac{\sum_{\mathbf{x}} B_{gt}(\mathbf{x}) B_o(\mathbf{x})}{\sum_{\mathbf{x}} B_{gt}(\mathbf{x})}, Precision = \frac{\sum_{\mathbf{x}} B_{gt}(\mathbf{x}) B_o(\mathbf{x})}{\sum_{\mathbf{x}} B_o(\mathbf{x})}. \quad (5)$$

As can be seen in Fig. 8(b), by our contextual propagation, the regions' recall is largely improved (10%) whilst keeping a high precision.

B. Qualitative Results

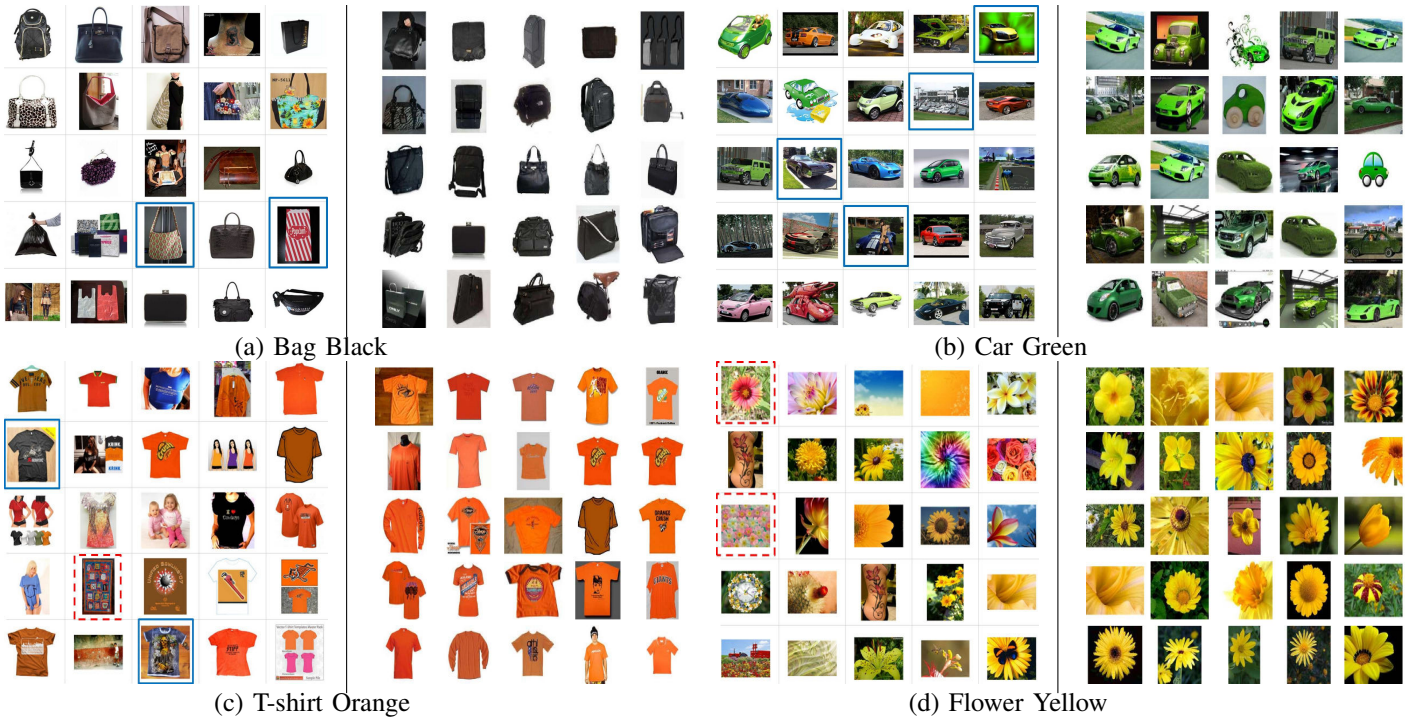
Fig. 9 shows the color name regions (showed by color) detected by our algorithm. Most of color regions are successfully detected and compact in our experiments. Fig. 10 shows four results simulating the search engines' color filter function mentioned in Fig. 1. We filter the collected searched web images of each query based on each DCN's area on ROI. Comparing with the thresholding method, our scheme shows significant improvement on visual pleasure just depending on the image content. From the comparison, we could see the images in which the demanding DCN regions are in a scattered layout (marked with dash box) and images in which the demanding DCN just dominates the background (marked with solid box) are failed to be excluded by thresholding. However, we handle those problems properly in most of the cases.

V. Conclusion

We proposed a new approach that effectively detects the human perceptually dominant color names of an image motivated by a real application. In our work, the color names learning, context-aware dominant color region extraction and ROI extraction from saliency maps are integrated together to identify perceptually dominant color names, showing promising results for web image search.

References

[1] S.-F. Chang, T. Sikora, and A. Puri, "Overview of the mpeg-7 standard," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 11, no. 6, pp. 688–695, 2001.



(a) Bag Black (b) Car Green (c) T-shirt Orange (d) Flower Yellow

Figure 10. Examples of the color filter results from thresholding (left column) and our method (right column). The rectangles mark out the images failed to be filtered by thresholding.

[2] S. Kiranyaz, S. Uhlmann, and M. Gabbouj, "Dominant color extraction based on dynamic clustering by multi-dimensional particle swarm optimization," *Content-Based Multimedia Indexing, International Workshop on*, vol. 0, pp. 181–188, 2009.

[3] J. Chamorro-Martínez, J. M. Medina, C. D. Barranco, E. Galán-Perales, and J. M. Soto-Hidalgo, "Retrieving images in fuzzy object-relational databases using dominant color descriptors," *Fuzzy Sets and Systems*, vol. 158, no. 3, pp. 312–324, 2007.

[4] N. Krishnan, M. S. Banu, and C. C. Christiyana, "Content based image retrieval using dominant color identification based on foreground objects," *Computational Intelligence and Multimedia Applications, International Conference on*, vol. 3, pp. 190–194, 2007.

[5] M. Rao, B. Rao, and A.Govardhan, "Article: Ctdcirs: Content based image retrieval system based on dominant color and texture features," *IJCA*, vol. 18, no. 6, pp. 40–46, 2011.

[6] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution*. Los Angeles: University of California Press, 1969.

[7] C. L. Hardin and L. Maffi, *Color Categories in Thought and Language*. Cambridge University Press, 1997.

[8] D. Conway, "An experimental comparison of three natural language colour naming models," in *Proceedings of the EWHCI*, 1992, pp. 328–339.

[9] J. van de Weijer, C. Schmid, J. J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1512–1523, 2009.

[10] R. Benavente, M. Vanrell, and R. Baldrich, "A data set for fuzzy colour naming," *Color Research and Application*, vol. 31, no. 1, pp. 48–56, Feb 2006.

[11] T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR*, 1999, pp. 50–57.

[12] Aiguo Li and Xiyi Bao, "Extracting image dominant color features based on region growing," in *WISM*, 2010, pp. 120–123.

[13] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 33, no. 2, pp. 353–367, 2011.

[14] H. Jiang, J. Wang, Z. Yuan, T. Liu, and N. Zheng, "Automatic salient object segmentation based on context and shape prior," in *BMVC*, 2011, pp. 110.1–110.12.

[15] P. Wang, J. Wang, G. Zeng, J. Feng, H. Zha, and S. Li, "Salient object detection for searched web images via global saliency," in *CVPR*. IEEE, 2012.

[16] B. Berlin, *Basic Color Terms: Their Universality and Evolution*. Berkeley/Los Angeles: University of California Press, 1969.

[17] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.