

On Comparing Color Spaces for Food Segmentation

Sinem Aslan^(✉), Gianluigi Ciocca, and Raimondo Schettini

Department of Informatics, Systems and Communication,
University of Milano-Bicocca, Milano, Italy
{sinem.aslan,ciocca,schettini}@disco.unimib.it

Abstract. Accurate segmentation of food regions is important for both food recognition and quantity estimation and any error would degrade the accuracy of the food dietary assessment system. Main goal of this work is to investigate the performance of a number of color encoding schemes and color spaces for food segmentation exploiting the JSEG algorithm. Our main outcome is that significant improvements in segmentation can be achieved with a proper color space selection and by learning the proper setting of the segmentation parameters from a training set.

Keywords: Automatic food segmentation · Color spaces · JSEG

1 Introduction

Measuring nutrition intake and food calorie in daily diets is important to not only treat and control food-related health problems, but also for people who want to be aware of their nutrition habits to maintain a healthy weight. Recent developments at vision-based measurement [1–3] has gained significant attention from community dealing with dietary assessment, since the process is quietly simplified for the users, i.e., they simply take the photo of their food with a mobile device and calorie calculation is achieved automatically in the pipeline of processes employing computer vision techniques.

A general pipeline of calorie calculation by vision-based measurement consists of four stages [1]: (i) Preprocessing for image enhancement; (ii) Food segmentation to determine the food regions inside dishes; (iii) Food recognition where representative features are extracted on the segmented regions and fed into a classifier; (iv) Calorie measurement where the mass of the food is estimated and corresponding calories are computed using existing nutrition tables. In this paper, we have focused on the second stage, i.e. food region segmentation that greatly influence the accuracy of the subsequent stages.

Plenty of papers for food segmentation has been published and a number of outstanding ones are presented in Table 1. The literature works on food segmentation have employed a variety of segmentation schemes, e.g., thresholding, active contours, JSEG, normalized cuts, mean shift, etc., by utilizing a different

color space, e.g., gray-scale, CIELUV, CIELAB. Moreover, performance of these algorithms have been evaluated on different food image datasets. In this work, we aim to make a comparative evaluation of different color encoding schemes and color spaces for food region segmentation on the same dataset and by using the same segmentation scheme.

The color encoding schemes and color spaces [4,5] that we have considered are $Y'IQ$, $Y'CbCr$, $Y'PbPr$, $Y'DbDr$, CIEXYZ, CIELAB and CIELUV, $O_1O_2O_3$, rgb (normalized RGB), and $I_1I_2I_3$. $Y'IQ$, $Y'CbCr$, $Y'PbPr$, and $Y'DbDr$ are the luma and chroma encoding systems that separates the sRGB into one luminance and two chrominance components. Taking advantage of human vision's sensitivity to changes on the luminance component, these systems are useful for compression applications. CIEXYZ, CIELAB and CIELUV colorimetric spaces are device independent, i.e., they do not depend on the parameter settings of the devices but represent the colors based on response of an ideal standard observer to wavelengths of light. CIELAB and CIELUV are perceptually uniform, i.e., the Euclidean distance between two colors in CIELAB and CIELUV is strongly correlated to the distance perceived in human vision. rgb is invariant to surface orientation, illumination direction and intensity [4]. O_1 and O_2 components of the opponent color space $O_1O_2O_3$ are independent of highlights, but sensitive to surface orientation, illumination direction and illumination intensity, while O_3 has no invariant property [4]. Color information is separated into three approximately orthogonal components at $I_1I_2I_3$ and it is reported as useful for segmentation in [5].

Table 1. Literature works on food region segmentation

Publication	Color space	Segmentation scheme
Shroff et al. [6]	Gray-scale	Adaptive thresholding
He et al. [7]	Gray-scale	Active contours, normalized cuts, and local variation performed on the detected foreground
Zhu et al. [8]	Gray-scale	Normalized cuts performed on the foreground region, that is tuned regarding to the resulted food recognition performance
Anthimopoulos et al. [1]	CIELAB	Mean-shift performed on the plate region which is detected by RANSAC
Bettadapura et al. [2]	CIELAB and gray-scale	Hierarchical segmentation with the GPS-based location heuristics providing assumptions for segmentation
Ciocca et al. [3]	CIELUV	JSEG on the detected food regions
Matsuda et al. [9]	CIELUV and gray-scale	Regions that are segmented by Deformable Part Model, a circle detector and JSEG are integrated according to the score of classification accuracy they provide

We have chosen the well-known JSEG automatic color segmentation algorithm [10] to integrate the computations in different color spaces. JSEG has been successfully used in many literature works and the published source code [11] yields modifications on the method conveniently.

The experiments are done on automatically cropped images of UNIMIB2016 food dataset [3] which includes a wide range of food types with both bounding box and polygon annotations.

2 JSEG

JSEG [10], illustrated in Fig. 1, accomplishes segmentation in two main stages, i.e., *color quantization* and *spatial segmentation*. In the first stage, the colors of images are coarsely quantized into several representing classes to obtain a class-map where each pixel is labeled with its corresponding color class label. It is suggested in [10] to use the color quantization algorithm developed by Deng et al. [12] which conforms to human perception sensitivity. According to this method [12], the images are initially smoothed by *Peer Group Filtering (PGF)* which avoids to blur the edges. Then, using the local statistics provided by PGF, the color quantization algorithm with the implementation steps as follows is performed: (1) Assign weights to pixels in a way that noisy regions are weighted less and smoothed regions are weighted more; (2) Estimate the initial number of clusters by considering the smoothness of the entire image, i.e., the less smooth the image is, the higher number of initial clusters is; (3) Determine the initial clusters by splitting initialization algorithm [12] and implement vector quantization by modified Generalized Lloyd Algorithm (GLA) which incorporates the weights that were computed at the first step; (4) Perform an agglomerative clustering algorithm [13] to merge the close clusters until the minimum distance between two centroids satisfies a preset threshold T_Q . The novelty of the algorithm in [12] lies at the weighting scheme employed at the first step which yields GLA to shift the centroids towards points with higher weights, i.e., smoother regions.

In the second stage of JSEG, a homogeneity measurement, called as *J-value*, is computed by using the obtained color class-map in a local window around each pixel. High and low J-values indicate possible region boundaries and centers, respectively. Computed J-values for all pixels form a gray-scale pseudo-image called as *J-Image* and computing J-values in N different window sizes results with J-images in N number of scales. Small and larger sized windows provide to

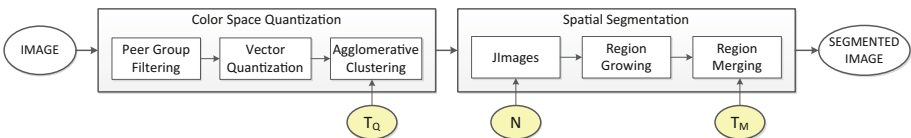


Fig. 1. Schematic of the JSEG algorithm. The ellipses in yellow represent the algorithm's parameters. (Color figure online)

localize color edges and detect texture boundaries, respectively, and it is useful to employ multiple scales of J-Images in the segmentation process in order to facilitate from both information. What is next is, the resulted multi scale J-Images are used by a region growing scheme in an iterative way to accomplish the initial segmentation which essentially constitutes to the over segmentation of the input image. In order to obtain the final segmentation, over segmented regions are merged by the agglomerative method [13] that was already employed at the color quantization algorithm. The most similar neighbour region pairs are merged until the minimum Euclidean distance between two histogram features satisfies a preset threshold T_M .

JSEG and the employed color quantization scheme process images in the CIELUV color space. Three parameters are set by users in the whole process, i.e., color quantization threshold (T_Q), number of scales of J-Images (N), and region merge threshold (T_M). These parameters directly influence the segmentation results. Low values of both the color quantization threshold T_Q and region merge threshold T_M encourage over segmentation. Finer details are segmented with higher values of N and vice versa.

3 Experimental Setup

Food Dataset. We have used UNIMIB-2016 dataset [3] since it includes a wide variety of food types, i.e. 1,027 tray images including 73 food categories; in addition to the bounding box annotations, the published polygon annotations provide evaluation with more precise ground truth compared to existing datasets; and it is sufficiently challenging for segmentation. The main challenges can be listed as (i) white colored placemats and plates complicates to segment the food regions in the same color, e.g., *riso in bianco* and *pasta pesto besciamella e cornetti* (see Fig. 2a), (ii) includes multiple food segmentation problem, since side and main dishes are served in the same plate (see Fig. 2b); (iii) images are acquired in an uncontrolled environment by a hand-held smart phone and includes illumination (see Fig. 2c).

Differently from [3], we assume that the photos of food regions on a tray were shot individually in this paper. In order to obtain such material, we cropped the tray images into subimages by exploiting the published bounding box annotations as a subimage would include the Region of Interest (ROI), i.e., food region. Each sub-image is cropped from a custom space $d_h/2$ and $d_v/2$ from the borders of the bounding box, where d_h and d_v are the distances (in pixels) from center to horizontal and vertical borders of the bounding box. We desire to crop main and side foods together in a single subimage and with a new bounding box annotation covering both. Thus, we co-cropped foods if their bounding boxes overlap in the ratio of 95%. Using this simple heuristic we obtain the new dataset that includes 2,679 images and with a quick check we eliminated 50 images which were not cropped at all due to very close positions of the foods on the trays. A new challenge as a result of automatic cropping is the “noise” objects around ROI (See Fig. 2d). The dataset of cropped UNIMIB-2016 images and their polygon and bounding box annotations will be published.

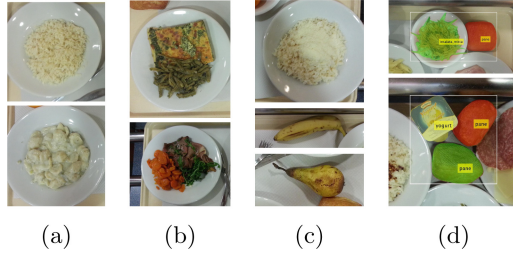


Fig. 2. Challenges of UNIMIB2016 dataset (automatically cropped images).

Parameter Setting Schemes for JSEG. The JSEG default values suggested in the published implementation [11] are $T_Q = 250$, $T_M = 0.4$, and although the parameter N can be set by the user, it is suggested in [10, 11] to use the automatic setting in JSEG which specifies N according to the input image size. It is mentioned in [10] that JSEG works well on a large set of images, i.e., 2500 images, with the mentioned fixed values of the parameters without any requirement for tuning. However, transforming the input images to other color spaces requires to update the fixed value of T_Q while N and T_M would not get affected from this operation. Thus, we have used the default values of $T_M = 0.4$ and $N(\text{automatic})$ [11] at the experiments, and we define another termination criterion for color quantization which is independent to underlying color space. The new criterion considers the *resulting number of clusters* after merging operation instead of minimum distance between quantized colors.

We have followed two approaches for setting of T_C : (i) *Fixed scheme of parameter setting.* We fix the T_C to the value which yields segmentation performance be most close to (or slightly better than) the performance obtained with the default parameter setting, i.e., $T_Q = 250$, for images in CIELUV color space [11]. (ii) *Optimized scheme of parameter setting.* We learn the value of T_C from a training set for each color space individually.

4 Results

We have resized the images as their smallest length would be 128 and 256 pixels to investigate performance at different image sizes. In order to assess the quality of the segmentation, we applied the evaluation benchmarks suggested in [14]. Specifically, we compute *boundary-based* measurements Precision (P), Recall (R) and Fscore (F) and *region-based* measurements, i.e., covering (of ground truth by segmentation), Probabilistic Rand Index (PRI), and the Variation of Information (VI). Differently from [14], we have one ground truth data and one scale of segmentation (since we do not perform hierarchical segmentation) for each image. P, R, F and segment covering are aggregated scores on the whole dataset, i.e., fractions are computed after aggregating statistics from all images, whereas PRI and VI are the averaged scores over number of images [14].

For the fixed scheme of parameter setting, we compute the performance scores on the whole dataset, i.e., 2,629 images. For the optimized scheme, we randomly sample 200 images to construct the training set and learning the optimal parameter value on the training set, we present the performance results on the remaining 2,429 testing set images.

4.1 Fixed Scheme for JSEG Parameter Selection

At the first stage of the fixing scheme, we have segmented 2,629 images in CIELUV color space by $T_Q = 250$ setting as suggested in [11], and with a number of T_C settings, i.e., $T_C = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$. We have obtained the performance results in Table 2. In this experiment, we evaluate the quality of segmentation with respect to the average of boundary and region based Fscores, i.e. $(F_{boundary} + F_{region})/2$, in order to include contribution of both region and boundary-based assessment. We observe in Table 2 that in comparison with $T_Q = 250$ setting, the closest and slightly better performance is obtained with $T_C = 4$.

Table 2. Performance results, in terms of $(F_{boundary} + F_{region})/2$, that are obtained with default setting of T_Q and different settings of T_C .

Image size at shortest side	$T_Q = 250$	$T_C = 2$	$T_C = 3$	$T_C = 4$	$T_C = 5$	$T_C = 6$	$T_C = 7$	$T_C = 8$	$T_C = 9$	$T_C = 10$
128 pix.	0.49	0.62	0.55	0.51	0.47	0.43	0.40	0.38	0.35	0.33
256 pix.	0.45	0.61	0.53	0.48	0.43	0.35	0.39	0.33	0.30	0.28

In the second stage, we fix $T_C = 4$, and segment the images in other color spaces. The obtained performance results are given in Table 3. It is observed that the highest boundary based Fscore is obtained with CIELUV, which is followed by Y'DbDr and rgb in both image sizes. Moreover, covering score of Y'DbDr is 3% and 2% better than CIELUV and rgb respectively in both image sizes. PRI and VOI scores are also compatible with this observation. Among all CIEXYZ is the worst in all experiments.

4.2 Optimized Scheme for JSEG Parameter Selection

We have measured the score of $(F_{boundary} + F_{region})/2$ with each $T_C \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ setting on training images, and the best performed setting is employed in segmentation of the testing images. We present the performance results with optimal T_C setting for each color space in Table 4. We also include the performance that we obtained with the published implementation of JSEG that works in CIELUV with fixed $T_Q = 250$ setting [10, 11].

Table 3. Performance results obtained by JSEG with fixed $T_C = 4$ setting varying the color spaces.

Color space	128 Pix.						256 pix.					
	Boundary-based			Region-based			Boundary-based			Region-based		
	P	R	F	Covering	PRI	VOI	P	R	F	Covering	PRI	VOI
Y'CbCr	0.27	0.45	0.33	0.57	0.65	1.82	0.20	0.51	0.29	0.54	0.63	1.14
Y'DbDr	0.34	0.48	0.40	0.69	0.73	1.34	0.28	0.55	0.37	0.67	0.72	1.55
Y'IQ	0.28	0.43	0.34	0.62	0.68	1.66	0.21	0.50	0.30	0.59	0.66	1.96
Y'PbPr	0.28	0.44	0.34	0.62	0.68	1.64	0.21	0.51	0.30	0.58	0.66	1.97
CIELAB	0.23	0.37	0.29	0.54	0.63	1.88	0.18	0.44	0.25	0.52	0.62	2.17
CIELUV	0.33	0.50	0.40	0.66	0.71	1.46	0.28	0.56	0.38	0.64	0.70	1.63
CIEXYZ	0.20	0.38	0.26	0.43	0.56	2.35	0.16	0.48	0.24	0.41	0.55	2.73
rgb	0.33	0.48	0.39	0.67	0.72	1.41	0.27	0.54	0.37	0.65	0.71	1.59
$O_1O_2O_3$	0.21	0.40	0.28	0.46	0.58	2.25	0.17	0.48	0.25	0.42	0.56	2.64
$I_1I_2I_3$	0.20	0.39	0.27	0.44	0.57	2.33	0.16	0.47	0.23	0.40	0.55	2.76

Table 4. Performance results obtained by the optimal value of T_C learned on the training set for each color space. (*)Benchmark using $T_Q = 250$

Color space	128 Pix.						256 pix.					
	Boundary-based			Region-based			Boundary-based			Region-based		
	P	R	F	Covering	PRI	VOI	P	R	F	Covering	PRI	VOI
Y'CbCr	0.30	0.32	0.31	0.66	0.70	1.26	0.24	0.34	0.28	0.65	0.68	1.39
Y'DbDr	0.49	0.37	0.42	0.79	0.81	0.79	0.45	0.40	0.42	0.78	0.81	0.82
Y'IQ	0.34	0.32	0.33	0.70	0.71	1.12	0.28	0.35	0.31	0.69	0.72	1.22
Y'PbPr	0.34	0.32	0.33	0.70	0.73	1.11	0.28	0.35	0.31	0.69	0.72	1.21
CIELAB	0.25	0.33	0.28	0.59	0.66	1.60	0.21	0.28	0.24	0.61	0.65	1.47
CIELUV	0.47	0.42	0.45	0.79	0.82	0.84	0.43	0.45	0.44	0.79	0.81	0.88
CIEXYZ	0.27	0.32	0.30	0.63	0.67	1.41	0.22	0.34	0.27	0.62	0.67	1.52
rgb	0.52	0.40	0.45	0.82	0.84	0.71	0.49	0.43	0.46	0.81	0.83	0.74
$O_1O_2O_3$	0.27	0.32	0.29	0.63	0.67	1.42	0.22	0.34	0.27	0.62	0.67	1.52
$I_1I_2I_3$	0.26	0.32	0.29	0.63	0.67	1.44	0.21	0.34	0.26	0.61	0.66	1.56
CIELUV(*)	0.32	0.51	0.39	0.64	0.70	1.57	0.26	0.58	0.36	0.60	0.68	1.85

We list our observations as follows: (i) *Comparison of color spaces*: rgb and CIELUV gives the same best boundary-based Fscore at the smaller sized images while rgb is 2% better than CIELUV for larger sized images. rgb outperforms others in all region-based scores. Y'DbDr follows them both in boundary and region based scores. The worst performances are obtained for CIELAB and $I_1I_2I_3$; (ii) *Comparison with Table 3*: Optimizing T_C improved boundary-based performance for most of the color spaces, e.g., $\sim 6\%$, $\sim 5\%$, and $\sim 2\%$ improvement in Fscore is obtained for rgb, CIELUV and Y'DbDr, respectively, for smaller sized images, and even more for larger size images. Besides, performance at

Y'CbCr, YIQ, Y'PbPr and CIELAB slightly ($\sim 1\%$) degrades with optimized T_C at smaller sized images, but same for all at larger sized images. Optimizing T_C improved region-based scores significantly for all color spaces, e.g., around 15%, 16%, 10% and 20% improvement in covering score is achieved for rgb, CIELUV, Y'DbDr and CIELAB at both image sizes, respectively; *(iii) Comparison with benchmark*: Default JSEG implementation with fixed $T_Q = 250$ at CIELUV gives better boundary-based recall, however since their precision is not good enough optimized scheme outperforms benchmark in the rates of $\sim 6\%$ and $\sim 10\%$ at boundary-based Fscore for small and larger sized images, respectively. Improvement in region-based performance is even more remarkable, i.e., in the rates of $\sim 20\%$.

5 Conclusion

In this paper we studied the segmentation algorithm of the processing pipeline for food dietary assessment. We focused on color space selection food segmentation. More precisely, an extensive comparative evaluation of ten color encoding scheme and spaces is made by using the well-known JSEG segmentation algorithm. We have also investigated the optimal parameter setting for JSEG to work in different color spaces. Experimental results show that representations in Y'DbDr and rgb is to be preferred for food segmentation.

References

1. Anthimopoulos, M., Dehais, J., Diem, P., Mougiakakou, S.: Segmentation and recognition of multi-food meal images for carbohydrate counting. In: Proceedings of the IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE 2013), pp. 1–4 (2013)
2. Bettadapura, V., Thomaz, E., Parnami, A., Abowd, G.D., Essa, I.: Leveraging context to support automated food recognition in restaurants. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV 2015), pp. 580–587 (2015)
3. Ciocca, G., Napoletano, P., Schettini, R.: Food recognition: a new dataset, experiments, and results. IEEE J. Biomed. Health Inform. **21**(3), 588–598 (2017)
4. Lee, D., Plataniotis, K.N.: A taxonomy of color constancy and invariance algorithm. In: Celebi, M.E., Smolka, B. (eds.) Advances in Low-Level Color Image Processing. LNCVB, vol. 11, pp. 55–94. Springer, Dordrecht (2014). https://doi.org/10.1007/978-94-007-7584-8_3
5. Ohta, Y.I., Kanade, T., Sakai, T.: Color information for region segmentation. Comput. Graph. Image Process. **13**(3), 222–241 (1980)
6. Shroff, G., Smailagic, A., Siewiorek, D.P.: Wearable context-aware food recognition for calorie monitoring. In: Proceedings of the 12th IEEE International Symposium on Wearable Computers (ISWC 2008), pp. 119–120 (2008)
7. He, Y., Khanna, N., Boushey, C., Delp, E.: Image segmentation for image-based dietary assessment: a comparative study. In: Proceedings of the IEEE International Symposium on Signals, Circuits and Systems (ISSCS 2013), pp. 1–4 (2013)

8. Zhu, F., Bosch, M., Khanna, N., Boushey, C.J., Delp, E.J.: Multiple hypotheses image segmentation and classification with application to dietary assessment. *IEEE J. Biomed. Health Inform.* **19**(1), 377–388 (2015)
9. Matsuda, Y., Hoashi, H., Yanai, K.: Recognition of multiple-food images by detecting candidate regions. In: *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2012)*, pp. 25–30 (2012)
10. Deng, Y., Manjunath, B.: Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(8), 800–810 (2001)
11. Deng, Y., Manjunath, B.: JSEG Project. <http://old.vision.ece.ucsb.edu/segmentation/jseg/software/> (1999). Accessed 27 June 2017
12. Deng, Y., Kenney, C., Moore, M.S., Manjunath, B.: Peer group filtering and perceptual color image quantization. In: *Proceedings of the IEEE International Symposium on Circuits and Systems, (ISCAS 1999)*, Vol. 4, pp. 21–24. IEEE (1999)
13. Duda, R.O., Hart, P.E., Stork, D.G., et al.: *Pattern classification*, vol. 2. Wiley, New York (1973)
14. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.* **33**(5), 898–916 (2011)