On Comparing Color Spaces for Food Segmentation

Sinem Aslan, Gianluigi Ciocca, Raimondo Schettini

Università degli Studi di Milano-Bicocca, {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.

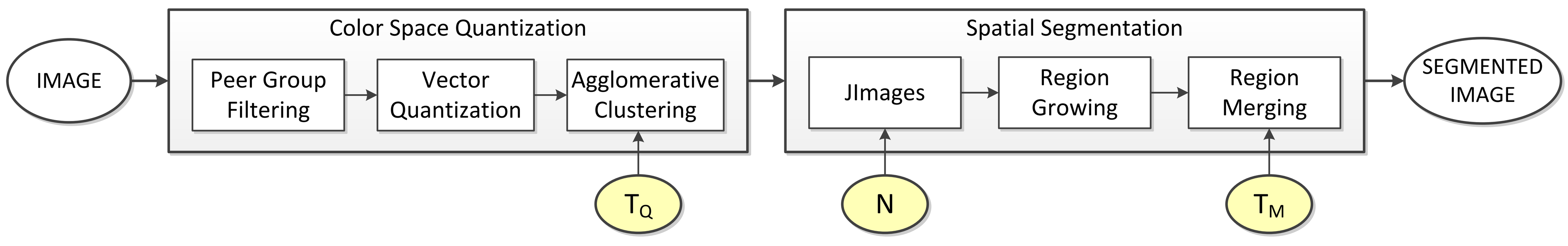
Motivation & System workflow

- Literature works uses a variety of segmentation schemes, each employs a different color space and evaluated on different datasets.
- We aim to make a comparative evaluation of *different color encoding schemes and color spaces* for food region segmentation, on the *same dataset* and using the *same segmentation scheme*.

The segmentation algorithm: JSEG
- Successfully used in many literature works
- Published source code yields modification on the method conveniently

Dataset: automatically cropped UNIMIB2016 images
- Includes a wide range of food types with both bounding box and polygon annotations.
- Sufficiently challenging for segmentation.

- **Schematic of the JSEG [1] algorithm:**



- User specified parameters directly infuence the segmentation results:
  - Low values of 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.
- Suggested default values [1] are T_Q = 250 (CIELUV), T_M = 0.4, N:automatic. 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.

Food dataset: Automatically cropped UNIMIB-2016 images [2]

- 1,027 tray images (2629 cropped images) including 73 food categories
- Bounding box & polygon annotations: Evaluation with more precise ground truth

foods in same color with the plates | multiple food in a plate | illumination changes | "noise" objects around ROI

Approach & Results

We employed a new criterion for color quantization which considers the resulting number of clusters (T_C) after merging operation instead of minimum distance (T_Q) between quantized colors.

We performed two schemes of parameter settings:

1. 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.

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

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	2.14
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

- The highest boundary based Fscore is obtained with CIELUV, which is followed by Y'DbDr and rgb in both image sizes.
- 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.

2. Optimized scheme of parameter setting

We learn the value of T_C from a training set for each color space individually.

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

	Boundary-based performance	Region-based performance
Comparison with fixed scheme	6%, 5%, and 2% improvement for rgb, CIELUV and Y'DbDr.	15%, 16%, 10% and 20% improvement for rgb, CIELUV, Y'DbDr and CIELAB.
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. Y'DbDr follows them both in boundary and region based scores.
Comparison with benchmark	Benchmark gives better boundary-based recall, however since their precision is not good enough optimized scheme outperforms benchmark in the rates of 6% and 10% at boundary-based Fscore.	Improvement in region-based performance is even more remarkable, i.e., in the rates of 20%

References

[1] Deng, Y., Manjunath, B.S., Unsupervised segmentation of color-texture regions in images and video. IEEE Trans. Pattern Anal. Mach. Intell., 23(8) pp. 800–810, (2001)

[2] Ciocca, G., Napoletano, P., Schettini, R., Food Recognition: A New Dataset, Experiments, and Results. IEEE journal of biomedical and health informatics, 21(3), pp. 588–598 (2017).