A Multimedia Database for Automatic Meal Assessment Systems

ICIAP MADiMa 2017

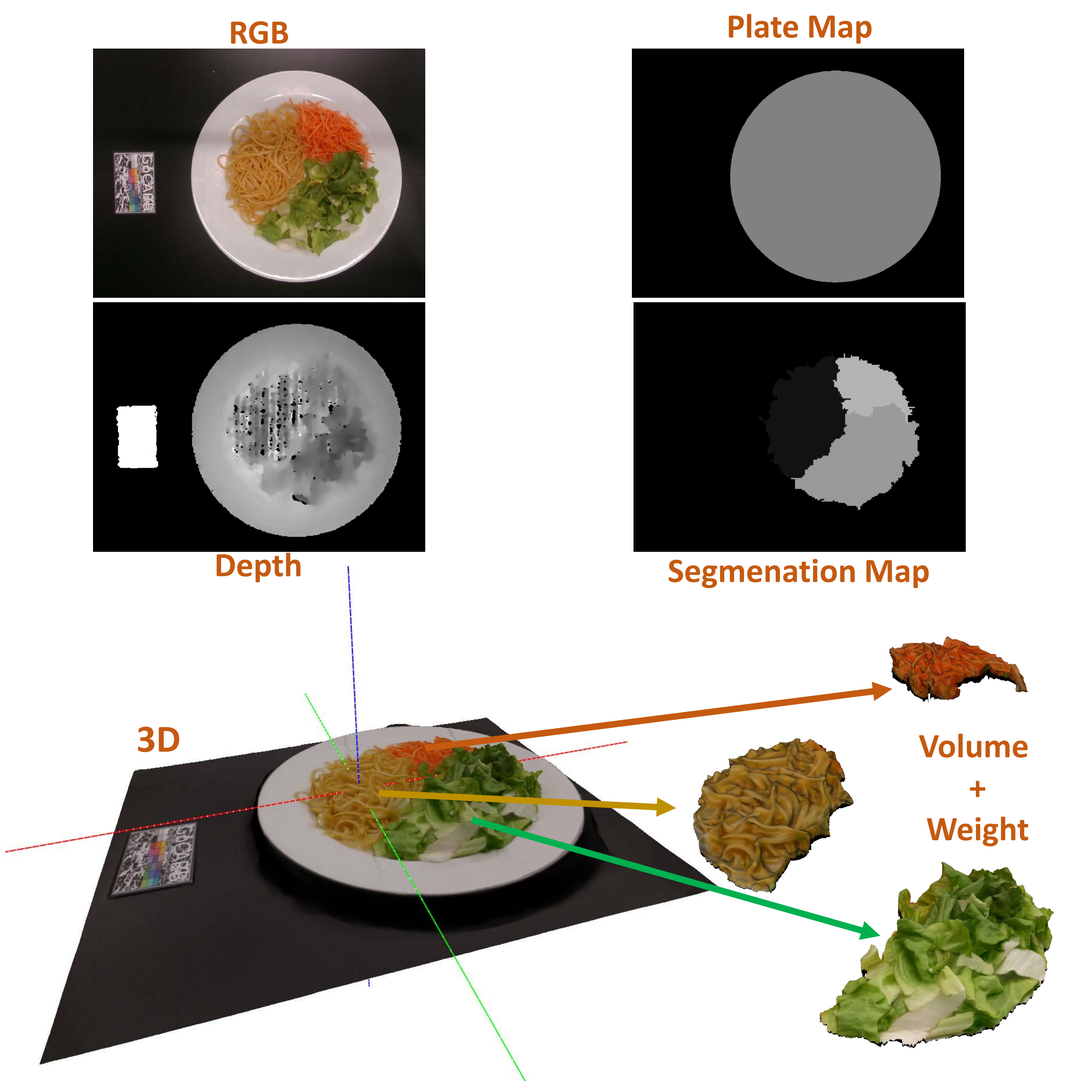
Dario Allegra, Marios Anthimopoulos, Joachim Dehais, Ya Lu, Filippo Stanco, Giovanni Maria Farinella and Stavroula Mougiakakou

{allegra, gfarinella, fstanco}@dmi.unict.it; {marios.anthimopoulos, joachim.dehais, ya.lu, stavroula.mougiakakou}@artorg.unibe.ch

Abstract

Diet management relies on the rather challenging task of food intake monitoring and assessment. To facilitate this procedure, several systems have been recently proposed for automatic meal understanding on mobile devices using computer vision methods. The development and validation of these systems requires large amounts of data and although some public datasets already exist, they don't cover the entire spectrum of inputs and/or uses. We introduce a new dataset, which contains RGB images of meals together with the corresponding depth maps, 3D models, segmentation and recognition labelled maps, weights and volumes. We also present a number of experiments on the new database to provide baselines performances in the context of food segmentation, depth and volume estimation.

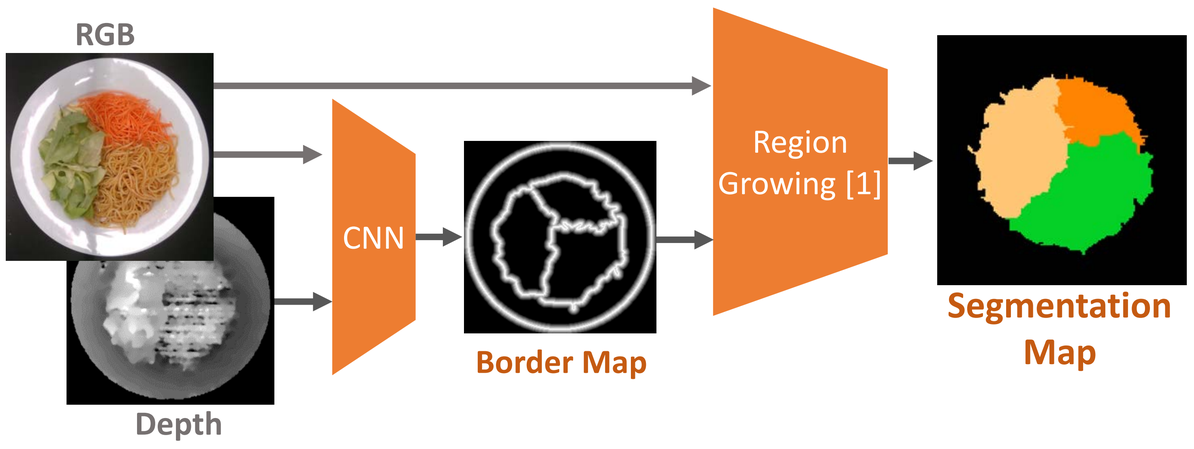
Dataset



Dataset's contents

Sensors	Setup	Images	Distance (cm)	Angle	Maps	
					Rec	Seg
Intel® RealSense™ Camera SR300	Constrained	8 RGB-D	40	0°	-	
				30°	-	
				60°	×	1
				90°	×	1
			60	0°	-	
				30°	-	
				60°	×	1
				90°	×	1
	Unconstrained	200 RGB-D	[40-60]	[45° -90°]	×	2
Samsung Galaxy S4 (using GoCARB)	Constrained	8 RGB	40	0°	-	
				30°	-	
				60°	×	1
				90°	×	1
			60	0°	-	
				30°	-	
				60°	×	1
				90°	×	1
	Unconstrained	6 RGB	[40-60]	[45° -90°]	×	1
LG Nexus	Unconstrained	~50 RGB	[40-60]	360° view	×	1
Total/ dish		**~272 (208 RGB-D; ~64 RGB)**			×	**12**
Total for the 80 dishes		**21807 (16640 RGB-D; 5167 RGB)**			×	**960**

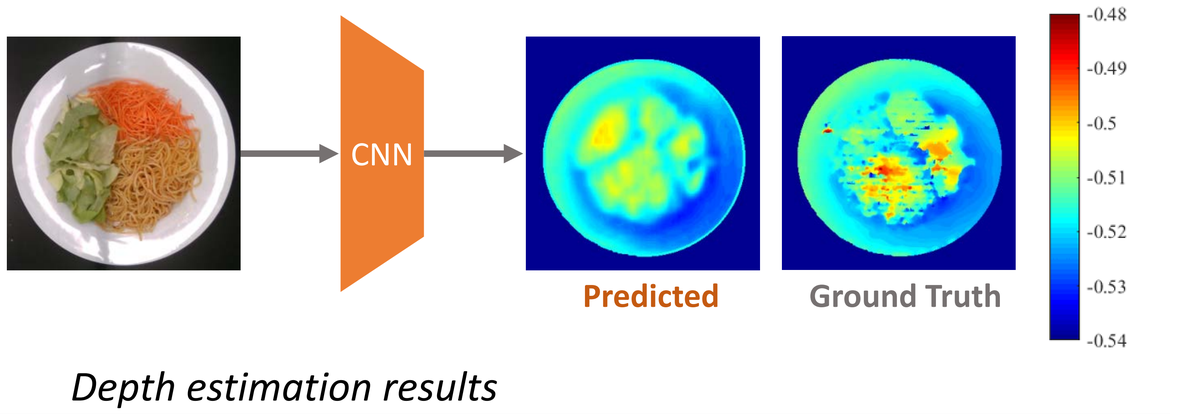
Segmentation



Segmentation results

CNN based on	Loss	AUG RGB-D	RGB		RGB-D	
			Min Fscore	Total Fscore	Min Fscore	Total Fscore
Segnet	MSE	No	0.7329	0.9326	0.7059	0.9288
Unet	MAE	No	0.6889	0.9268	0.7351	0.9342
Unet	MSE	No	0.6875	0.9247	0.7332	0.9328
Unet	MAE	Yes	0.6893	0.9281	**0.7426**	**0.9369**

Depth Estimation



Depth estimation results

CNN	Mean Absolute Difference - MAD (mm)	Absolute Relative Difference - ARD (%)
Eigen [2]	37.09	7.53
Proposed	**8.64**	**1.76**

Volume Estimation

We replace the depth estimation step as calculated in [3] (stereo reconstruction) with the depth obtained from the RGB-D images captured by depth sensor. In these conditions, the average error on the proposed dataset was **13.8%** using the method in [3], and **14%** using RGB-D images. Results indicate that a monocular RGB-D image can replace stereo pairs for volume estimation

References

[1] J. Dehais, M. Anthimopoulos, S. Mougiakakou. *"Food Image Segmentation for Dietary Assessment".* MADiMa, 2016

[2] D. Eigen, C. Puhrsch, R. Fergus, *"Depth Map Prediction from a Single Image using a Multi-Scale Deep Network".* NIPS, 2014

[3] J. Dehais, M. Anthimopoulos, S. Shevchik, S. Mougiakakou., *"Two-View 3D Reconstruction for Food Volume Estimation".* IEEE TMM, 2017.