Traditional programming to count bites automatically

Michele Tufano, Marlou Lasschuijt, Edith Feskens, Guido Camps

Wageningen University & Research
Human Nutrition and Health, Global Nutrition
Contact: michele.tufano@wur.nl
T + 39 334 74 09 720

Hungry Robot Lab
Contact: guido.camps@wur.nl
T + 31 (0)317 480 080
https://hungryrobots.guidocamps.com/

Background

Eating behavior is a key factor for weight management and eating disorders. The current method to study eating behavior is manual annotation of meal videos, which is time-consuming, laborious, lacks objectivity, and precludes real-time feedback. To replace manual annotation, we employ computer vision methods that can automatically analyze meal videos and predict the number of bites per meal.

Objective

We aim to use facial keypoints and traditional programming to automatically count the number of bites in a meal from video recordings.

Introduction

Facial keypoints (Fig.1) can localize and track points in the human face, body, and hands. We used traditional programming because it was an unexplored approach in the field. We developed algorithms that can calculate the distance between the lips and mouth ratio (light blue lines in Fig.1). The algorithms count a bite if the mouth ratio is higher than a set threshold. 12 algorithm variations (bitecounters 1-12) were tested for accuracy in 170 meal videos from 15 participants. Manual annotation of meal videos provided the ground truth.

Figure 1. The 468 facial keypoints (white) are applied on a face. The lips contour (pink) and their distance (light blue) are highlighted.

Algorithms

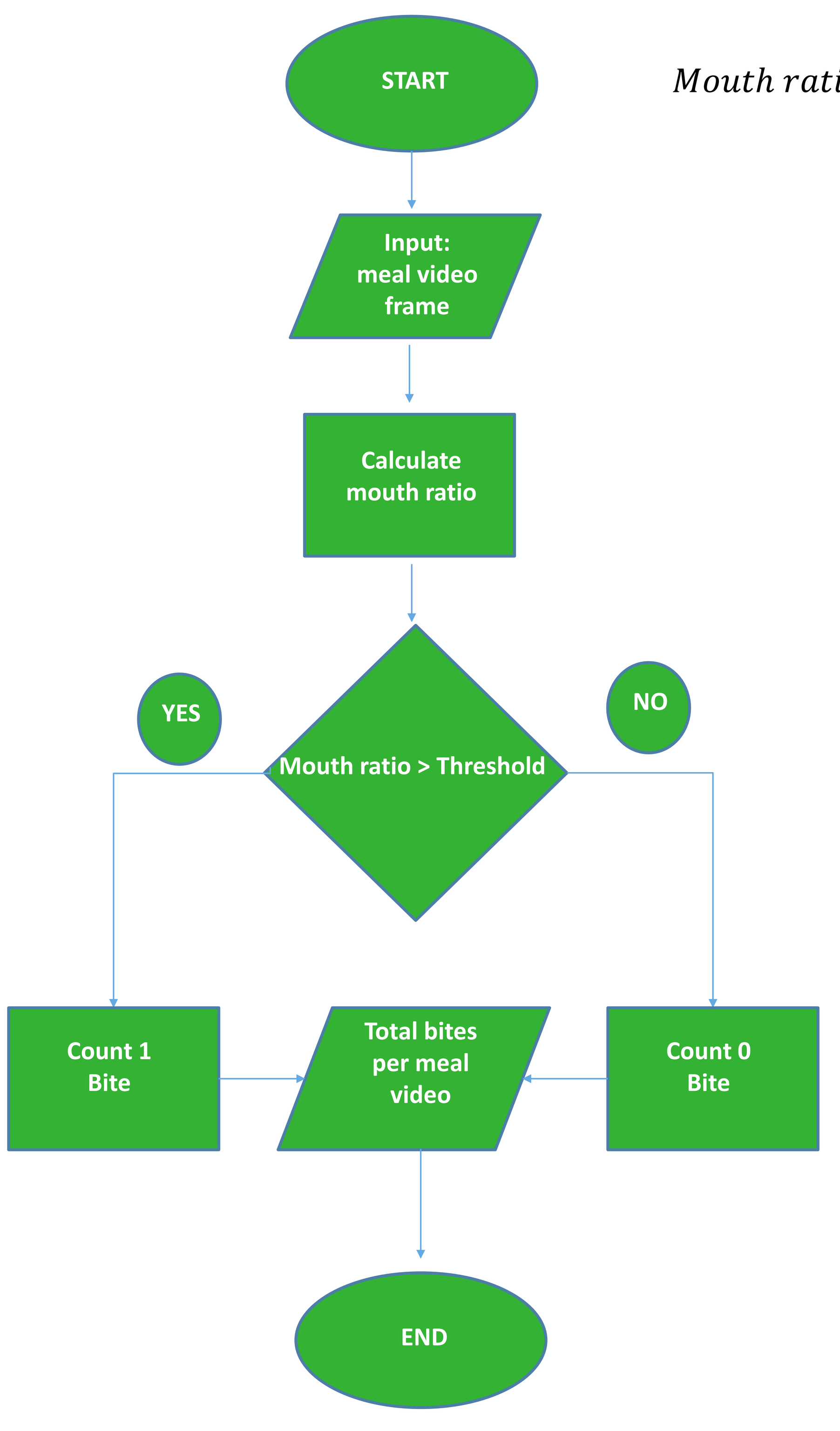

Figure 2. Flowchart of the bitecounter algorithm

$$\textit{Mouth ratio} = \frac{\textit{Distance between upper and lower lips}}{\textit{Distance between left and right side of the mouth}}$$

ALGORITHM	THRESHOLD
Bitecounter 1	60
Bitecounter 2	60 (forward); 75 (left, right); 65 (up); 30 (down)
Bitecounter 3	60 (forward); 75 (left, right); 65 (up)
Bitecounter 4	60 (forward); 75 (left, right); 65 (up); 40 (down)
Bitecounter 5	75 (left, right); 0.46 lips/face ratio (up, down, forward)
Bitecounter 6	60 (forward); 75 (left, right)
Bitecounter 7	60 (forward); 75 (left, right); 40 (down)
Bitecounter 8	60 (forward); 75 (left, right); 35 (down)
Bitecounter 9	30% increase from mouth ratio in the first video frame
Bitecounter 10	40% increase from mouth ratio in the first video frame
Bitecounter 11	50% increase from mouth ratio in the first video frame
Bitecounter 12	60% increase from mouth ratio in the first video frame

Table 1. Threshold used in the algorithms. Using a gaze estimator, threshold were set for the gaze direction (left, right, up, down, forward) for bitecounter2-8.

Results

Algorithm	Pearson r	P-value	Correct predictions (%)
Bitecounter 1	0.73	1.84E-26	4.7
Bitecounter 2	0.67	1.17E-20	3.36
Bitecounter 3	0.66	5.19E-20	4.03
Bitecounter 4	0.67	9.54E-21	4.03
Bitecounter 5	0.64	7.21E-19	6.04
Bitecounter 6	0.7	6.18E-23	4.7
Bitecounter 7	0.69	2.59E-22	4.03
Bitecounter 8	0.7	2.58E-23	5.37
Bitecounter 9	0.45	1.00E-08	0.67
Bitecounter 10	0.44	2.42E-08	2.01
Bitecounter 11	0.51	3.09E-11	2.68
Bitecounter 12	0.52	1.28E-11	2.68

Table 2. Pearson correlation coefficient r, relative p-value, and correct predictions (%) between each algorithm and the manual annotation.

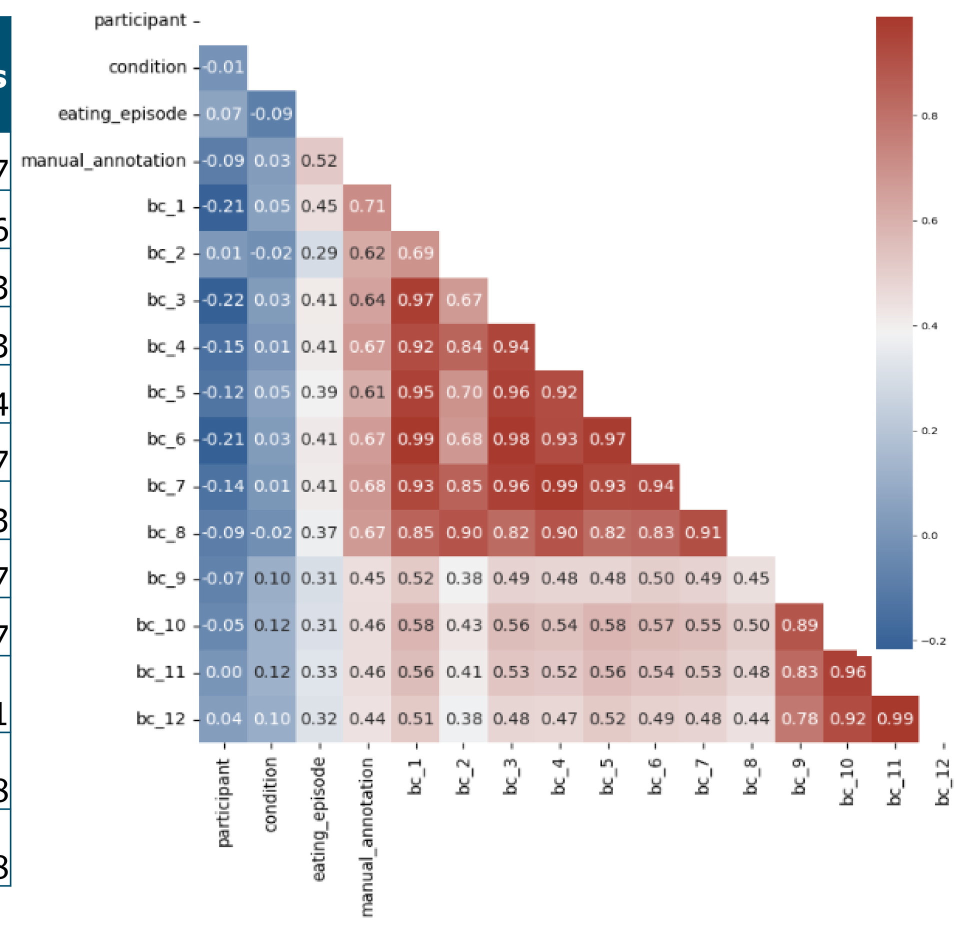

Figure 3. Heatmap of the analyzed dataset. bc_1-bc_12 correspond to the bitecounter algorithms.

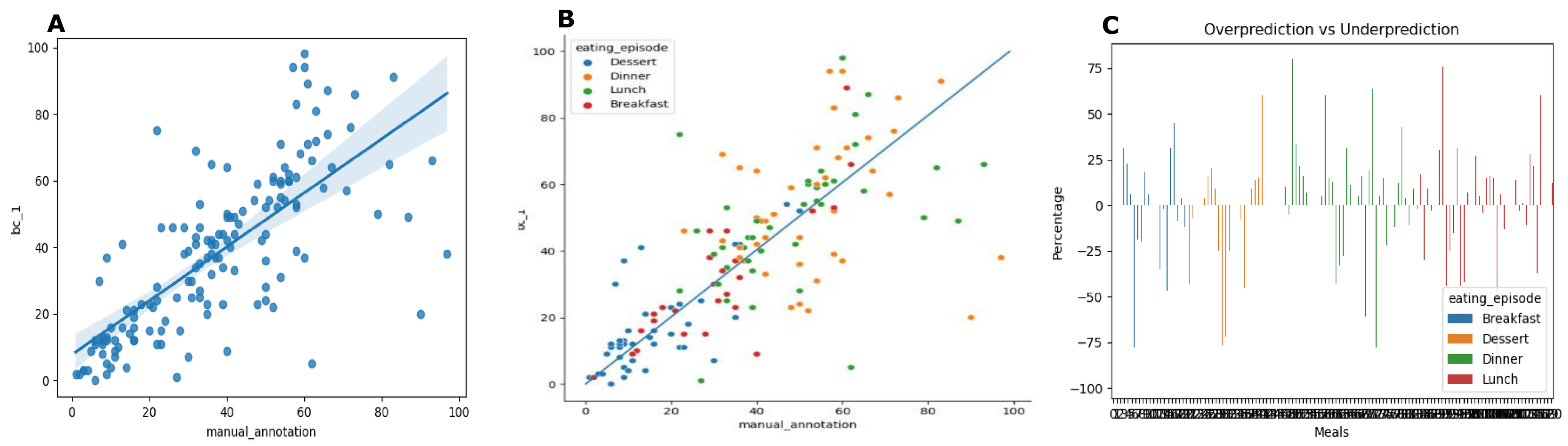

Figure 4. A) Regression plot between manual annotation (x) and bitecounter1 (y). **B)** Scatterplot between manual annotation (x) and bitecounter1 (y) with meals in color code. **C)** Percentage (x) of overpredicted (positive) and underpredicted (negative) bites for every meal (y) analyzed by bitecounter 1. For example, an underprediction of 20% corresponds to 80 bites predicted in a meal where 100 were observed. A percentage of 0 corresponds to a correct prediction.

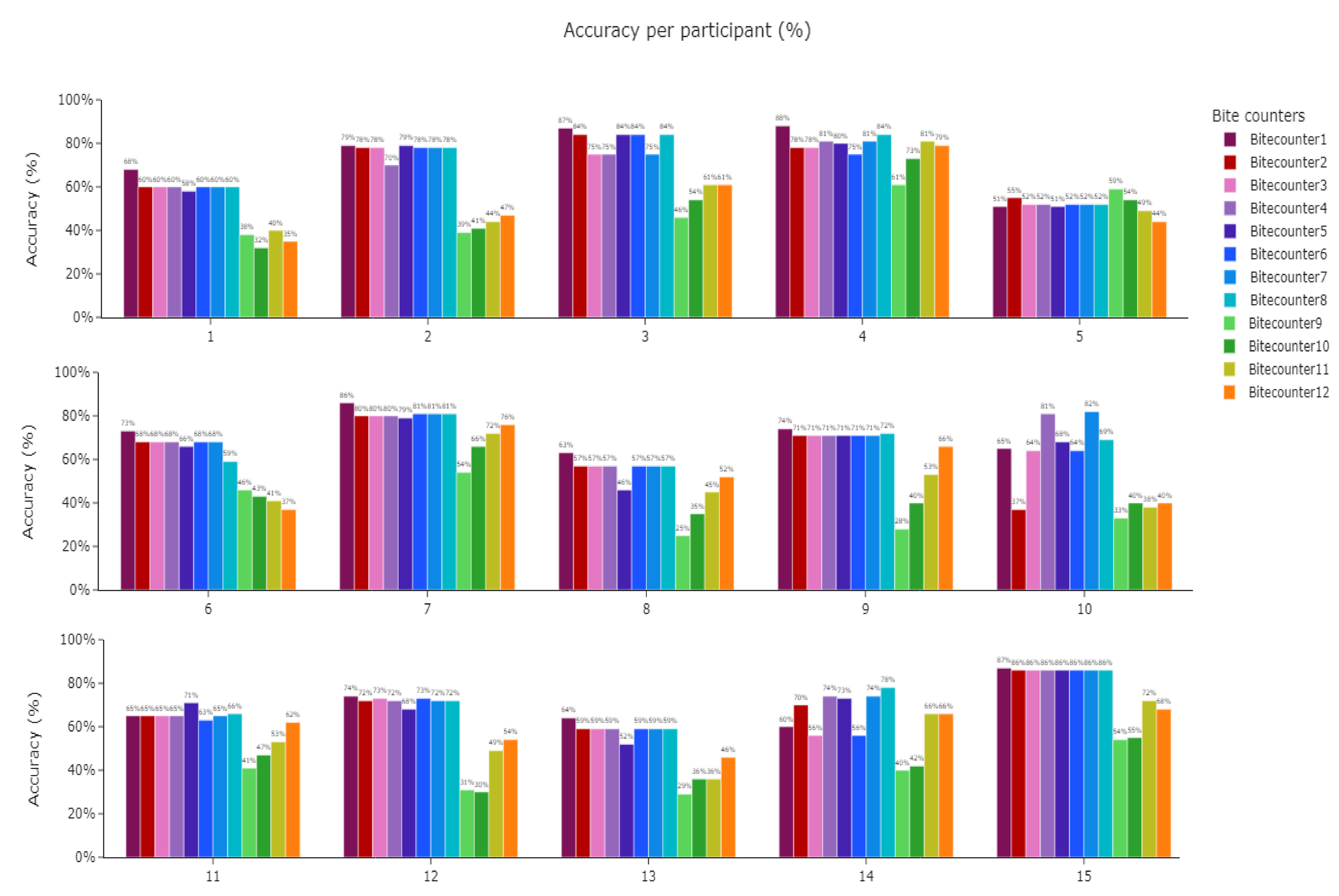

Figure 5. Accuracy (y) of the algorithms per participant (x).

Discussion

Traditional programming is not accurate enough to replace the manual annotation because:

- Correct bite predictions per videos are too low (~5%)
- Although a small error in bite predictions could be acceptable, the percentage of over/underprediction is too high (±75%) especially in meals with higher number of bites (i.e., dinner and lunch)
- The algorithms do not adapt to different faces, as shown by the differences in accuracy per participant

Machine learning models are required to increase accuracy in bite predictions