Automatic Cooking of Porridge Based on Human Visual Perception

Yujuan Sun
School of Information Science and Engineering, Ocean University of China, Qingdao, China,266100

School of Information and Electrical Engineering, Ludong University, Yantai, China, 264025
syj_anne@163.com

Junyu Dong
School of Information Science and Engineering, Ocean University of China, Qingdao, China.266100
dongjunyu@ouc.edu.cn

Lin Qi, Mike Chantler
Heriot-Watt University
Edinburgh,UK lq12@hw.ac.uk
M.J.Chantler@hw.ac.uk

ABSTRACT
When cooking porridge, humans can easily spot how vigorously the porridge is boiled and accordingly adjust the cooking power. The change of the porridge surface appearance can be regarded as dynamic textures. This paper studies the correlation between several computational descriptors and human perception using image sequences of boiling porridges. By comparing the chosen features with the results of psychophysical experiments, we found that the texture directionality descriptor proposed by Tamura et al and the inverse difference moment descriptor proposed by Haralick et al best coincide with human's perception. Our findings can be further used to judge how vigorously the porridge is boiled and accordingly adjust the cooking power in automatic cooking systems.

Keywords
Porridge, dynamic texture, texture descriptor, directionality, inverse difference moment

1. INTRODUCTION
Automatic control for cooking porridge can be improved by involving finding how vigorous the boil is based on human visual perception. Since the changes of the porridge surface appearance can be regarded as dynamic textures, texture features can be utilized, such as those introduced by Tamura et al [1] and Haralick et al[2], which were perception based and computational based respectively. In [3], an image model with a new set of features that aims to address the challenge of perceptual similarity was presented, and three perceptual properties i.e. periodicity, directionality and randomness were extracted. Some literature has proposed texture features which are robust to scale and rotation changes, general geometrical transforms and illumination variations [4]. The abundant textural features from co-occurrence matrix approach can be reduced to the only significant component [5]. Further, machine learning techniques such as artificial neural network and fuzzy c-means were used with a new rotation- invariant and scale-invariant image descriptor based on steerable pyramid decomposition [6].

In order to realize an automatic cooking system in terms of the specific type of food, this paper studies how computational features from porridge surface texture are related with human's visual perception. During cooking process, the continuous changing of the porridge surface texture can be easily observed by human. However, whether the computational descriptor coincides with the judgment of human's visual perception has not been well investigated. The visual cues perceived by human or the texture descriptors that can well reflect human's visual perception is studied in this paper. We study the correlation between human visual perception and eight texture descriptors, i.e. three primary texture descriptors from Tamura et al [1] and another five descriptors from Haralick et al [2], using image sequence captured during porridge cooking. We evaluated these descriptors to find out the most consistent ones with human's perception. The derived descriptor can be used to judge how vigorously the porridge is boiled and then to control the cook power.

Predicting Perceptions: The 3rd International Conference on Appearance, 17-19 April, 2012, Edinburgh, UK. Conference Proceedings Publication ISBN: 978-1-4716-6869-2, Pages: 104-107

2. DESIGH OF THE PSYCHOPHYSICAL EXPERIMENTS
The surface of porridge being cooked was captured as texture images. During a cooking process, porridge surface exhibits obvious textural change. In our experiments, a camera was placed above the pot. In order to prevent steam from covering the camera lens, the camera was installed at a position with a slant angle of 30 to 40 degrees above the pot. The camera captured images of the porridge being cooked in a fixed time interval δ. To facilitate real- time computation, we set δ=1 second. The image resolution was set to 400 * 600dpi, which is sufficient for capturing the appearance change. Figure 1 shown, from left to right and top to bottom, the gradual change of one kind of porridge surface appearance with cooking time. The images shown in figure 1 were the samples of the captured image sequence.

For each image (frame) in Fig 1, Tamura's three descriptors are calculated and the results are shown in Fig 2. The Haralick's first five descriptors are also calculated and the results are shown in Figure 3. We can see that the values of some descriptors, such as coarseness, directionality, ASM, correlation and inverse difference moment (IDM), change relatively slow, whereas the curves of the other descriptors show step changes. Since we know that once the porridge is in the boiling stage, its surface should not changerapidly, we select coarseness, directionality, ASM, correlation and IDM as candidate descriptors.

Our purpose is to test for correlations between frame-to-frame differences in human judgments of boiling vigor and values of feature descriptors. In order to find out which candidate descriptors are more consistent with the human visual perception, we conducted a set of psychophysical experiments, which are designed as follows: images of five kinds of porridge captured during the cooking process were shown to observers in five trial groups according to the type of porridge. Ten observers participated in our experiment, each of whom needs to complete two tasks. The first task is judging the boiling-point. Participants judged whether the porridge is in the boiling state or not from the given images. The order of the images in each group was randomized using Latin squares. If each group has 25 images, a Latin square is a 25×25 array filled with 25 different symbols. All participants performed the first task twice in random image sequence, so we need 20 random sequences. The first 20 rows of Latin square were selected. Thus 20 judgments were produced for each image. For each sequence, the image with the minimum distance to the 50th percentile was regarded as the reference of boiling point.

The second task is to estimate the difference of the vigor of the boil between adjacent frames in each group. If each group has 25 images, then there will be 24 pairs of adjacent frames, i.e. (1,2),(2,3),(3,4)…(24,25). Each pair was repeated twice; thus each image group produces 48 trials. At each trial the participant was given one pair of images. If the difference of the vigor of the boil in a pair is large, then a higher score should be given; otherwise a lower score should be provided. There are no restrictions on the range of scores. For each group, a total of 10*2*24 comparative scores were produced. The scores of each group were normalized before averaging across observers.

3. RESULTS

In the first task the boiling point was observed at the 14^{th},15^{th},13^{th},16^{th} and 12^{th} frame respectively for each kind of porridge.

Since the boiling point is the inflection point in the feature curves, the perceived data and the calculated data are the negative correlation at the left side of the boiling point and the positive correlation at the right side of boiling point. In order to ensure the consistency of the correlation, we performed incremental integration at the left side of the boiling point and decremental integration at the right side of the boiling point, deriving curves of human perceived information. We compared the human perceived data with the value of candidate descriptors. The correlation coefficients were computed between the values of the descriptors and the perceived data for five kinds of porridges. The results are shown in Table 1.

Table 1. The correlation coefficients between different descriptors and perceived scales.(*:p<0.05,:p<0.01))**

	directionality	IDM	coarseness	ASM	correaltion
Babao	0.8084**	0.7997**	0.5974**	0.7059**	0.7430**
Dacha	0.6050**	0.5784**	0.6823**	-0.1123	0.6876**
Rice	0.6812**	0.6946**	-0.0223	0.7514**	0.5198**
Oatmeal	0.9315**	0.9480**	-0.0027	0.7708**	0.7443**
Millet	0.6349**	0.6453**	0.4332*	0.7295**	0.4583*

Larger coefficients mean that the behavior of the corresponding descriptors is more consistent with that of human visual perception.

We find from Table 1 that the correlation coefficients of directionality and the IDM descriptors are larger than others. Then these descriptors may be selected as the control signal for automatic cooking of porridge.

According to the boiling point detected in the first task, we can set the threshold value for selected descriptors to control the cooking power. The result is shown in Fig 4. We compare the result of first task with the actual images in Fig 1, and find that the corresponding image is exactly in the boiling state. In practice, we may turn off or reduce the cooking power at certain time after the boiling point. From the descriptors' aspect, when the highest point is detected in the descriptor curve, we give the control signal after the descriptor has dropped down for specific threshold. We conservatively set the threshold to 0.4 in our experiments and got satisfactory cooking results. In Fig 4, the asterisks shows the time of sending the control signal.

4. CONCLUSIONS

Based on the principle of human visual perception, an automatic power control method for cooking porridge is presented. A camera is installed above the cooking pot, and in a fixed time interval porridge being cooked is filmed. Then the values of the texture descriptors are calculated and analyzed to decide whether the changes of the curve are adapt to the requirements of the threshold. If the change meets the requirement, the cooking power can be reduced. The scheme is reliable, accurate and simple for implementation. An integrated device may be developed based on the scheme for fully automatic cooking of porridge without human supervision.

5. REFERENCES

[1] Hideyuki Tamura, Shunjimori, and Takashi Yamawaki, J. 1978. Textural Features Corresponding to Visual Perception.IEEE Transactions on Systems, Man and Cybernetics, 8(6): 460 – 473.
[2]Robert M.Haralick,K, Shanmugam, and ITS'Hak Dinstein, J. 1973. Textural features for image classification. IEEE Transactions on Systems,man and cybernetics, 3(6):610-621.
[3]F.Liu, R.W. Picard, J. 1996. Periodiity, directionality, and randomness: Wold features for image modeling and retrieval. IEEE Tansactions on PAMI, 18(7):722-733.
[4]Yong Xu, Xiong Yang, Haibin Ling, and Hui Ji, J.A. 2010. A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid.IEEE Conf. on Computer Vision and Pattern Recognition (**CVPR**), San Francisco.
[5]Jayashri Joshi, Mrs.A.C. Phadke, J.A.2010. Feature Extraction and Texture Classification in MRI. International Conference on ICCT 2010.
[6] Javier A. Montoya-Zegarra, Joao Paulo Papa, Neucimar, J. Leite, J.2008. Learning How to Extract Rotation-Invariant and Scale-Invariant Features from Texture Images, EURASIP Journal on Advances in Signal Processing. doi:10.1155/2008/691924.

Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance, pp. 104-107, Edinburgh, UK, ISBN 978-1-4716-6869-2, April 2012.

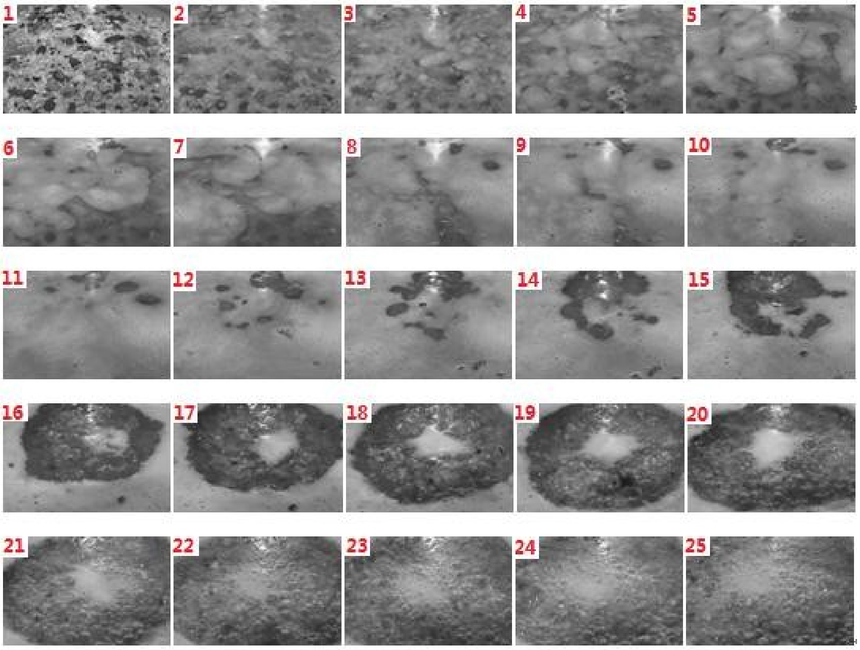

Figure 1. The process of cooking porridge

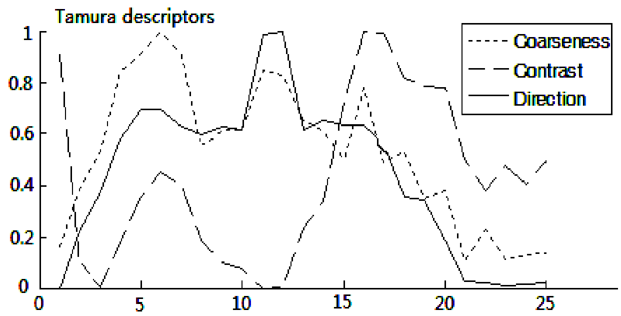

Figure 2. Tamura's descriptors calculated from the boiling porridge

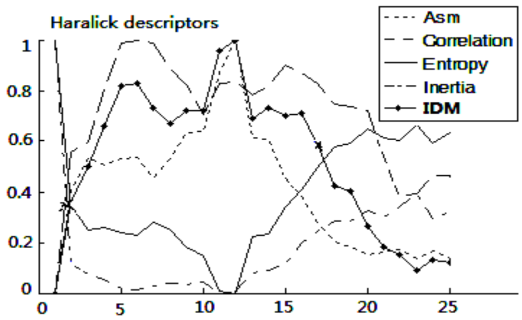

Figure 3. Haralick's descriptors calculated from the boiling porridge

(In Fig 2 and Fig 3, the horizontal axis is the frame label, the vertical axis represents descriptor values of the corresponding images. The descriptor values have been scaled to the range between zero and one.)

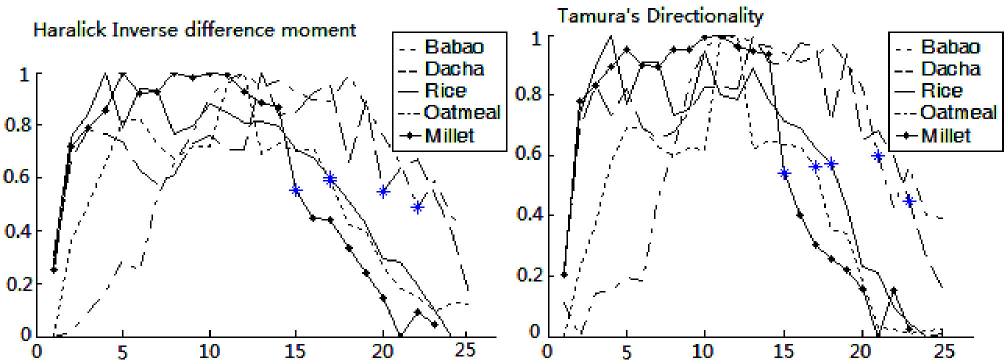

Figure 4. The curve of Haralick's inverse difference moment and Tamura's directionality for five kinds of porridges (Blue asterisk represent that the descriptor has dropped down for specific threshold(0.4 in our experiments)after detecting the highest point)