Simple Real-time Facial Expression Recognition for Interactive Arts

저자 (Authors)	Seong-Gi Oh, Chang-Young Jang, Tae-Yong Kim
출처 (Source)	TECHART: Journal of Arts and Imaging Science 6(3), 2019.8, 17-20(4 pages)
발행처 (Publisher)	중앙대학교 첨단영상대학원 TECHART Journal of Arts and Imaging Science
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE08771933
APA Style	Seong-Gi Oh, Chang-Young Jang, Tae-Yong Kim (2019). Simple Real-time Facial Expression Recognition for Interactive Arts. TECHART: Journal of Arts and Imaging Science, 6(3), 17-20
이용정보 (Accessed)	중앙대학교 219.255.207.*** 2019/12/09 15:55 (KST)

Simple Real-time Facial Expression Recognition for Interactive Arts

Seong-Gi Oh[1], Chang-Young Jang, and Tae-Yong Kim[2]*

[1] Chung-Ang University / Seoul, Korea
[2] Chung-Ang University / Seoul, Korea
**Corresponding Author (kimty@cau.ac.kr)*

Abstract: In this paper, technical parts of our emotion recognition based Interactive Art are introduced. Simple multiclass SVM model based on ECOC designed for facial expression recognition. 4 methods by 2 features and 2 classifiers are tested for the best pipeline for real-time. 4 case's accuracy after number of features(cell size) is tested but it has each appropriate rate instead of proportionality. Comparison of 2 vector features (HOG, LBP) for extracting facial features, HOG shows better accuracy. In other comparison of 2 classifier(ECOC, KNN), ECOC shows better accuracy. The fastest pipeline of 4 methods is HOG + ECOC with 60.368 calculation/s score and best accuracy(97.44%).

Keywords: Emotion Recognition, Image processing, Interactive art.

Received May. 23, 2019; accepted for publication Aug. 9, 2019; published online Aug. 31, 2019. DOI: 10.15323/techart.2019.8.6.3.17 / ISSN: 2288-9248.

1. Introduction

Interactive art has own special functions, works in real time and provide a better immersion experience than others. It performed by desktop computers, laptops or embedded systems; Arduino, Raspberry Pi, Intel Edison presently. These systems have lower cost but better accessibility to sensors and ease of build pipeline than desktop systems. However, due to architectural limitations of the mobile processor and low memory capacity, it is difficult to process high-level functions such as 3d rendering, complexed neural networks etc. in real-time. This problem also occurred on common computer systems with not high performance. Therefore, even if the same function is performed, a method that is structurally simple and has a low processing time is better for interactive art. In this paper, we compare facial expression detection methods and proposes a pipeline suitable for real-time processing.

2. Related work

P Michel et al. [1] suggested real time model using support vector machines in 2003 paper. This pipeline consists of recognition + features tracking. The reason why using tracking is lack of computer power in 2003s. Nowadays, tracking steps for classification is no more needed in single face case because it makes pipeline faster but accuracy lower by dependency of prediction between previous and present features on images. R Vedantham et al. [2] proposed a pipeline for real-time facial expression recognition using Garbor filter + KNN method. This method shows better accuracy (96.67%) than P Michel et al. (60.7%) even its pipeline is simpler. I Kotsia et al. [3] multiclass SVM model for facial expression is known as most accurate (99.7%) but its pipeline is too complexed for real-time processing. Its real-time model suggested by Vinícius S et al. [4] with lower accuracy (93.6%) using depth camera input.

3. Pipeline Design

In this section, details of our 3-step pipeline are introduced. These steps consist of face detection, facial feature extraction and building expression classifier.

3.1 Object Detection Methods

The most popular method for face detection is Ada-boost of binary features proposed by Viola-Jones [5]. Convolution Neural Network using ImageNet [6] is strong method also but requires high-level processing system. In this paper, Viola-Jones method is selected because prefer to make the pipeline without a Neural Network.

A. Viola-Jones method for face detection

The first step of our pipeline is detecting face regions to extracting facial feature. Viola-Jones method using binary features is one of the fastest and simplest detectors to find objects in an image

B. K- Nearest Neighbor (KNN) Classifier [7]

KNN matching inputs to most related existing data and classifying its class without pre-training. It is the simplest and basic classifier for feature matching. Since it doesn't need training step, effective method when training data is mutating in real-time such as reinforcement learning.

C. Support Vector Machine (SVM) [8]

SVM is the most used to training vector features. It has better sensitivity to images containing moving objects comparing with Viola-Jones. However, it can be slower than Viola-Jones method because vector feature has much more elements than binary features. In the multi-class case, it needs more pipeline (LIBSVM etc.) because performs binary classification.

D. Multi-model SVM; Error Correcting Output Codes (ECOC) [9]

ECOC is the simplest method for multi-model SVM. We build 15-bit coding for 7 emotion classes consist of following Table. 1

Table. 1
Proposed bits of ECOC

Emotion Class	Code Word														
	f0	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14
Neutral	1	1	0	0	0	0	1	0	1	0	0	1	1	0	1
	0	0	1	1	1	1	0	1	0	1	1	0	0	1	0
Disgusted	1	0	0	1	0	0	0	1	1	1	1	0	1	0	1
Fear	0	0	1	1	0	1	1	1	0	0	0	0	1	0	1
Sad	1	1	1	0	1	0	1	1	0	0	1	0	0	0	1
Surprised	0	1	0	0	1	1	0	1	1	1	0	0	0	0	1
Happy	1	0	1	1	1	0	0	0	0	1	0	1	0	0	1
Angry	0	0	0	1	1	1	1	0	1	0	1	1	0	0	1

Since the base facial expression is Neutral, double classes are taken to Neutral emotion for perfect opposite case of each bits.

3.2 Facial Expression Features

We use following 2 features because can find different between 2 classifier SVM and KNN.

A. Histogram of Gradients (HOG) [10]

Through edge image input, vector features are obtained through the difference between each region centers and edges. Feature extractor consisted of blocks, cells, and overlapped blocks. Cells consists of pixels and block consists of cells. Each block has one feature point, and the block created by overlap also gets one feature point. Each feature point contains un-normalized vectors by the commanded number of directions, and each vector represents the magnitude of the gradients for each direction.

B. Local Binary Pattern (LBP) [11]

LBP obtaining vector features through each region level property from Gray level (N x N x 1) image input. Same with HOG, it consists of blocks, cells, and overlapped blocks. Each vector represents the magnitude of the gradients obtained through the level difference with that pixel relative to the region center.

3.3 Datasets for Facial Expression

Following 4 datasets have 7 classes of basic emotions; Neutral, Angry, Disgusted, Fear, Happy, Sad, Surprised

A. Japanese Female Facial Expression JAFFE [12]

JAFFE contains 192 training images and 27 test images of 7 facial expressions posed by 10 Japanese female models. Source URL is: http://www.kasrl.org/jaffe.html

B. Karolinska Directed Emotional Faces KDEF [13]

KDEF contains 967 train images and 138 test images of 7 facial expressions from 5 different angles. Source URL is: http://www.emotionlab.se/kdef/

C. Multimedia Understanding Group MUG [14]

MUG contains 1746 train images and 255 test images of 7 facial expressions form 35 women and 51 men all of Caucasian origin between 20 and 35 years of age. Source URL is: https://mug.ee.auth.gr/fed/

D. Warsaw Set of Emotional Facial Expression Pictures WSEFEP [15]

WSEFEP contains 204 train images and 29 test images of 7 facial expressions form 14 men and 16 women. Source URL is: http://www.emotional-face.org/

4. Experiments and Results

We compared 2 vector features (HOG, LBP) with KNN (non-trained classifier) and SVM. Used database is 100x100 regularized image set of 4 datasets.

4.1 Features & Classifiers comparison

4 pipelines using HOG and LBP features, KNN and ECOC classifier are tested by different cell sizes (2, 4, 6, …, 32) because cell size is critical variance for the numbers of vectors. Overlap rate of 2 features is 0.5 of block size (2 x 2) same with 2 extractors. All tests performed in MATLAB 2018b using single thread of i7-8750h (2.20Ghz) processor without GPU assistant and multi-threading.

A. LBP - KNN

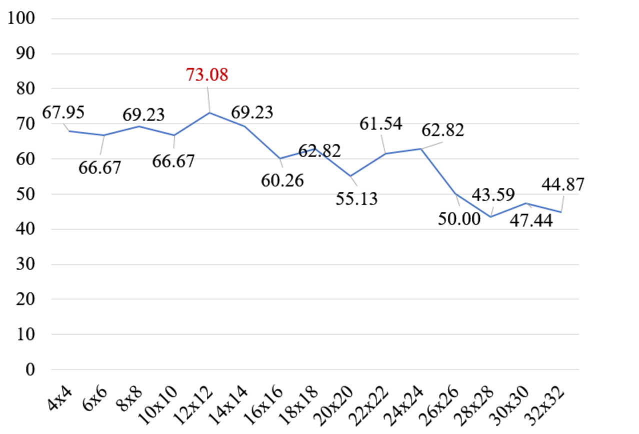

Fig. 1. Accuracy of KNN with LBP feature 4~32 cell size range.

In Figure.1, 12 x 12 is the best cell size for accuracy with 73.08% Accuracy

B. LBP - ECOC

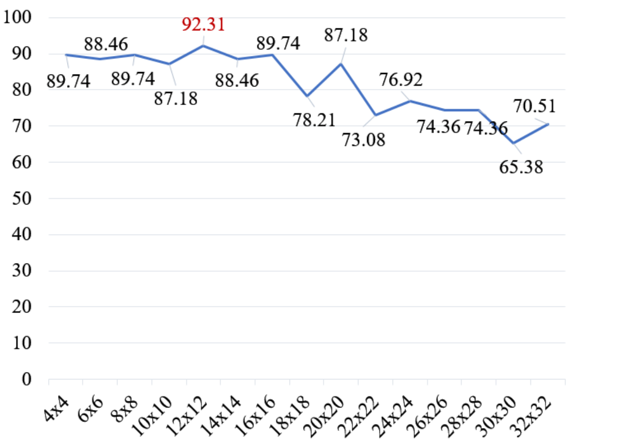

Fig. 2. Accuracy of ECOC with LBP feature 4~32 cell size range.

In Figure.2, 12 x 12 is the best cell size for accuracy with 92.31% Accuracy

C. HOG - KNN

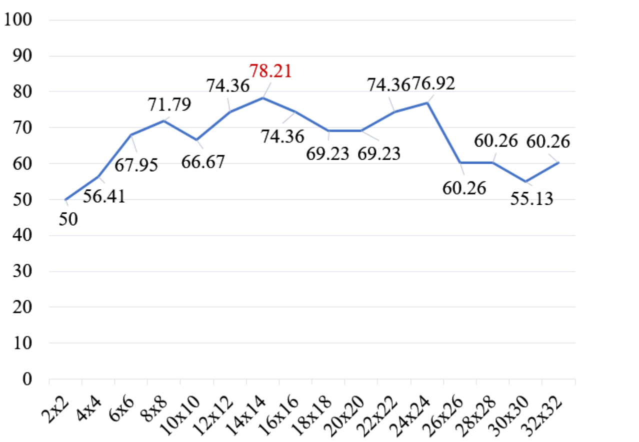

Fig. 3. Accuracy of KNN with HOG feature 2~32 cell size range.

In Figure.3, 14 x 14 is the best cell size for accuracy with 78.21% Accuracy

D. HOG - ECOC

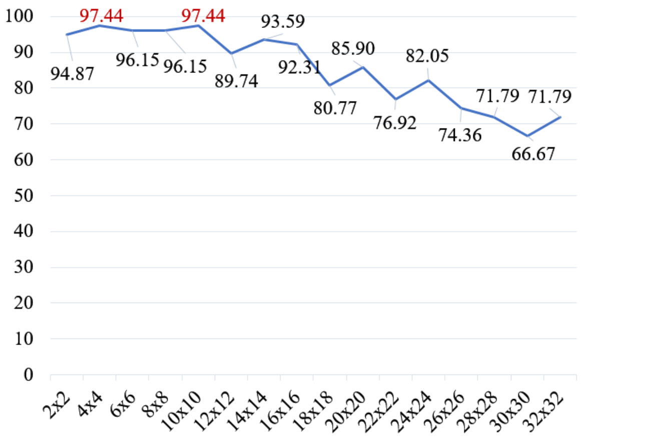

Fig. 4. Accuracy of ECOC with HOG feature 2~32 cell size range.

In Figure.4, 10 x 10 is the best cell size for accuracy with 97.44% Accuracy

4.2 Summary of Results

Table. 2
Summary table

Feature	Classifier	Max Accuracy (%)	Cell size for the best	Feature extraction (ms)	Total time (ms)	Detection/s
HOG	KNN	78.21	14 x 14	3.014	23.748	42.109
HOG	ECOC	97.44	10 x 10	3.366	16.565	60.368
LBP	KNN	73.08	12 x 12	3.290	48.101	20.790
LBP	ECOC	92.31	12 x 12		16.893	59.196

Table.2 shows maximum accuracy and its cell size. Total time consists of sum that feature extraction time + classifying time. Detection/s score is 1000 (1s)/Total time and shows how many processing can be repeated in a second.

- Best Feature: HOG
- Best Classifier: ECOC
- Best Method: 97.44% using ECOC & HOG, 10x10 cell size

We can see training accuracy is going lower when away from fit cell size. But training and classifying is going faster with lager cell size because the number of features getting less counts. In Table. 2, 10x10 is better cell size for feature extraction because 10x10 is faster than 4x4 with same accuracy. ECOC has better accuracy and less time cost than KNN classifier. The minimum Detection/s score for real-time processing is over than 30 frames per seconds. Except for LBP–KNN, all methods satisfying real-time requirement.

5. Conclusion

The best pipeline is trained 15bit ECOC classifier with 10x10 cell sized HOG feature. It shows 97.44% classification accuracy with 60.368 fps that double score of real-time. The number of features has irregular relation with accuracy in results Table.2 by cell sizes. Cell size near 0.1 rate of image size is effective to accuracy of classification and time cost.

Acknowledgement

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF2018R1D1A1B 07044286) and BK21 plus.

References

[1] Michel, P., & El Kaliouby, R. (2003, November). Real time facial expression recognition in video using support vector machines. In *Proceedings of the 5th international conference on Multimodal interfaces* (pp. 258-264). ACM.

[2] Ramachandran V., Lavanya S., & E.Sreenivasa R. (2018). Real Time Facial Expression Recognition in Video Using Nearest Neighbor Classifier. *International Journal of Pure and Applied Mathematics*. (Vol. 118, No.9, pp. 849-854).

[3] Kotsia, I., & Pitas, I. (2006). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE transactions on image processing*, *16*(1), 172-187.

[4] Silva, V., Soares, F., Esteves, J. S., Figueiredo, J., Leão, C. P., Santos, C., & Pereira, A. P. (2016, October). Real-time emotions recognition system. In *2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (pp. 201-206). IEEE.

[5] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

[7] Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

[8] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, *2*(2), 121-167.

[9] Dietterich, T. G., & Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of artificial intelligence research*, *2*, 263-286.

[10] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *international Conference on computer vision & Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE Computer Society.

[11] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12), 2037-2041.

[12] Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J., & Budynek, J. (1998, April). The Japanese female facial expression (JAFFE) database. *In Proceedings of third international conference on automatic face and gesture recognition* (pp. 14-16).

[13] Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, *CD ROM from Department of Clinical Neuroscience, Psychology section*, Karolinska Institutet, ISBN 91-630-7164-9.

[14] N. Aifanti, C. Papachristou and A. Delopoulos, The MUG Facial Expression Database, *in Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Desenzano, Italy, April 12-14 2010.

[15] Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., & Ohme, R. K. (2015). Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. *Frontiers in psychology*, 5, 1516.

Biographies

First author Seong-Gi Oh is an M.S. student at Chung-Ang University Graduate School of Advanced Imaging Science and Multimedia. He is a Game Technology and Application (GTA) Lab student. In 2018, he received a BS degree from ChungAng University; he is interested in Artificial Intelligence using Image Processing.

Corresponding author Tae-Yong Kim is an professor at Chung-Ang University Graduate School of Advanced Imaging Science and Multimedia. He is a chief at Game Technology and Application (GTA) Lab. In 1998, he received a doctor degree from POSTECH; he is interested in Computer Vision and Game Science

Second author Chang-Young Jang is an M.S. student at Chung-Ang University Graduate School of Advanced Imaging Science and Multimedia. He is a Game Technology and Application (GTA) Lab student. In 2018, he received a BS degree from SangMyung University; he is interested in Gesture recognition.