Unsupervised Estimation of Facial Expression Intensity for Emotional Scene Retrieval in Lifelog Videos

Shota Sakaue, Kyoto Institute of Technology, Kyoto, Japan
Hiroki Nomiya, Kyoto Institute of Technology, Kyoto, Japan
Teruhisa Hochin, Kyoto Institute of Technology, Kyoto, Japan

ABSTRACT

This article describes how in order to facilitate the retrieval of impressive scenes from lifelog videos, a method to estimate the intensity of a facial expression of a person in a lifelog video is proposed. The previous work made it possible to estimate the facial expression intensity, but the previous method requires some training samples which should be manually and carefully selected. This makes the previous method quite inconvenient. This article attempts to solve this problem by introducing an unsupervised learning method. The proposed method estimates the facial expression intensity via a clustering on the basis of several facial features computed from the positional relationship of a number of facial feature points. For the evaluation of the proposed method, an experiment to estimate the facial expression intensity is performed using a lifelog video data set. The estimation performance of the proposed method is compared with that of the previous method.

KEYWORDS

Clustering, Facial Expression Intensity, Facial Expression Recognition, Facial Feature Points, Fuzzy c-Means, Lifelog Video, Unsupervised Learning, Video Retrieval

1. INTRODUCTION

Recent improvement of multimedia recording devices enables us to easily create a large amount of the records of daily life. The records are known as lifelog (Aizawa, 2004) (Aizawa, Hori, Kawasaki, & Ishikawa, 2004) (Gemmell et al., 2002). Lifelog is stored as various types of data such as text, image, and video. Particularly, video data will be widely used because they can be easily created using, for example, smartphones and contain a variety of useful information. Therefore, we focus on lifelog videos in this study.

Due to the easiness of creating lifelog videos, a large amount of lifelog video data can be stored in a video database. This makes the retrieval of lifelog videos quite difficult. As a result, a considerable amount of valuable lifelog video data will not be utilized. Therefore, an efficient and accurate retrieval method for lifelog videos is needed.

DOI: 10.4018/IJSI.2018100103

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
In order to facilitate the utilization of lifelog videos, we have proposed some methods to retrieve impressive scenes from lifelog videos (Nomiya & Hochin, 2013) (Nomiya, Morikuni, & Hochin, 2015) (Nomiya & Hochin, 2015). These methods detect impressive scenes from a lifelog video using facial expressions of a person in the video. They can find emotional scenes that a person expresses a certain facial expression. However, they are not able to estimate the intensity of the facial expression. For example, they can distinguish a smiling face and a neutral one, but cannot distinguish a smile and a big laugh.

Morikuni et al. attempted to estimate the intensity of a facial expression and proposed a measure called facial expression intensity (Morikuni, Nomiya, & Hochin, 2015). The facial expression intensity is computed using several salient points on a face called facial feature points. However, the preciseness of the facial expression intensity is insufficient. Moreover, they focused only on “happiness” and “surprise,” and did not provide any estimation methods for such as “anger” and “sadness,” which are relatively difficult to estimate the intensity.

Sakaue et al. ameliorated the accuracy of Morikuni’s facial expression intensity by adding some effective facial features and improving the computation method of the facial expression intensity (Sakaue, Nomiya, & Hochin, 2017). In addition, they evaluated their facial expression intensity using some data sets containing basic six facial expressions (anger, disgust, fear, happiness, sadness, and surprise). Sakaue’s facial expression intensity has, however, a crucial problem that several training samples are required to compute the facial expression intensity. The training samples should be manually and carefully selected because they affect the preciseness of the estimation of the facial expression intensity. Considering that lifelog videos are often managed by a few persons due to privacy issues, preparing sufficient training samples will be a considerable burden for the owners of lifelog videos.

In this paper, we propose a new estimation method of the facial expression intensity on the basis of a clustering technique. Since the clustering is a kind of unsupervised learning method, the proposed method does not require any training samples. We evaluate the proposed facial expression intensity using a lifelog video data set and a widely-used facial expression data set. Additionally, we show the fact through an experiment that the accuracy of the proposed facial expression intensity is comparable to Sakaue’s facial expression intensity.

The remainder of this paper is organized as follows. Section 2 introduces Sakaue’s facial expression intensity as the previous research. Section 3 explains the computation method of the proposed facial expression intensity. Section 4 describes the experiment to evaluate the proposed facial expression intensity. Section 5 gives a consideration to the result of experiment. Finally, Section 6 concludes this paper.

2. PREVIOUS RESEARCH

In this section, the estimation method of Sakaue’s facial expression intensity is explained as the proposed method of the previous research (Sakaue, Nomiya, & Hochin, 2017). The facial expression intensity of the previous research is computed based on the positional relationship of several facial feature points.

2.1. Facial Feature Points

The facial feature points are extracted by using Luxand FaceSDK 4.0 (Luxand Inc., 2011). The facial expression intensity is computed based on the positional relationship of the following fifty-four facial feature points. Figure 1 shows the facial feature points.

- Left and right eyebrows: 10 points \(P_{1}, \cdots, P_{10}\)
- Left and right eyes: 18 points \(P_{11}, \cdots, P_{28}\)
Backstepping Control for Synchronizing Fractional-Order Chaotic Systems
[www.igi-global.com/chapter/backstepping-control-for-synchronizing-fractional-order-chaotic-systems/204799?camid=4v1a](www.igi-global.com/chapter/backstepping-control-for-synchronizing-fractional-order-chaotic-systems/204799?camid=4v1a)

Network Support Data Analysis for Fault Identification Using Machine Learning