Computer Vision and Image Understanding 147 (2016) 102–113

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Crowdsourcing facial expressions for affective-interaction

Gonçalo Tavares, André Mourão, João Magalhães*

NOVA Laboratory for Computer Science and Informatics, NOVA-LINCS, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

ARTICLE INFO

Article history:
Received 31 March 2015
Accepted 3 February 2016

Keywords:
Affective-interaction
Realistic facial expressions
Crowdsourcing

ABSTRACT

Affective-interaction in computer games is a novel area with several new challenges, such as detecting players facial expressions robustly. Many of the existing facial expression datasets are composed of a set of posed face images not captured in a realistic affective-interaction setting. The contribution of this paper is an affective-interaction dataset captured while users were playing a game that reacted to their facial-expressions. This dataset was the result of a framework designed for gathering affective-interaction data and annotating this data with high-quality labels. The first part of the framework is a computer game [15] planned to elicit a particular facial expressions that directly control the game outcome. Thus, the game creates a true and engaging affective-interaction scenario where facial-expressions data were captured. The proposed dataset is composed of a series of sequential video frames where faces were detected while users interacted with a game with their facial expressions. The second part of the framework is a crowdsourcing process designed to ask annotators to identify the facial-expression present in a given face image. Each face image was annotated with a facial-expression: *happy*, *anger*, *disgust*, *contempt*, *sad*, *fear*, *surprise*, and *neutral*. We examined how the annotators performance was affected by multiple variables, e.g., reward, judgment limits, golden questions. Once these parameters were tuned, we gathered 229,584 annotations for the whole 42,911 images. Statistical consensus techniques were then used to merge the annotators judgments and produce high-quality image-labels. Finally, we compared different classifiers trained on both ground-truth (*expert*) labels and crowdsourcing labels: we observed no differences in classification accuracy, which confirms the quality of the produced labels. Thus, we conclude that the proposed affective-interaction dataset provides a unique set of images of people playing games with their facial expressions and labels with a quality similar to that of expert labels (differences are less than 9%).

1. Introduction

Novel input techniques enabling human–computer interaction using the user body are becoming increasingly popular. Microsoft Kinect introduced full body gesture-based interaction to the mainstream public. With Microsoft Kinect, users can play games with their body by performing actions like jumping, boxing, amongst others. Although the body movements and gestures provide a rich source of input, they can be augmented with affective-interaction through the recognition of a player's facial expression for a more natural interaction. Fig. 1a shows the same player performing the same action, *a punch*, but with different facial expressions: a *neutral* and an *angry* expression. An affective-interaction aware application would consider both the body-motion and the facial expression when scoring or recognizing actions, giving a higher score to the *punch* action on the right of Fig. 1a.

Computer games are a primary example of where computational actions can be adjusted to the player's facial expression, however, interest in facial expressions has moved far beyond computer vision and interaction research, reaching areas like media consumption or health applications [14,20]. Mcduff el al. [14] showed the effectiveness of a smile for rating videos; they developed a video recommendation system where viewers' facial expressions are used to rate videos. Moreover, their experiments were validated with manual affective-feedback data collected through crowdsourcing. These applications require a realistic set of annotated images for researchers to improve current algorithms in the context of affective-interaction.

To address this challenge, we devised a framework to gather affective-interaction data through a computer game and a crowdsourcing process to acquire high-quality judgments of that data, see Fig. 1b. Crowdsourcing services are increasingly explored in

* Corresponding author.
E-mail addresses: g.tavares@campus.fct.unl.pt (G. Tavares), a.mourao@campus.fct.unl.pt (A. Mourão), jm.magalhaes@fct.unl.pt, jmc.mag@gmail.com (J. Magalhães).

http://dx.doi.org/10.1016/j.cviu.2016.02.001

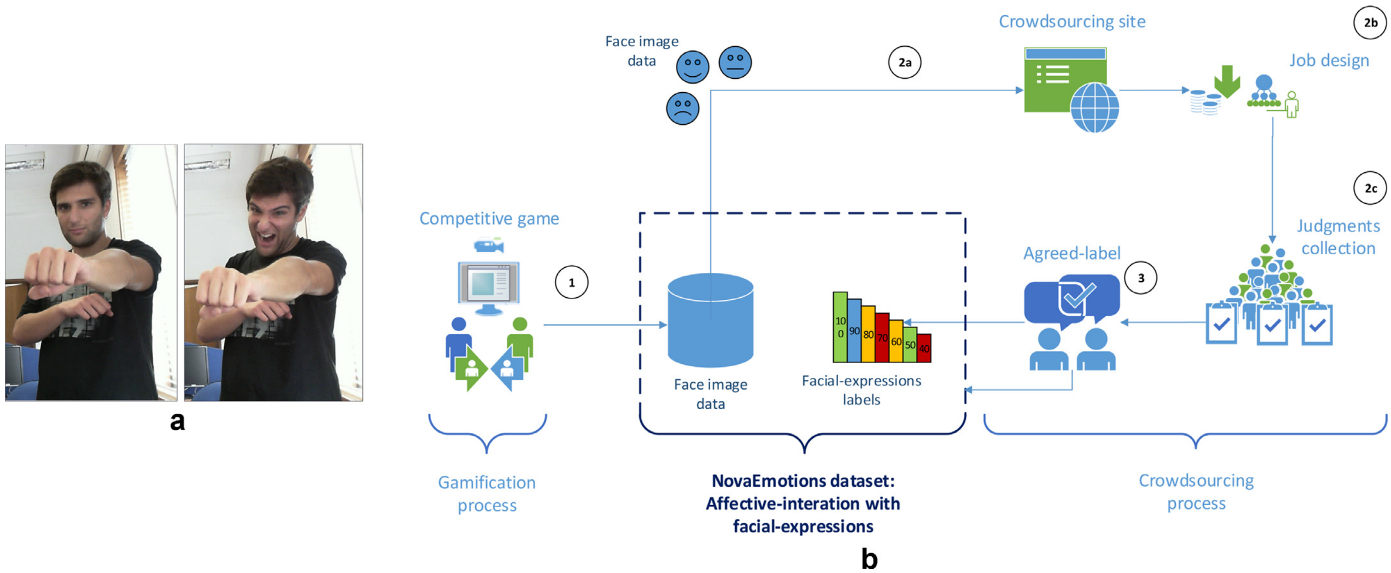

Fig. 1. Facial-expressions for affective-interaction: (a) the same action with different facial expressions, and (b) the gamification and crowdsourcing processes to generate the facial expressions dataset for affective-interaction.

Table 1
Crowdsourcing terminology.

Term	Defnition
Annotator	An annotator is a human that is being paid to provide judgments about a certain image
Judgments	A judgment is an annotator's decision about his/her perceived association between an image face and a facial expression
Gold question	A gold question is a control question to assess the quality of the judgments provided by the annotator
Job	A job is a set of judgments that an annotator must complete
Label	A label is the output of a statistical consensus method that aggregates the judgments provided by different annotators

research tasks to generate large volumes of human-level annotated knowledge [20]. In this paper, we will follow the terminology defined in Table 1. In our case, crowdsourcing enables the collection of an overcomplete set of facial expression judgments. In step 1, we implemented a game [15] that captures players' faces while interacting with the game. In this scenario, players were controlling the game-play through their facial expression – what we call an affective-interaction scenario. During the span of several game rounds, a large set of unlabeled interaction images was collected. Next, a crowdsourcing process was used to annotate each image with a facial expression. The design of the crowdsourcing job was carefully planned: we obtained several judgments per image (facial expression and corresponding intensity). This is particularly important as facial expression classification is a multi-class problem, instead of a binary.

The data was then uploaded to the crowdsourcing site[1] (step 2a), a crowdsourcing job was designed (step 2b), and a set of judgments were obtained for each face image (step 2c). These judgments were later merged with a statistical consensus methods to obtain the optimal labels, this corresponds to step 3. This last step is particularly relevant to get the maximum quality out of the crowdsourcing judgments. To improve the quality of the obtained judgments, i.e. inter-annotator agreement, we need to model the annotator's behavior when judging facial expression images. We profiled multiple state-of-the-art statistical consensus methods, in the facial-expression domain, using a dataset annotated with expert labels, the CK+ (Cohn–Kanade) [12]. Considering the label estimates provided by each method, several classifiers were then trained to recognize facial-expressions. This allowed us to further analyze the quality of the crowdsourcing labels.

In summary, the key contribution of this paper is a novel facial-expression dataset[2] to foster the affective-interaction field, in particular computer-games interaction. From this point forward, we will refer to the released dataset as the *NovaEmotions dataset*. The two other contributions confirm the value of the proposed dataset: building on previous work by Sheshadri and Lease [21] we benchmarked several state-of-the-art statistical consensus methods in the domain of affective-interaction, and compared the performance of facial-expression classifiers trained with expert labels and crowdsourced labels. These experiments confirmed that, although the crowdsourced labels are less than 9% different from the expert labels, the facial-expression classifiers present no significant difference. In the end, we obtained a unique facial expression dataset of users playing an affective-interactive computer game.

The remainder of this paper is organized as follows: Section 2 discusses related work and Section 3 details the acquisition of the 42,911 affective-interaction images. The crowdsourcing process for obtaining over 229,584 judgments of facial-expressions is presented in Section 4 and the results are discussed in Section 5. Finally, in Section 6, we assess the data quality by comparing different classifiers trained on the proposed dataset and the widely known CK+ dataset.

2. Related work

2.1. Human computation

Although computing power has exponentially increased in recent years, humans still achieve better results in understanding human languages, image semantics and many other tasks. Researchers have been studying ways of using humans as a source of human computation [18]. However, unlike computers, humans

[1] The Crowdflower service, http://www.crowdflower.com, was used in all experiments presented in this paper.

[2] The NovaEmotions dataset is available at: http://novasearch.org/datasets/.