Image and Vision Computing 72 (2018) 24–38

Contents lists available at ScienceDirect

Image and Vision Computing

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

Development and utilization of a disgusting image dataset to understand and predict visual disgust☆,☆☆

Sang-Il Oh, Hang-Bong Kang*

Department of Digital Media, Catholic University of Korea, 43-1, Yeoggok 2-dong, Wonmmi-gu, Bucheon-si, Gyeonggi-do, Republic of Korea

ARTICLE INFO

Article history:
Received 3 October 2016
Received in revised form 19 February 2018
Accepted 4 March 2018
Available online 12 March 2018

Keywords:
Visual disgust
Image dataset
Visual perception
Subjective assessment
Machine learning

ABSTRACT

When viewers search for images on the internet, they may unexpectedly encounter disgusting or explicit images. As such images may result in mental suffering or trauma, predicting whether images will induce disgust in order to avoid such issues is desirable. However, formal definitions or insights as to what constitutes disgust-inducing visual factors do not exist. Consequently, eliminating disgusting images from retrieval results is still a challenge. In this paper, we collect a large-scale disgust-inducing image dataset containing approximately 60,000 images, each labeled with disgust scores and divided into image categories. Subsequently, using our dataset, we explore various attributes of disgust-inducing images, such as score distributions, categories of disgusting images, and relationships with other visual attributes. Then, we develop a new Convolutional Neural Network (CNN), called DiNet, that uses more than two pre-trained convolutional layers to consider local to global features for image representation. Experimental results indicate that the developed CNN architecture outperforms both feature-based learning models and state-of-the-art deep learning models with an accuracy of 67.58%. Furthermore, disgust maps extracted using the developed model facilitate an understanding of the disgust-inducing regions of images.

1. Introduction

The number of users of image retrieval or photo sharing systems such as Google, Yahoo, Facebook, Instagram, Tumblr, and Flickr is increasing. Simultaneously, users are often confronted with explicit images that can invoke feelings of disgust. Several researchers [1,2,3,4,5] have investigated the symptoms observed when people see disgusting images. They found that minimizing exposure to unexpected disgusting images through diverse application studies is necessary for mental health and a healthy internet culture. Although the degree of disgust differs from person to person, people usually feel anger and negative emotions when they unexpectedly encounter disgust-inducing images. Our objectives are to facilitate an understanding of disgust-inducing images and to estimate visual disgust by introducing a large-scale disgust-inducing image dataset and related applications such as localizing the disgust-inducing regions.

To the best of our knowledge, no work has been done on understanding visual disgust as a high-level attribute. Much of the research conducted on feelings of disgust has been conducted from psychological or medical perspectives [1,2,3,4,5]. Previous computer vision and pattern recognition research related to exposure to unexpected images only dealt with specific disgust-inducing domains, like sexual [6,7,8,9] and violent images [10,11,12,13]. Additionally, previous studies on unexpected images gauged each image or video sequence on an absolute scale based on the presence or absence of disgusting content to filter them. In contrast, our goal is not only to find disgust-inducing images from image data, but also to understand the various cues of visual disgust to accurately estimate disgust-inducing scores. Existing methods that determine visual disgust from image data cannot insightfully understand the visual disgust. Consequently, they have severe limitations in finding fundamental solutions to the above issues.

To address this problem, we need a large-scale dataset containing disgust-inducing images. Although the International Affective Picture System (IAPS) [14] dataset does contain disgust-inducing images, the number of disgust-inducing images it contains is insufficient for our task. The IAPS dataset is also providing other various

☆ This paper has been recommended for acceptance by Hatice Gunes.
☆☆ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (No. 2015R1A2A1A10056304) and also supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2017-0-00819).
* Corresponding author.
E-mail addresses: nicolas0@catholic.ac.kr (S.-I. Oh), hbkang@catholic.ac.kr (H.-B. Kang).

https://doi.org/10.1016/j.imavis.2018.03.002
0262-8856/

images (in total of 504 images). Among them, the disgust-inducing images it contains are only double digits. Therefore, to understand visual disgust and build a prediction model, we first developed a new dataset comprising 61,573 disgust-inducing images obtained from photo sharing sites (green box in Fig. 1). The images in the dataset are grouped into four categories per the level of disgust induced in internet users, which we ascertained via an internet survey form. Additionally, we conducted a user study using a crowdsourcing program, in which subjects were asked to identify the more disgusting image in a pair of images and to comment on the pair, from which we extracted a ground truth score of visual disgust for each image.

To understand visual disgust, we explored our dataset from several perspectives (red box in Fig. 1). We conducted our investigation through key attributes by analysis of disgust scores with static analysis and categories of disgust-inducing images. The categories of disgust-inducing images are diverse and each has its own disgust-inducing attributes. For example, a dotted pattern can be a feature of small bug images, and a bloody red color can be a feature of blood images. On the basis of comments received in the user study, we selected frequently chosen disgust-inducing components in each category to determine which image transformation methods can reduce the feeling of disgust.

Most visual perception can be expressed by other visual subjective aspects; visual aesthetics do not correlate with memorability [15]. To explore the relationship between disgust score and other visual subjective attributes, we measured the memorability [15], aesthetics [16], photonics [17], interestingness [18], and ugliness of our dataset. The visual subjective attributes used in the analysis can be applied to normal images, but not to specific attribute domains such as fashionability [19]. We also measured the Spearman's correlation between each extracted visual subjective score and the disgust score of our dataset.

Subsequently, we developed a new Convolutional Neural Network (CNN) architecture to automatically predict disgust scores, called DiNet (blue box in Fig. 1). The main objective of DiNet is simultaneously considering image features from fine-grained local features to globally represented features and to accurately estimate a disgust score. Previous models that used CNNs [20,21,22,23,24] fed the final output of the fully connected layers into the last loss layer for their tasks. However, this resulted in the possibility of losing the local information of the image by bypassing a layer. Additionally, the outputs of early layers are omitted from learned representations. In contrast, DiNet generates a convolutional cube (ConvCube) from more than one convolutional layer of a pre-trained CNN model as representations, as opposed to an all-passed output. A given ConvCube is fed into another fine-tuned CNN comprising two convolutional layers, two fully connected layers, and a loss layer to estimate the disgust score of an input image.

Additionally, we extracted a disgust map using DiNet. The disgust map from DiNet can be used to find and explore disgust-inducing regions to automatically filter or transform them. To extract the disgust map using DiNet, we divide the image into image patches to be overlapped. Then, each overlapped image patch is fed into DiNet to estimate their disgust scores. The given set of image patches with the disgust scores are then rendered back into the form of the original image to obtain the disgust map.

Our contributions in this paper are as follows: 1) we present a developed large-scale disgust-inducing image dataset, 2) we explain our understanding of disgust-inducing factors from several perspectives, and 3) we model these factors to automatically predict disgust scores. We collected 60,000 disgust-inducing images from websites and categorized them manually. Then, using the images in our dataset, we measured the disgust score via a crowdsourced user study to understand visual disgust per image category, the distribution of disgust scores, and disgust-inducing regions. We also investigated and developed a machine learning model for accurately estimating the disgust score of input images. Finally, from the proposed machine learning model for estimating disgust score, we extracted disgust maps of input images that can be used to detect disgust-inducing regions and transform them.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our disgusting image dataset and the image annotations therein. Section 4 outlines the analysis conducted of the disgust-inducing images to understand them. Section 5 analyzes machine learning models to determine which models can suitably predict visual disgust. Section 5 demonstrates extraction of a disgust map using DiNet. Section 6 presents the results of machine learning models used to estimate disgust score

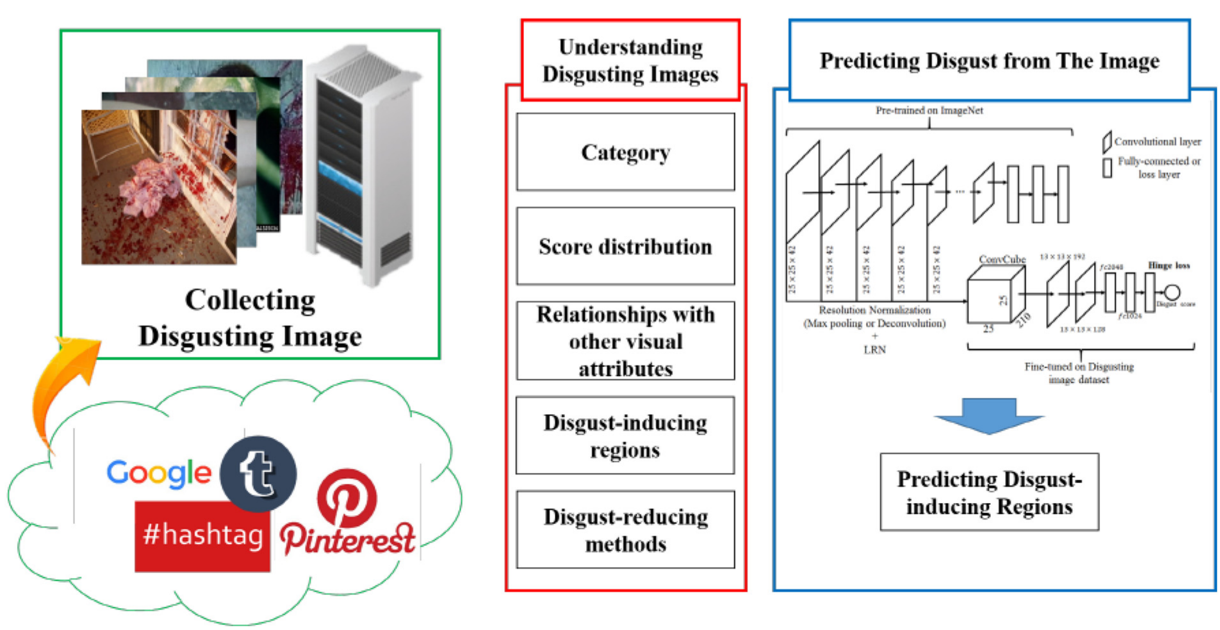

Fig. 1. The three phases comprising our study: 1) (green box) collection of disgusting images from websites, 2) (red box) evaluation of various perspectives from our dataset, and 3) (blue box) creation of a new model to predict disgust scores and extract disgust-inducing regions. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)