J. Vis. Commun. Image R. 38 (2016) 276–283

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Evaluation of local and global descriptors for emotional impact recognition ☆

Syntyche Gbèhounou, François Lecellier *, Christine Fernandez-Maloigne

University of Poitiers, XLIM Laboratory, UMR CNRS 7252, Poitiers, France

ARTICLE INFO

Article history:
Received 17 February 2015
Revised 10 February 2016
Accepted 7 March 2016
Available online 7 March 2016

Keywords:
Images
Local descriptors
Subjective evaluations
Emotions
Classification

ABSTRACT

In order to model the concept of emotion and to extract the emotional impact from images, one may search suitable image processing features. However, in the literature, there is no consensus on the ones to consider since they are often linked to the application. Obviously, the perception of emotion is not only influenced by the content of the images, it is also modified by some personal experiences like cultural aspects and semantic associated to some colours or objects. In this paper, we choose low level features frequently used in CBIR especially those based on SIFT descriptors. To take into account the complex process of emotion perception, we also consider colour and texture features and one global scene descriptor: GIST. We supposed the chosen features could implicitly encode high-level information about emotions due to their accuracy in the different CBIR applications of the literature.

We test our methodology on two databases: SENSE and IAPS.

1. Introduction

In past decades, many achievements have been made in computer vision in order to replicate the most amazing capabilities of the human brain, for example image classification according semantic content or people tracking in video surveillance. However, there are some aspects of our behaviour or perception which remain difficult to apprehend, for example emotion prediction from an image or a video. This has several application such as: film classification, road safety education, advertising or e-commerce, by selecting appropriate images depending on the situation.

In order to predict the emotional impact of an image or a video, one first need is to describe what an emotion is and how to categorize them. There are two emotion classifications used in the literature [1]:

1. **Discrete approach**: emotional process can be explained with a set of basic or fundamental emotions, innate and common to all human (sadness, anger, happiness, disgust, fear, ...). There is no consensus about the nature and the number of these fundamental emotions. This modelling is usually preferred in emotion extraction based on facial expressions.
2. **Dimensional approach**: on the opposite, the emotions are considered in this model as the result of fixed number of concepts such as pleasure, arousal or power, represented in a dimensional space. The chosen dimensions vary depending to the needs of the model. Russel's model is the most considered, Fig. 1, with the dimensions valence and arousal:
   - ***The valence*** corresponds to the way a person feels when looking at a picture. This dimension varies from negative to positive and allows to distinguish between negative emotions and pleasant ones.
   - ***The arousal*** represents the activation level of the human body.

The advantage of the dimensional approach is to define a large number of emotions without the limitation of a fixed number of concept as the discrete ones. In spite of this advantage, some emotions can be confused (such as fear and anger in the circumplex of Russel) or unrepresented (among others surprise in Russel's model).

In the literature, a lot of works are based on the discrete modelling of the emotions; for example those of Paleari and Huet [2], Kaya and Epps [3], Wei et al. [4] or Ou et al. [5–7]. In this paper, our goal is to obtain a classification into three different classes

☆ This paper has been recommended for acceptance by Chang-Su Kim.

* Corresponding author.

E-mail addresses: syntyche.gbehounou@univ-poitiers.fr (S. Gbèhounou), francois.lecellier@univ-poitiers.fr (F. Lecellier), christine.fernandez@univ-poitiers.fr (C. Fernandez-Maloigne).

http://dx.doi.org/10.1016/j.jvcir.2016.03.009
1047-3203/

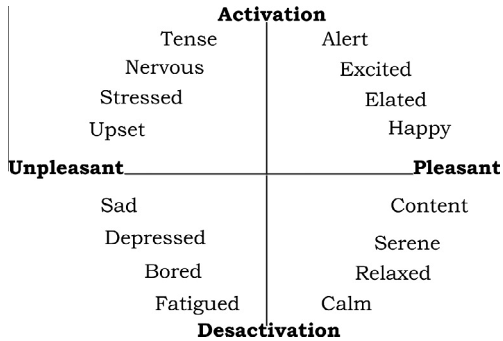

Fig. 1. Russel's emotions modelling. The axe Unpleasant/Pleasant corresponds to the arousal and the second one to the valence.

"Unpleasant", "Neutral" and "Pleasant". To tackle this objective, we choose a dimensional approach, since in discrete one the number and nature of emotions remain uncertain. Moreover, there are concepts which cannot be assigned to a specific class (surprise for example can be pleasant or unpleasant).

The extraction of emotional impact is an ambitious task since the emotions are not only content related (textures, colours, shapes, objects, ...), but also depend on cultural and personal experiences.

In the past decade, lots of papers have been devoted to the links between emotions and colours [4–11]. Several of them consider emotions associated with particular colours through culture, age, gender or social status influences. There is a consensus among the authors to conclude that a link exists between colours and particular emotions. As stated by Ou et al. [5], colours play an important role in decision-making, evoking different emotional feelings. The research on colour emotion or colour pair emotion is now a well-established area of research. Indeed, in a series of publications, Ou et al. [5–7] studied the relationship between emotions, preferences and colours and have established a model of emotions associated with colours from psychophysical experiments.

Another part of the literature is dedicated to facial expression interpretation [2]. In this work, emotions are associated with facial features (such as eyebrows, lips). Since facial expressions are common among humans to express basic emotions (happy, fear, sadness, surprise, ...), it seems to be the easiest way to predict them. Nevertheless, in this case, the authors extract emotions carried by the images and not really those felt by someone looking at these pictures.

More recently some authors looked at the emotion recognition as a Content Based Image Retrieval (CBIR) task [12–14]. Their underlying idea consists in considering the traditional image retrieval techniques to extract the emotional impact of images. To achieve this goal, the authors used a multistage method, at first by extracting traditional image features (colours, textures, shapes) and then combined those features into a classification system after a learning step. For example, Wang and Yu [15] used the semantic description of colours to associate an emotional semantic to an image. Concerning textures, the orientation of the different lines contained in the images is sometimes considered. According to Liu et al. [1], oblique lines could be associated with dynamism and action; horizontal and vertical ones with calm and relaxation.

Our work is part of this last family of approaches. We evaluated some low level features well adapted for object recognition and image retrieval [16–22] and conducted our work on two databases:

- A set of natural images that was assessed during subjective evaluations: Study of Emotion on Natural image databaSE (SENSE) [23].
- A database considered as a reference on psychological studies of emotions: International Affective Picture System (IAPS) [24].

This paper is organized as follow. We describe the image databases in Section 2 and the features used for emotion recognition in Section 3. The classification process is explained in Section 4. In Section 5 we summarize our results. We conclude about our study and provide some perspectives in Section 6.

2. Image databases

In the domain of emotion extraction, the choice of the database is not trivial since there is no reference for all emotion studies and applications, some authors even built their own dataset without spreading it. We choose in this study to considered two databases: the first one is composed of low semantic images and the second of more semantic ones. In this paper, "low-semantic" means, that the images do not shock and do not force a strong emotional response. We think that even low semantic images, including abstract representation, may produce emotions according to the viewer sensitivity and the viewing time. We define this "low-semantic" criteria in response to some high semantic images on IAPS [24], a reference in psychological studies on emotions. In this database, which will be described in Section 2.1, there are images with blood, dirt, high semantic photomanipulations or naked people which might induce a bias in the assessment of the images. There is a risk of overreaction to neutral images viewed right after strong pleasant or unpleasant ones. Our aim in this paper is to evaluate the behaviour of our strategy developed for a low semantic database on a more semantic one.

2.1. The International Affective Picture System (IAPS)

This dataset is developed since the late 1980s at NIMH Center for Emotion and Attention (CSEA) at the University of Florida [24], which is composed of photographs used in emotion research. It is considered as a reference in psychological studies and many papers present results on this base [1,13,14].

The images of IAPS are scored according to the affective ratings: pleasure, arousal and dominance. It corresponds to a dimensional representation of emotions. The affective norms for the pictures in the IAPS were obtained in 18 separate studies involving approximately 60 pictures for each session. Each of the 1182 images from their dataset was evaluated by about 100 participants.[1] The emotion of the images we have chosen is based of those defined in different papers [25–27] using three class classification: "Pleasant", "Neutral" and "Unpleasant".

2.2. SENSE

Study of Emotion on Natural image databaSE (SENSE) is the database used in [23,28]. It is a low semantic, natural and diversified database containing 350 images free to use for research and publication. It is composed of animals, food and drink, landscapes, historic and tourist monuments as shown in Fig. 2 with some examples.

Some transformations were applied on some images of the database: geometric modifications and changes on colour balance, so some images are repeated twice or more. In this database, only 17 images (4.86%) contain human faces to ensure that the facial expression does not induce bias. It is composed of low semantic images since they minimize the potential interactions between emotions on following images during subjective evaluations. This

[1] It is the size of the database when we received it.