Enriching Image Datasets with Unrestrained Emotional Data: A Study with Users

Soraia M. Alarcão and Manuel J. Fonseca
LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
salarcao@lasige.di.fc.ul.pt, mjfonseca@ciencias.ulisboa.pt

***Abstract*—Elicitation of emotions is typically done through the presentation of emotionally salient material, like images or videos, thus requiring reliably annotated datasets. Although there are datasets with emotional information, these only describe either emotional polarities or discrete emotions. The only available dataset with both types of information restrained the participants during the study by separating a priori the images according to their polarity (positive or negative). In this paper, we describe an unrestrained study with 60 participants, where we asked them to rate the polarities and discrete emotions elicited by a set of images. The analysis of the emotional ratings made by the users revealed the most frequent correlations between the basic emotions. Furthermore, the analysis of the ratings' agreement among participants and existing datasets shows that our results are aligned with the existing ones. As a result of our study, we make available to researchers a more informative picture dataset annotated with emotional polarities and multiple emotions, as a complement to existing datasets.**

I. Introduction

The role of emotions in human cognition is essential given their importance in the daily life of human beings. Emotions play a critical role in rational decision-making, perception, human interaction, and intelligence [1], [2].

In the last decade, there has been an increasing body of work involving emotions: to improve content-based classification for both music and video, using photos and emotions conveyed by multimedia [3]; to gather emotional information from images through their visual content [4]; to observe the emotional state of a person using Electroencephalography [5]; to improve interactive experiences using user emotional expressions [6]; and finally, to enhance the quality of recommendation systems [7].

Besides these examples, many studies in psychology and computer science involve manipulating emotions via emotional stimuli [8]. If a stimulus is relevant enough, an appraisal is automatically executed and will trigger reactions in measurable components of emotion, such as physiological responses, expressivity, action tendencies, and subjective feelings. Several methods have been introduced for priming participants, such as the presentation of emotionally salient material like images [9], audio [10], video [11], or text [12]. The use of the visual channel remains the most common to convey emotional stimulation [13].

DOI reference number: 10.18293/VLSS2018-033

In the different areas of research based on visual stimulation, reliable datasets are important for the success of emotion induction. To that end, in 1997, the International Affective Picture System (IAPS) dataset was presented [14]. Later, in 2011 and 2014, two new datasets were created: Geneva Affective PicturE Database (GAPED) [13], and Nencki Affective Picture System (NAPS) [15]. These increased the availability of visual emotion stimuli, while trying to solve the problem of a limited number of pictures for specific themes. IAPS only provides valence and arousal, while GAPED has some information about the emotional polarity (negative, neutral or positive) of their images, but it is not enough for the cases where there is the need to use discrete emotions.

To minor the lack of emotional information, in 2005 and 2016, Mikels [16]–[18] and NAPS Basic Emotions (NAPS-BE) [19] were presented. Mikels collected descriptive emotional data on a subset of the IAPS to identify the elicited discrete emotions. Although this work enriched the emotional information associated to the IAPS dataset, we believe that the authors have restrained the choices of the participants by asking them to select discrete emotions only in a specific polarity (positive or negative), according to the subset where the image was placed a priori by the authors. This restriction prevented mixtures of positive and negative emotions. However, it is possible that an image arouses positive emotions in a person and negative in another. Finally, authors did not consider that images could be neutral.

In this paper, we present a study about the experience of viewing a set of images from the IAPS and GAPED datasets. We focused on the process of rating the images according to the emotions and polarities they elicited in the viewer, as well as the participants' insights during the experience. Although it would be interesting to use images from the NAPS-BE, it was not yet available when we conducted the study. Our contributions are: 1) a more complete and realistic picture dataset composed of 169 images, each annotated with information about the predominant emotional polarity, the intensity of each discrete emotion elicited by the image, and the valence and arousal values from the original datasets; 2) the relationship between multiple emotions that arise when visualizing images, that are in line with the literature, thus confirming the quality of our dataset emotional annotation; 3) our experimental procedure designed to provide more comfort to the users, avoiding stress and fatigue.

II. Background and Related Work

In this section, we briefly explain what are emotions and how we can represent them. We also describe the most commonly used datasets of images to elicit emotions.

A. Emotions

Polarity provides a coarse indication of the emotional image content (positive, neutral, and negative). Emotions, on the contrary, give a more detailed description of the emotional information conveyed. These have been described as discrete and consistent responses to external or internal events with particular significance for the human organism [20]. This finer distinction of emotions provides a richer emotional classification, making it suitable for specific research purposes, such as studying the neuroanatomical correlations among basic emotions when a person is exposed to multimedia stimuli [21].

When talking about emotions, it is important to mention the subjectivity inherent, since multiple emotions can appear in the same subject while looking, for example, at a picture, as well as different subjects can feel different emotions when viewing the same picture, mainly due to each subject's current emotional state and "life experiences" [22], [23]. However, the expected affective response can be considered objective, as it reflects the more-or-less unanimous response of a general audience to a given stimulus [24].

Regarding the existence of multiple emotions while viewing an image, these correlations of basic emotions are a well-known phenomena in the field of psychology. One of the most important results was that when happiness rises, all other emotions decline; another one is that fear correlates positively with sadness and anger [25], [26].

B. Emotions Representation

There are two different perspectives towards emotion representation: categorical and dimensional. The first indicates that basic emotions have evolved through natural selection. Plutchik proposed eight basic emotions (acceptance, anger, curiosity, disgust, fear, joy, sadness, and surprise), from which we can define all the others [27]. Ekman based his work in the relationship between facial expressions and emotions derived from the universal basic emotions (anger, disgust, fear, happiness, sadness, and surprise) [28]. These emotions are considered universal since their external manifestation seems to be independent of culture and personal experiences [29].

In the dimensional perspective, which is based on cognition, the emotions are mapped into the Valence, Arousal and Dominance (VAD) dimensions. Valence goes from unpleasant to pleasant, arousal goes from states like sleepy to excited, and finally, dominance corresponds to the strength of the emotion [14], [30]. The most common model used is the Circumplex Model of Affect (CMA), where all affective states arise from cognitive interpretations of core neural sensations that are the product of valence and arousal [31].

In this work, we used Ekman's set of universal emotions (anger, disgust, fear, happiness, sadness, and surprise) complemented with the neutral emotion.

TABLE I
Comparision among the most commonly used datasets of images.

Dataset	#Images	V-A	Polarities	Emotions
IAPS	1182	Yes	No	No
EmoPics	378	Yes	No	No
GAPED	730	Yes	Yes	No
NAPS	1356	Yes	No	No
POFA	110	No	No	Yes
KDEF	4900	No	No	Yes
NimStim	646	No	No	Yes
ArtPhoto	807	No	No	Yes
Abstract	228	No	No	Yes
Mikels	330	Yes	Yes[2]	Yes
NAPS-BE	510	Yes	No	Yes

C. Image Datasets

In all the different areas of research based on visual stimulation, reliable databases are important for the success of emotion induction. In Tables I and II, we briefly present the most commonly used datasets of images to elicit emotions.

As we can see in Table I, only GAPED and Mikels provide information about the polarity of an emotion, i.e., negative, neutral or positive (Mikels does not consider the neutral polarity). In Mikels, the authors defined the emotional polarity of an image before the participants performed their rating about the discrete emotions. Given the subjectivity inherent to emotions, this could have restrained the results since it did not allow people to express positive emotions for "negative" images, and vice-versa. For example, Yoon *et al.* concluded that some of images did not have agreement between the tags assigned by the image creators and the ones given by image viewers [32].

Machajdik datasets (Art Photo and Abstract Paintings) [33], Mikels, and NAPS-BE discriminate the emotions elicited by images. However, Abstract Paintings is focused in a very specific type of images that are not usually found in personal collections, while the ratings for images of the Art Photo were only done by the artists. IAPS, Emotional Picture Set (EmoPicS) [34], and NAPS do not provide any information about the emotional content of their images, offering only valence and arousal information or physical characteristics of the images. Karolinska Directed Emotional Faces (KDEF) [35], NimStim Face Stimulus Set (NimStim) [36], and Pictures of Facial Affect (POFA)[1] were only labeled with facial expressions and corresponding emotions.

Some datasets have Valence and Arousal (VA) information, but no emotional data; others have emotional information, but no VA; and finally, only GAPED, NAPS-BE, and Mikels have both, but they are restrained and limited.

[1] http://www.paulekman.com/product/pictures-of-facial-affect-pofa/

[2] The emotional polarity (negative or positive) for each image was defined by the authors, not collected from the participants.

TABLE II
DESCRIPTION OF THE MOST COMMONLY USED DATASETS OF IMAGES TO ELICIT EMOTIONS.

Dataset	Description
IAPS	It contains 1182 images, and provides a set of normative emotional stimuli for experimental investigations of emotion and attention. The authors rely on a dimensional view, in which emotions are defined by a coincidence of values on a number of VAD dimensions. Each picture is characterized in terms of their valence and arousal ratings. They were made by males, females and children using Self-Assessment Manikin (SAM) questionnaires during 10 years [37].
EmoPicS	It contains 378 standardized color images with different semantic contents, such as social situations, animals, and plants, selected from public online photo libraries and archives. Each image of the database was rated with their corresponding dimensional information: valence and arousal, and also with some physical characteristics of the given image: color composition, contrast, and luminance.
GAPED	It contains 730 pictures: 121 representing positive emotions using human and animal babies as well as natural sceneries, 89 for the neutral, mainly using inanimate objects, and 520 for the negative, using spiders, snakes, human rights violation, and animal mistreatment. The pictures were rated according to valence, arousal, and the congruence of the represented scene with moral and legal norms regarding Swiss legislation, since the study was conducted in Switzerland. These ratings were made by 60 subjects, where each subject rated 182 images.
NAPS	It contains 1356 realistic, high-quality images divided into five categories: animals, faces, landscapes, objects, and people. Besides valence, arousal and motivational direction (avoidance-approach) ratings, each image was annotated with some physical characteristics, namely color composition, contrast, and luminance. 204 subjects made the ratings, where each one rated 362 images, pseudo-randomly chosen from all the categories with the constraint that no more than three stimuli of the same category were presented in succession.
POFA	This dataset consists of 110 photographs of facial expressions that have been widely used in cross-cultural studies, and more recently, in neuropsychological research. All images are in black and white, and each image has a set of norms associated. It is important to note that the images are not identical in intensity or facial configuration.
KDEF	It is a set of 4900 pictures of human facial expressions of emotion suitable for perception, attention, emotion, and memory. Thus, special attention was given to photograph expressions at different angles, with soft light, and using t-shirts with uniform colors. A grid was used to center the face of the users during shooting, as well as position the eyes and mouth in certain coordinates of the image during scanning. The set contains 70 individuals, each displaying seven different emotional expressions, which were photographed from five different angles.
NimStim	It consists of 646 facial expression stimuli. Images include fearful, happy, sad, angry, surprised, calm, neutral, and disgusted expressions displayed by a variety of models of various genders and races. Examples of facial expressions were shown to the actors, for them to get an idea of what was the aim, and then they posed for each facial expression. Muscles were adjusted until the desired expression was achieved.
Art Photo	It contains 807 artistic photographs that were obtained by using the emotion label as search terms in the deviantArt site. The emotion label was determined by the artist who uploaded the photo, that was trying to evoke a certain emotion in the viewer of the photograph through the conscious manipulation of the image composition, colors, etc.
Abstract	It contains 228 images with combinations of color and texture, without any recognizable objects. To obtain ground truth, images were peer rated in a web-survey where the users could select the emotional category from amusement, anger, awe, contentment, disgust, excitement, fear and sad, for 20 images per session. 230 people rated approximately 280 images, where each image was rated about 14 times.
Mikels	This dataset is composed of 330 images from the IAPS, annotated with positive (amusement, awe, contentment, and excitement) and negative (anger, disgust, fear, and sadness) emotions. Thirty males and 30 females made the emotional category ratings in two studies, using a subset of negative images and a subset of positive images, with a constrained set of categorical labels.
NAPS-BE	This dataset contains 510 images from the NAPS, annotated with the emotions anger, disgust, fear, happiness, sadness, and surprise. It has 98 images depicting animals, 161 faces, 49 landscapes, 102 objects, and 100 people. Sixty seven females and 57 males made the emotional ratings, where each subject rated around 170 images.

III. EMOTIONAL USER STUDY

In this section, we describe the study carried out, in which participants identified both the emotional polarity and emotions they felt while visualizing each image.

A. Participants

Sixty participants completed the study: 26 females and 34 males, with 70% of them belonging to the 18-29 age group, and almost 60% having a BSc Degree. None of the participants had participated in any study using the IAPS or GAPED, and the overwhelming majority had no knowledge about these datasets.

Regarding their emotional state at the beginning of the study, 31 participants classified themselves as neutral, 25 as positive, and only 4 as negative. Considering the discrete emotions (anger, disgust, fear, happiness, neutral, sadness, and surprise), the majority of the participants were feeling moderately happy or moderately neutral, both with a median of 3 in a scale of 1-5, with 1 corresponding to a weak feeling, and 5 to a strong feeling.

B. Apparatus and Material

A MacBook Pro (13-inch) computer was used with an application for participants to see the images and rate the emotions and polarities elicited by each image.

The dataset used in the study was composed of 86 images from the IAPS, 76 images from the GAPED, and 7 images from Mikels' dataset. It contained images with animals (cats, dogs, horses, sharks, snakes, spiders, tigers, among others), car accidents, children, death situations, diseases, fire, mutilation, natural catastrophes, poverty, and war scenarios. We chose a set of images that we believed to represent in a balanced way the discrete emotions throughout the valence-arousal space (see Figure 1).

Since it was impractical and even unpleasant for participants to annotate all the images in our dataset, and also due to the time it would take, we randomly divided our dataset into four subsets: DS0 to DS3. DS0 contained 57 images (30 IAPS, 20 GAPED, 7 Mikels), DS1 contained 40 images (20 IAPS, 20 GAPED), while DS2 and DS3 contained 36 images each (18 IAPS, 18 GAPED).

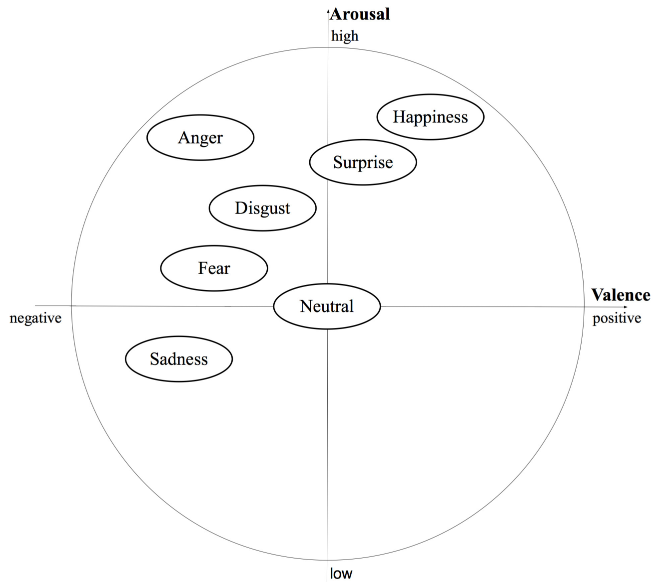


Fig. 1. Adaptation of the Circumplex Model of Affect, mapping the discrete emotions into the Valence-Arousal plane [38].

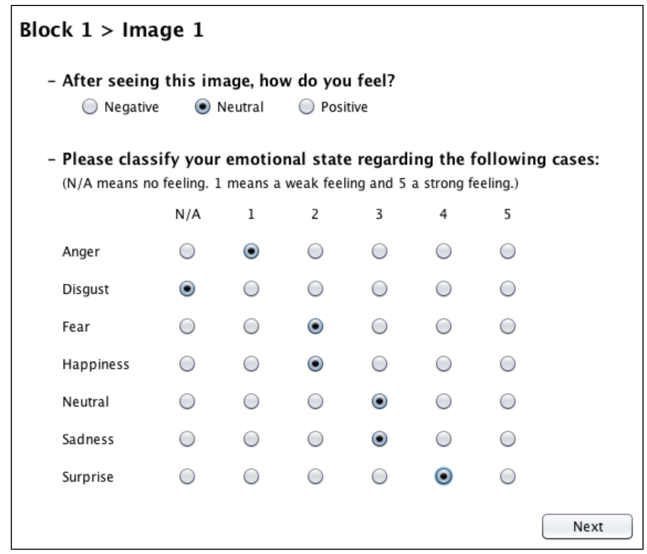


Fig. 2. Rating screen of the application with the 5-point Likert scale.

All the participants rated each image of DS0, while images from DS1, DS2 and DS3 were rated by 20 participants. With this process, we managed to get a larger number of annotated images in the shortest time possible.

C. Design and Procedure

The experimental sessions took place in a room properly prepared for the task, aiming at providing comfort to participants, with adequate lighting and isolation from external noises. The option for the solo exhibition seeks to contribute to better control of external interference (e.g., comments from other participants, noise) that could interfere with emotional participant's experience [39].

We started by explaining the purposes of the study and how it would be held. To ensure the willingness of the subjects regarding negative images, we showed three images as examples of what could be expected. After that, the subjects could decide whether to continue or not the study. One participant (not included in the 60) decided not to continue the study due to medical issues. If they accepted, they should fill the participants' questionnaire with their personal information (age, gender, etc.), and the classification of their current emotional state (polarity and emotions).

The first screen of the application presented a summary of the most important aspects of the study. Then, seven blocks of images were presented sequentially, with about 14 images on each block. Each image (with a resolution of 640x480 pixels) was displayed randomly during 5 seconds, and after the visualization, participants evaluated their emotional state (regarding the polarity felt), and rated it for each of the emotions (see Figure 2). To obtain the participants' emotional reactions without practical limitations (e.g. specialized equipment for collecting physiological signals), we adopted a 5-point Likert scale for each emotion.

This process was repeated for each image of the seven blocks of images of our study. Although in similar studies participants usually had a limited time to answer, we decided not to do it. This way, we allowed participants to spend the time they needed, without feeling pressured to respond or even stressed out. We also provided a 30 seconds interval between each block of images, during which only a black screen was displayed, to relax the user and avoid fatigue.

To verify and validate if our procedure had any error and if it was completely clear to the subjects, we performed a pilot test with a 27 years old male and a 18 years old female. With the exception of an image that was duplicated, none of the subjects had any doubt or detected any error in our study. An interesting aspect identified in this pilot test was the different sensitivities of the participants to the negative images. One subject considered the majority of the images very violent, while the other considered them almost neutral, and in some cases he enjoyed the consider negative content.

IV. Emotional Classification Procedure

In this section, we describe the procedure used to classify each image based on the participants' ratings both in terms of the dominant polarity and discrete emotions.

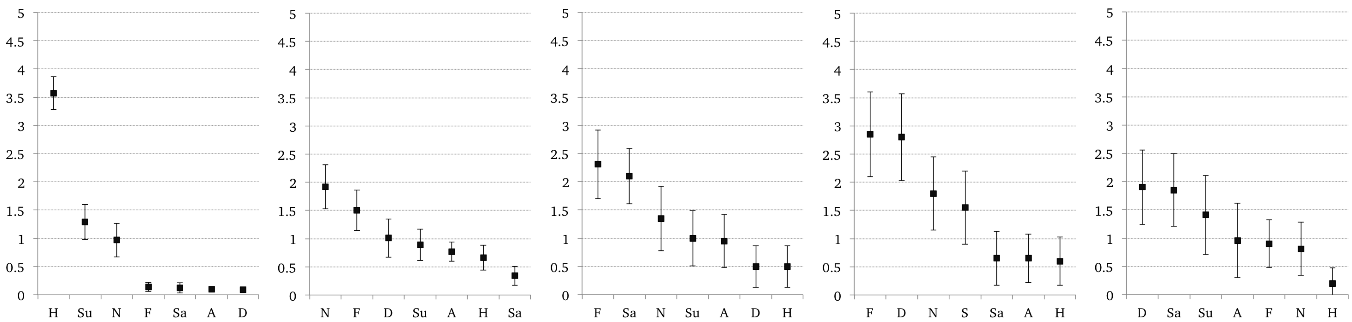
To assign an emotional polarity to an image, we chose the polarity with the highest number of votes. In Table III, we present examples of the distribution of votes across each polarity, while Figure 3 depicts the corresponding images.

TABLE III
Examples of the distribution of votes across each polarity.

Image	Negative	Neutral	Positive	Assigned Polarity
1460.jpg	0.0%	13.3%	86.7%	Positive
Sn087.jpg	20.0%	68.3%	11.7%	Neutral
9925.jpg	40.0%	50.0%	10.0%	Neutral
Sp044.jpg	40.0%	50.0%	10.0%	Neutral
3017.jpg	75.0%	20.0%	5.0%	Negative

(a) 1460.jpg (b) Sn087.jpg (c) 9925.jpg (d) Sp044.jpg (e) 3017.jpg

Fig. 3. Examples of images from our dataset depicting: (a) kitten, (b) snake, (c) fire, (d) spider, and (e) mutilation.



(a) single (b) blended: 2 emotions (c) blended: 3 emotions (d) blended: 4 emotions (e) undifferentiated

Fig. 4. Examples of Confidence Intervals of images from our dataset, and how they are classified according to our procedure: (a) happiness emotion, (b) neutral and fear emotions, (c) fear, sadness and neutral emotions, (d) fear, disgust, neutral, and sadness emotions, and (e) undifferentiated.

We considered that an image could transmit up to four emotions, with no constraints about their polarity. We made this decision because Posner *et al.* stated that "*individuals do not experience, or recognize, emotions as isolated, discrete entities, but that they rather recognize emotions as ambiguous and overlapping experiences*" [31].

To identify the dominant emotions for each image, we followed the procedure from Mikels *et al.* [16]. However, since we are considering more emotions per image than Mikels (four vs three), our procedure is slightly different. For each image, we computed the mean of the ratings assigned by participants to each emotion, and a 90% t-based Confidence Interval (CI) around each mean. Then, the emotions' label was determined according to the overlap of the CIs for each emotion.

If the mean for one emotion is higher than the means of all the other emotions, and if the CI for that emotion does not overlap with the CIs for the other emotional labels, it is classified as a single emotion (see Figure 4a). If two, three or four means are higher than the rest, and the intersection between their CIs is not empty, the image is categorized as blended (see Figures 4b - 4d). If more than four CIs overlap, the image is classified as undifferentiated (see Figure 4e).

In our study, and contrary to what Mikels did, we could have images with a mix of negative and positive emotions.

V. Results

In this section, we present the polarities agreement and emotional labels assigned to each image. We also present the most elicited emotions together. Finally, we present observations made by our participants during the study.

A. Agreement of Polarity Among Users

In Figures 5 and 6 we can observe, in detail, the votes of the users for each image in our dataset. From the 82 images classified as negative, 77 images had more than 50% of negative votes. The remaining votes were mainly neutral (45 images were rated with at most 30% of neutral votes, while 47 images had at most 5% of positive votes).

Regarding the 66 images classified as neutral, 62 of images had more than 50% of neutral votes. The remaining votes were usually rated more often as negative than positive (37 images with at most 30% of negative votes, while 41 had at most 15% of positive votes). Finally, all the 21 images classified as positive had more than 50% of positive votes. Eighteen images had at most 5% of negative votes, while 10 were rated with at most 30% of neutral votes.

In summary, all polarities were very well identified. When there was some mixing with either the positive or negative polarity, they were mixed with the neutral polarity. For the neutral polarity, it was mainly mixed with the negative polarity.

B. Agreement of Polarity Among Datasets

We compared our results only with GAPED because IAPS does not provide polarity information, and although Mikels provides information about the polarity, it was classified by the authors not by the participants.

We analyzed 76 images (33 negative, 9 positive, and 34 neutral) from the GAPED. For the neutral and positive polarities, we achieved an agreement of 100% for each. For the negative, the achieved agreement was 69%. The biggest mixed was with the neutral polarity (28%), while the mix with the positive polarity was very small (3%).

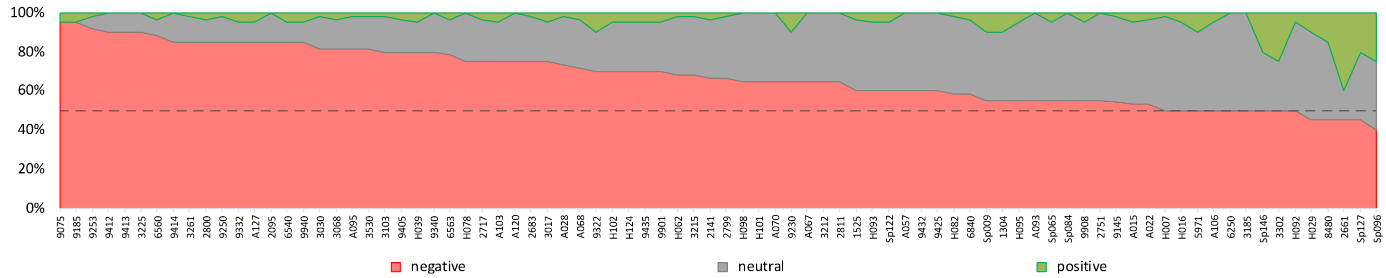


Fig. 5. Images classified as negative in our dataset. We show the percentage of votes that users assigned to each polarity. (best seen in color)

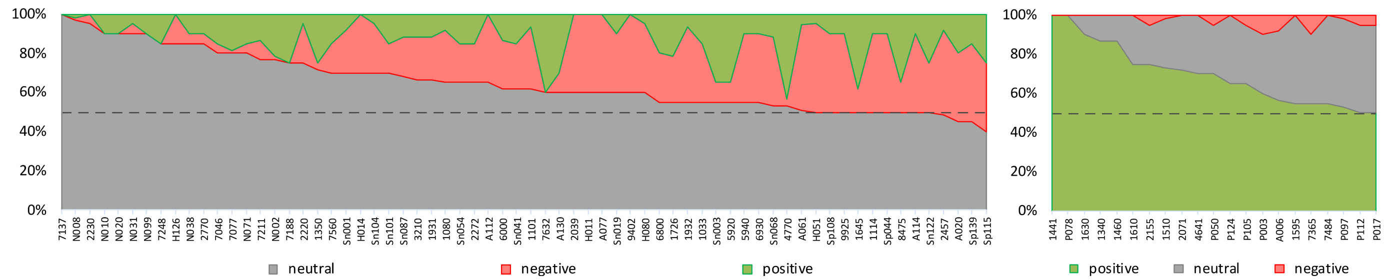


Fig. 6. Images classified as neutral (left) and positive (right) in our dataset. We show the percentage of votes that users assigned to each polarity. (best seen in color).

Dan-Glauser *et al.* also reported that their results in GAPED had a high percentage of negative valence ratings overlapping with the neutral for animal mistreatment, spider, human concern, and snake pictures [13].

C. Valence and Arousal Space

In Figure 7, we present the distribution of the ratings in the valence and arousal space.

For each polarity, a polygon delimits the space in which all pictures of the same polarity are found. If we compare the distribution of the polarities with the emotions displayed in Figure 1, we can see that there is a clear correspondence between the negative emotions with the negative polarity, as well as between the neutral emotion and the neutral polarity.

For the positive polarity, this correspondence with the happiness emotion is not so obvious, but it is easy to see that there is no overlap between the negative and positive polarities. Finally, we can see that there is some confusion between the neutral and negative polarities, as well as between the neutral and positive ones, however less significant.

D. Emotional Labels

From the 169 images of our dataset (see Table IV), we obtained 60 images annotated with a single emotion (35.5%), 87 classified as blended (51.5%), with 29 referring to the combination of two emotions (17.2%), 31 to three emotions (18.4%), and 27 for four emotions (16.0%). Finally, we only had 22 images classified as undifferentiated (13.0%).

If we compare our results with those presented in Mikels dataset (see Table V), we obtained more 6% of images classified with a single emotion, and less 8% undifferentiated images (29.0% (15.98+13.02) vs 36.9% in Mikels considering only three emotions [16]).

If we consider up to four emotions in an image, we have less 24% undifferentiated images (13.0% vs 36.9% in Mikels), while in the case of blended images we have around 21% more images (51.6% vs. 30.5% in Mikels).

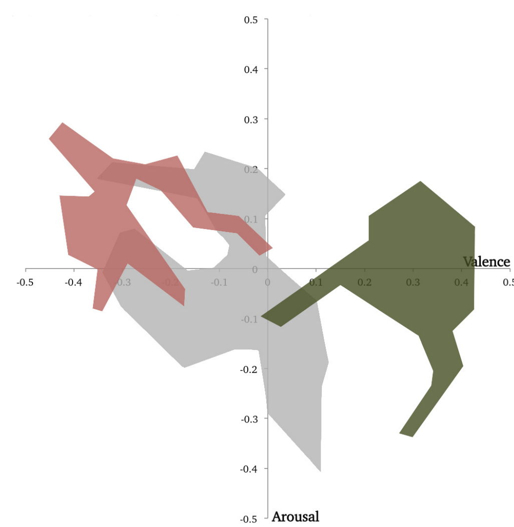


Fig. 7. Representation of the ratings in the valence/arousal space for each polarity. The red area (on the left) corresponds to the negative polarity. The grey area (at the center) corresponds to the neutral polarity, while the green (on the right) corresponds to the positive polarity. (best seen in color).

TABLE IV
DISTRIBUTION OF UL-EPS DATASET CONCERNING THE EMOTIONAL LABELS: SINGLE, BLENDED, AND UNDIFFERENTIATED.

Single	Blended 2	Blended 3	Blended 4	Undifferentiated
60	29	31	27	22
35.50%	17.16%	18.43%	15.98%	13.02%

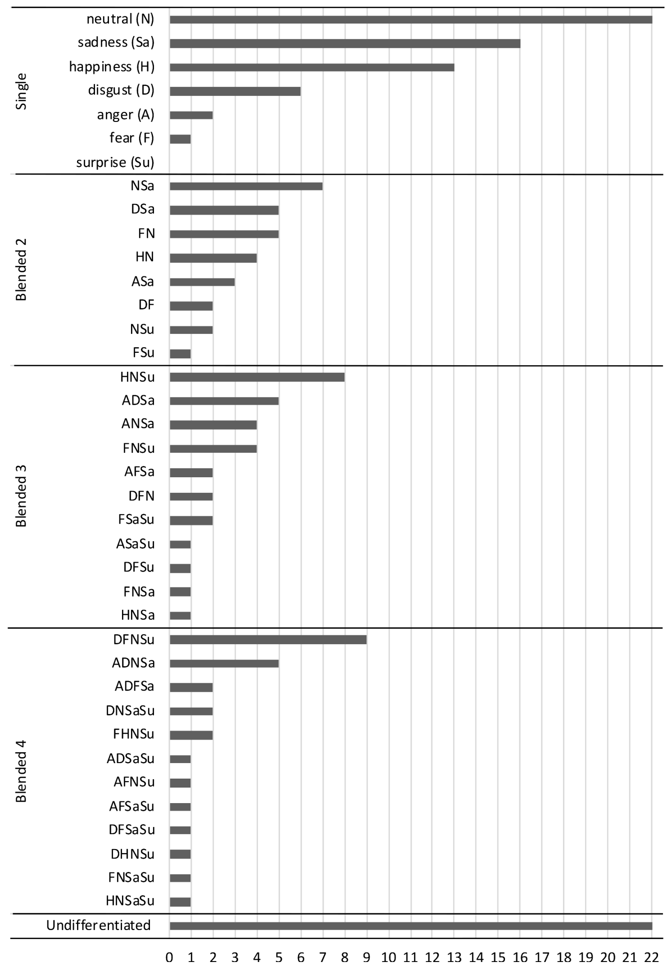


Fig. 8. Emotional labels that result from the classification process.

On the whole, we have a larger number of images annotated with emotional data. Thus, our dataset is more informative about the emotional labels assigned to images.

In Figure 8, we can see the different emotional labels resulting from the classification process. For the blended, we have for example DFSu, i.e., an image that contains the emotions disgust, fear, and surprise. The resulting label does not take into account the weight of each of the emotions present in an image, i.e., label DFSu includes the following combinations: DFSu, DSuF, FDSu, FSuD, SuDF, and SuFD.

E. Relationship Between Emotions

In Figure 9, it is possible to analyze the most elicited single emotions, and the relationship between two emotions. For that, we considered the frequency of occurrence of each emotion.

TABLE V
COMPARISON BETWEEN UL-EPS AND MIKELS DISTRIBUTION OF EMOTIONAL LABELS.

Dataset	Single	Blended (2 and 3)	Undifferentiated
ULEPS	35.50%	35.59%	29.00%
Mikels	30.00%	30.51%	36.92%

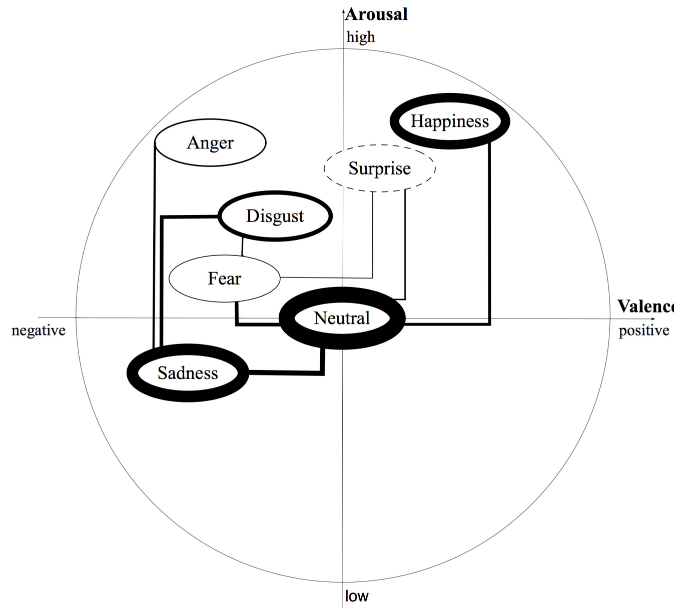


Fig. 9. Relationship between single emotions, and between the blending of two emotions.

The thicker the line, the bigger is the number of images that elicited that emotion (single) or the greater is the relationship between two emotions (blended). A dashed line indicates that there were no images that elicited that emotion.

The most elicited single emotions were neutral, happiness, sadness, and disgust. The most obvious relations are between the emotions neutral and sadness, neutral and fear, neutral and happiness, and sadness and disgust. Anger, fear, and surprise are the emotions less elicited alone. However, surprise tends to appear in conjunction with fear and neutral emotions. In the case of anger, there is some relation with the elicitation of sadness, while fear was elicited together with disgust, as well as with surprise and neutral.

Regarding the correlations between basic emotions, and considering Figures 8 and 9, we confirm the results reported in the literature. Happiness negatively correlates with all the other basic emotions. Anger shows correlation with fear and disgust. There is also correlation between sadness and fear, and between sadness and disgust. Finally, fear was also correlated with disgust. Overall, our results are in line with those reported in previous studies [26], [40]–[42].

F. Observations from Participants

During each session, participants were encouraged to share with us their opinions/comments about the experience. More than 40% of the participants mentioned some type of difficulty in understanding the content of some of the images, leading to confusion about their feelings.

The majority identified the lack of context as the main reason for this, e.g., some participants did not understand if an animal in front of a car will be hit by it or not. In this case there is confusion between feeling negative if the animal is hit, and neutral or positive otherwise.

Five participants claimed that surprise is subjective, difficult to understand, and also difficult to elicit from an image. There seemed to be some exceptions to this, such as a shark moving as it is attacking a person or images with unexpected content like a lamp or stairs. A couple of participants indicated us that none of the images was able to trigger anger.

Regarding the personal taste of the participants, some appreciated specific content such as snakes (4), spiders (3), or aquatic animals (1), while others did not appreciate it at all. However, some of them considered images with those animals "beautiful", mainly due to the colors in them. Three participants declared that they were not sensitive to some images, such as a children smiling, leading them to feel neutral, although they considered that they should feel "happy". Finally, some participants also mentioned that the emotional content of the previous visualized image may interfere in the way they were feeling at that moment.

VI. Conclusion

We described an unrestrained study performed with 60 participants to annotate a dataset of images with the polarity and discrete emotions elicited by each image. During our study there were no restrictions in the selection of the emotions, being possible for a user to associate a positive and a negative emotion to the same image.

We presented the relationship between multiple emotions that arouse when visualizing an image, and we verified that they were in line with existing literature. Moreover, we also presented our experimental procedure designed to avoid stress and fatigue, providing more comfort to the users.

We made a more complete and realistic picture dataset composed of 169 images publicly available to the community[3], as a new contribution to complement the already existing datasets. Each image was annotated with the emotional polarities (positive, neutral, and negative), discrete emotions (anger, disgust, fear, happiness, neutral, sadness, and surprise), and the original valence and arousal information.

Having in mind all the inherent subjectivity of emotions, the different constraints that could affect the participants judgement (current emotional state of the user, user's ability to evaluate what they felt, among others), the overall good agreement among participants, and between our dataset and the GAPED dataset, we can consider that the results achieved by our study are reliable and useful for the elicitation of emotions.

As future work, we intend to use our procedure to annotate more images with polarities, discrete emotions, and the physiological signals collected from the users while viewing the images.

Acknowledgment

This work was supported by national funds through Fundação para a Ciência e Tecnologia, under LASIGE Strategic Project - UID/CEC/00408/2013.

[3]http://www.di.fc.ul.pt/~mjf/research/ul-eps/UL-EPS_2018.xlsx

References

[1] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. Harper Perennial, 1995.

[2] R. Picard, "Affective computing," MIT Media Laboratory, Perceptual Computing Section, Tech. Rep. 321, 1995.

[3] P. Dunker, S. Nowak, and C. Lanz, "Content-based Mood Classification for Photos and Music," in *Multimedia Information Retrieval*, 2008.

[4] D. Joshi, R. Datta, E. Fedorovskaya, Q.-t. Luong, J. Z. Wang, L. Jia, and J. Luo, "Aesthetics and Emotions in Images [A computational perspective]," *Signal Processing Magazine*, 2011.

[5] D. O. Bos, "EEG-based Emotion Recognition: The Influence of Visual and Auditory Stimuli," *Capita Selecta Paper*, 2006.

[6] L. Axelrod and K. S. Hone, "Affectemes and allaffects: A novel approach to coding user emotional expression during interactive experiences," *Behaviour & Information Technology*, 2006.

[7] M. Tkalčič, A. Kosir, and J. Tasic, "Affective recommender systems: the role of emotions in recommender systems," in *Workshop on Human Decision Making in Recommender Systems*, 2011.

[8] K. C. Klauer, "Affective priming," *European Review of Social Psychology*, 1997.

[9] S. Wang and X. Wang, "Emotion semantics image retrieval: An brief overview," in *Affective Computing and Intelligent Interaction*, 2005.

[10] T. Li and M. Ogihara, "Detecting emotion in music," in *Music Information Retrieval*, 2003.

[11] K. Sun, J. Yu, Y. Huang, and X. Hu, "An improved valence-arousal emotion space for video affective content representation and recognition." Multimedia and Expo, 2009.

[12] V. L. Rubin, J. M. Stanton, and E. D. Liddy, "Discerning Emotions in Texts," in *Exploring Attitude and Affect in Text: Theories and Applications*, 2004.

[13] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance." *Behavior Research Methods*, 2011.

[14] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual." NIMH Center for the Study of Emotion and Attention, Tech. Rep., 1997.

[15] A. Marchewka, Ł. Zurawski, K. Jednoróg, and A. Grabowska, "The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database." *Behavior Research Methods*, 2014.

[16] J. A. Mikels, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz, "Emotional category data on images from the International Affective Picture System." *Behavior Research Methods*, 2005.

[17] A. Smith, "A new set of norms." *Behavior Research Methods*, 2004.

[18] ——, "Smith2004norms.txt," *Retrieved October 2, 2004 from Psychonomic Society Web Archieve*, 2004.

[19] M. Riegel, Ł. Zurawski, M. Wierzba, A. Moslehi, Ł. Klocek, M. Horvat, A. Grabowska, J. Michałowski, K. Jednoróg, and A. Marchewka, "Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE)." *Behavior Research Methods*, 2016.

[20] E. Fox, *Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions*. Palgrave Macmillan, 2008.

[21] R. D. Lane, E. M. Reiman, G. L. Ahern, G. E. Schwartz, and R. J. Davidson, "Neuroanatomical correlates of happiness, sadness, and disgust." *The American journal of psychiatry*, 1997.

[22] Y. Choi and E. M. Rasmussen, "Searching for images: The analysis of users' queries for image retrieval in American history," *Journal of the American Society for Information Science and Technology*, 2003.

[23] K. A. Olkiewicz and U. Markowska-kaczmar, "Emotion-based image retrieval - An artificial neural network approach," in *Computer Science and Information Technology*, 2010.

[24] A. Hanjalic, "Extracting Moods from Pictures and Sounds: Towards Truly Personalized TV," *Signal Processing Magazine*, 2006.

[25] C. E. Izard, *The Psychology of Emotions*. Plenum Press, 1991.

[26] S. Schmidt and W. Stock, "Collective indexing of emotions in images. A study in emotional information retrieval," *Journal of the American Society for Information Science and Technology*, 2009.

[27] R. Plutchik, "The Nature of Emotions," *American Scientist*, 2001.

[28] P. Ekman and R. J. Davidson, *The nature of emotion : fundamental questions.* New York : Oxford University Press, 1994.

[29] O. da Pos and P. Green-Armytage, "Facial expressions, colours and basic emotions," *Journal of the International Colour Association*, 2007.

[30] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-Time EEG-Based Emotion Recognition and Its Applications," *Transactions on Computational Science*, 2011.

[31] J. Posner, J. a. Russell, and B. S. Peterson, “The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology.” *Development and psychopathology*, 2005.

[32] J. Yoon, “Utilizing quantitative users’ reactions to represent affective meanings of an image,” *Journal of the American Society for Information Science and Technology*, 2010.

[33] J. Machajdik and A. Hanbury, “Affective image classification using features inspired by psychology and art theory,” *International Conference on Multimedia*, 2010.

[34] M. Wessa, P. Kanshe, P. Neumeister, K. Bode, J. Heissler, and S. Schoenfelder, “EmoPics: Subjektive und psychophysiologische Evaluationen neuen Bildmaterials für die klinisch-bio-psychologische Forschung,” *Zeitschrift für Klinischer Psychologie und Psychotherapie, Supplement, 1/11, 77*, 2010.

[35] D. Lundqvist, A. Flykt, and A. Öhman, *The Karolinska Directed Emotional Faces - KDEF.* CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.

[36] N. Tottenham, A. Borscheid, K. Elertsen, D. Marcus, and C. Nelson, “Categorization of facial expressions in children and adults: Establishing a larger stimulus set.” *Cognitive Neuroscience Society*, 2002.

[37] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *Journal of Behavior Therapy and Experimental Psychiatry*, 1994.

[38] M. Tkalčič, U. Burnik, and A. Košir, “Using affective parameters in a content-based recommender system for images,” *User Modeling and User-Adapted Interaction*, 2010.

[39] P. Arriaga and G. Almeida, “Fábrica de emoções : A eficácia da exposição a excertos de filmes na indução de emoções,” Instituto Superior de Psicologia Aplicada, Tech. Rep., 2010.

[40] I. Bretherton and M. Beeghly, “Talking about internal states: The acquisition of an explicit theory of mind,” *Developmental Psychology*, 1982.

[41] B. Fehr and J. Russell, “Concept of emotion viewed from a prototype perspective,” *Journal of Experimental Psychology: General*, 1984.

[42] P. Shaver, J. Schwartz, D. Kirson, and C. O’Connor, “Emotion knowledge: Further exploration of a prototype approach.” *Journal of Personality and Social Psychology*, 1987.