UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

Identifying Areas of Emotional Interest in Images Through Content Tags and Eye Gaze: A Study With Users

Cláudia Gervásio Desidério

Mestrado em Informática

Trabalho de Projeto orientado por:
Manuel João Caneira Monteiro da Fonseca

2021

Acknowledgments

I would like to begin by thanking my supervisor, Professor Manuel João da Fonseca, for his support, guidance and availability throughout this past year, which were essential for the completion of this dissertation. I would also like to thank my tutor, Soraia Meneses Alarcão Castelo, for her availability to clarify any doubts I had throughout the process, as well as the support given to the development of this work.

I would also like to thank my parents, Ana Cristina Gervásio and António Desidério, for being the people who supported me unconditionally during all this work. Without them, and without all the love, help and sacrifices they made throughout all these years, it would not have been possible to get where I am.

I would also like to thank my friends Bárbara, Beatriz, Cláudia, Eduardo, Miguel and Rita, for being available to support and help me whenever I asked.

Finally, I would also like to thank all the volunteers who participated in the three studies, without whom I would not have been able to carry out this work.

To my parents Ana Cristina and António

Resumo

Ao avaliarem uma imagem, as pessoas tendem a prestar atenção às várias zonas de forma seletiva. Esta atenção, é influenciada pelas propriedades específicas da própria imagem, como por exemplo a cor [17], mas também pela existência de rostos ou até texto [11]. Além disso, como tem vindo a ser constatado em várias pesquisas, esta é também influenciada pela capacidade da imagem conseguir provocar ou não uma reação emocional no seu observador [17]. Tirando partido desta capacidade de priorização da atenção, poder-se-á identificar a área da imagem responsável por determinada reação emocional e proceder-se à identificação do conteúdo que terá provocada essa reação.

Tendo isto em conta, procurámos através de este trabalho, tirar partido das reações emocionais experienciadas por um utilizador ao observar um conjunto de imagens, para as conseguir categorizar emocionalmente, e identificar o conteúdo específico responsável por essas reações. De modo a atingirmos estes objetivos, dividimos a nossa investigação em três fases distintas, cada uma das quais correspondentes a um estudo com utilizadores: i) identificação do conteúdo mais relevante em cada imagem; ii) categorização emocional das imagens, tendo em conta as reações emocionais que a sua avaliação provocaria, e também a identificação do conteúdo emocional da mesma; iii) identificação das zonas contendo cada conteúdo emocional registado na imagem. Devido ao contexto pandémico em que nos encontrávamos devido ao coronavírus SARS-COV-2, todos os estudos foram desenvolvidos online.

No primeiro estudo, procurámos identificar o conteúdo mais relevante. Para que tal fosse possível, procuramos identificar as cinco tags de conteúdo mais votadas pelos utilizadores. Como tal, de modo a possibilitarmos essa escolha, começamos por selecionar um conjunto de 252 imagens, representativas das seis emoções básicas de Ekman (anger, disgust, fear, happiness, sadness e surprise), através de um processo de seleção, que permitiu criar um dataset com um igual número de imagens para cada emoção e com formatos variados. Estas imagens foram depois avaliadas pelo modelo General da API Clarifai, o qual devolveu um conjunto de 30 conceitos representativos do conteúdo dessas imagens. Esses conceitos, foram depois filtrados por nós, de modo a ficarem apenas os 15 mais prováveis de se encontrarem em cada imagem, e selecionados posteriormente pelos voluntários deste estudo. No final, analisamos o nível de concordância entre utilizadores, cujo valor médio para o conjunto de imagens foi de 0.51, com um desvio padrão de

0.23, indicativo de uma concordância moderada para o conjunto do dataset. Além disso, a análise permitiu ainda averiguar a existência de uma concordância de moderada a muito boa para mais de 50% das imagens do dataset criado. Por fim, procedemos à verificação das tags selecionadas para cada imagem, e à identificação das 5 com maior quantidade de votos.

No segundo estudo, tivemos dois objetivos: i) verificar se existia alguma conexão entre as zonas da imagem olhadas durante mais tempo, no momento de visualização das mesmas e as reações emocionais experienciadas pelos utilizadores nesse momento; ii) verificar a existência de uma conexão entre as reações emocionais e o conteúdo da imagem, representado pelas 5 tags identificadas como as mais relevantes no estudo anterior. De modo que tal fosse possível, começamos por apresentar as imagens e respetivos conceitos aos utilizadores, os quais tiveram de realizar uma avaliação emocional de cada uma das imagens. Esta avaliação, incluiu não só a identificação da polaridade e emoção(ões) sentidas durante o momento de visualização, como também a identificação do conteúdo responsável pela(s) emoção(ões) experienciadas, através da seleção da(s) tag(s) de conteúdo adequada(s) às mesmas. Além disso, durante o momento de visualização das imagens, foram ainda retiradas as coordenadas do olhar do utilizador, de modo a perceber qual a zona que registou maior atenção do mesmo. Adicionalmente, foi ainda realizada a avaliação das expressões faciais do utilizador, enquanto visualizava a imagem. No final do estudo, verificámos para cada imagem qual a emoção e polaridades mais votadas, onde percebemos que existiam imagens associadas a cada uma das polaridades emocionais (negativa, neutra e positiva) e que, como esperado, para a maioria das imagens associadas a uma determinada emoção, a emoção mais votada seria aquela à qual se encontravam originalmente associadas. Contudo, verificou-se também a existência de imagens onde não houve concordância nem em relação à polaridade mais adequada, nem em relação à emoção, o que levou a casos com mais do que uma polaridade e emoções associadas. Além disso, verificou-se ainda, que no caso das imagens que se encontravam associadas originalmente a Anger, nenhuma foi associada a esta emoção, e que no caso das imagens de Surprise, apenas uma pequena percentagem de imagens, foi associada à emoção original. Para além disso, em ambos os casos a emoção mais votada para a maioria das imagens foi Happiness. Quanto à verificação das zonas que receberam maior atenção durante a visualização do estímulo, percebemos que para a maioria das imagens de todas as emoções, estas correspondiam ao centro das imagens.

Quanto ao conteúdo identificado como o mais relevante emocionalmente, analisamos o mesmo quanto ao tipo, polaridade e emoções mais votadas. Os resultados destas análises, mostraram que a maioria do conteúdo assinalado se tratava de conteúdo generalista, e que os votos para as polaridades e emoções foram um reflexo dos obtidos para as imagens. Adicionalmente, verificamos ainda se a polaridade associada ao conteúdo, era a esperada tendo em conta a emoção mais votada para o mesmo. Os resultados indicaram

que na maioria dos casos havia a correspondência esperada, com exceção de dois casos onde a polaridade negativa foi associada à emoção Happiness. Por fim, ao contrário do planeado, as informações resultantes do sistema de reconhecimento de expressões faciais, acabaram por ser descartadas, devido ao facto de o vídeo captado pela webcam dos dispositivos dos utilizadores, não ter permitido fazer uma avaliação desta informação. Por consequência, as informações provenientes deste software, não puderam ser comparadas com: i) os registos do eye tracker, de modo a perceber se a zona olhada mais tempo teria sido ou não a responsável pelas reações emocionais registadas; ii) nem com a emoção mais votada para cada imagem, de modo a perceber se a mesma era de facto a mais adequada. Desta forma, acabou por não nos ser possível completar o primeiro objetivo deste estudo.

No último estudo deste trabalho, tivemos como objetivo a identificação da zona emocionalmente mais relevante das imagens avaliadas, e perceber qual a zona com maior carga emocional. Para que nos fosse possível concretizar este objetivo, solicitamos a um grupo de voluntários, que procedesse à seleção das zonas de cada imagem, que melhor representavam cada um dos conteúdos emocionalmente relevantes, que tinham sido identificados no estudo anterior. No fim, os resultados obtidos foram por nós avaliados, de modo a perceber qual a concordância entre os vários utilizadores, em relação às zonas selecionadas. Esta análise, permitiu-nos perceber que existiu bastante variabilidade na escolha das zonas, o que resultou numa concordância fraca e até mesmo pobre, para a maioria das imagens. Além disso, procedemos ainda à identificação das zonas com maior carga emocional, ou seja, aquelas às quais foram atribuídas uma maior quantidade de votos, assim como também as zonas onde existia uma menor carga emocional, por serem as que possuíam menor quantidade de votos atribuídos. Os resultados obtidos desta análise, permitiram-nos perceber, que a zona mais votada para a maioria das imagens da maioria das emoções, ou na maioria das imagens com mais do que uma emoção atribuída, se tratava do centro (zona_E5). Contudo, existiram algumas exceções: Anger – em uma imagem era também o centro, e na outra existia um empate entre o canto inferior esquerdo e a zona em baixo ao centro; Surprise – em duas imagens voltava a ser o centro, em outras duas a zona_E6, e as restantes quatro imagens eram a zona_E1, zona_E2, zona_E3 e zona_E8 respetivamente; anger/disgust/sadness – zona_E4. Quanto às que possuíam menor carga emocional, na maioria dos casos associados à maioria das emoções, foi a zona_E3. Contudo, existiram novamente exceções: Neutra e Sadness – zona_E7; Surprise – zona_E3 e zona_E9; imagens com mais do que uma emoção associada – zona_E3, zona_E9 e zona_E7. Por fim, fomos ainda verificar se as zonas olhadas durante mais tempo correspondiam às que possuíam maior carga emocional. Ao contrário do que esperávamos, esta correspondência, apenas ocorreu em 64 imagens. Além disso, verificamos ainda a existência de seis imagens onde a zona olhada durante mais tempo, correspondeu a zonas onde não existia qualquer tipo de conteúdo emocional registado.

Quanto às restantes, existiu correspondência com zonas onde estava registado conteúdo, mas cujas zonas não eram as que possuíam maior carga emocional.

Deste trabalho, acabou por resultar um procedimento para a categorização de uma imagem de acordo com o conjunto de emoções básicas e universais definidas por Ekman e a emoção Neutral. O procedimento criado, permite também, a identificação do conteúdo emocional de cada imagem, e sua anotação tendo em conta o conteúdo considerado emocionalmente relevante para a mesma. Este procedimento, tira partido de informações como: i) avaliação emocional das imagens; ii) identificação do conteúdo emocionalmente relevante, através de uma tag de conteúdo; iii) coordenadas do olhar do utilizador registadas por eye tracking; iv) identificação das zonas da imagem que mais se adequam a cada conteúdo emocional.

Adicionalmente, deste trabalho resultou também um dataset composto por 252 imagens, categorizadas emocionalmente, e anotadas com dois tipos de informações: média de coordenadas registadas por eye tracking em cada zona da imagem e conteúdo emocional associado a cada uma das zonas, o qual vem acompanhado pelo número de votos.

Palavras-chave: emoções, imagens, eye tracking, tags de conteúdo, reações emocionais

Abstract

The attention of a user on an image is influenced by factors such as its specific properties, the existence of faces or text and the ability of the image to provoke an emotional reaction. This work, aimed to take advantage of the emotional reactions of an individual when observing an image, to categorize that image according to emotional reactions, and identify the concrete content responsible for them. Our work was divided in three studies with users, all developed online, due to the pandemic resulting from the appearance of the SARS-COV-2 coronavirus. In the first study, we attempted to understand, for each image, what could be considered the most relevant content. In the second study, we defined two objectives: i) to understand if there was a connection between the areas looked at for longer, and the emotional reactions; and ii) to verify if emotional reactions, were related to the most relevant content. In the last study, we tried to understand if the most emotionally charged zone corresponded to the zone looked at the longest. This work resulted in a procedure that allows us to: i) categorize an image according to the set of basic and universal emotions defined by Ekman and Neutral emotion; ii) identify the most relevant emotional content; and iii) annotate the image according to the most relevant content. This procedure, takes advantage of information such as: i) emotional evaluation of the image; ii) identification of the emotionally relevant content (content tag); iii) coordinates of the user's gaze registered by eye tracking; and iv) identification of the zones that best suit each emotional content. Additionally, a dataset was created with 252 images, emotionally categorized and annotated with two types of information: average eye tracking coordinates for each zone of the image and emotional content.

Keywords: emotions, images, eye tracking, content tags, emotional reactions

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we present our motivation for the development of this research, as well as the main objectives we intend to achieve with it. Furthermore, we describe shortly the studies performed and the main contributions of our work.

1.1 Motivation

When evaluating an image, humans have the ability to pay selective attention to certain areas of that image. This attention often falls on a certain property of the image, such as colour [17], but also on faces and text that is present in the image under evaluation [11]. Some researches have also been indicating over the years that the user's attention during the evaluation of an image can also be influenced by its emotional relevance. As such, an image that has an object or scene capable of eliciting an emotional response from an observer will tend to provoke a greater attention from the observer than an image that is emotionally neutral to the observer [17].

By taking advantage of this ability to prioritize attention, it is possible to understand which area of the image received more attention, and consequently identify the type of content that was responsible for triggering the observer's emotional reaction. Taking advantage of this information, it is then possible to classify the image through an emotional tag, which is in accordance with the emotion triggered by its observation. By integrating this kind of information, for instance, in the context of a photo presentation application, it would be possible to make the images be presented in a certain order, which would depend on the user's emotional state.

With this idea as a starting point, this work aimed to understand how to take advantage of an individual's emotional reactions, to classify images according to the emotions resulting from their observation, and also to correctly identify the content responsible for those reactions. In order for that to be possible, we defined three distinct phases for our research: i) identification of the most probable content in each image; ii) classification of the images according to the reactions triggered by them; iii) identification of the location

where the emotional content responsible for the emotional reactions is present.

1.2 Objectives

Through this work, we propose to achieve two main goals: i) understand if a photograph, used as a visual stimulus, is able to provoke an emotional reaction in its observer; ii) in case there is an emotional reaction, identify the specific content that could have been responsible for that reaction. In order to achieve these two objectives, we will divide our approach into three distinct phases, which will have differentiated objectives:

1. Content Tags - firstly we want to identify what type of content is found in the various images and understand which might best describe each of them;

2. Emotional Reactions and Emotional Tags - next we want to understand if users develop some emotional reaction when looking at the images, and try to understand what content may have been responsible for them, using an eye traker and the content identified by the users themselves through content tags;

3. Tags location - finally, we intend to understand which might be the most emotionally relevant areas, correlating the areas looked at for the longest time with the areas considered most relevant for each of the emotional contents present in the images.

1.3 Studies Performed

As mentioned, our work was divided into three phases. Each of these phases corresponded to a study with users, which, due to the pandemic context in which we were living, resulting from the appearance of the SARS-COV-2 coronavirus, were entirely developed online. As mentioned in the previous section of this chapter, each of these studies/phases had its specific objective, which were described above. In order for these objectives to be achieved, throughout each of these studies, we sought to obtain the necessary information by implementing procedures developed for this purpose.

In *Study I: Content Tags*, we started by obtaining a set of images representing Ekman's six basic and universal emotions, which were selected from the EmotionROI dataset [34]. This selection process, which was composed of a set of criteria, resulted in the selection of a set of 252 images, 42 of each of the six emotions. The set of images was then analysed by the API Clarifai General model, which returned a set of 30 concepts representative of the content of each image, which were accompanied by the probability of their presence. The concepts obtained went through a filtering process carried out by us, which allowed the identification of the 15 most representative concepts of the content of each image. After this initial phase, the images and respective concepts were presented

to users through an online platform, which allowed the selection of between three and 10 concepts that best represented the content of each image analysed. After the end of the study, we identified the five most voted tags for each image. The process of identification was composed of several selection criteria. Although it worked for most of the images, there were two exceptional cases, for which it was necessary to select a tag, in order to obtain the five desired tags. Taking the results obtained, and in order to verify their quality, we performed two analyses. The first analysis consisted in verifying the average number of tags selected per user for each image at a global level of the dataset, and its respective standard deviation. This same analysis also allowed us to have a notion of was been the minimum and maximum average number of selected tags. Finally, we also verified what was the level of agreement between users, in relation to the tags chosen for each image. This analysis was performed using the Fleiss Kappa inter-rater agreement measure.

In *Study II: Emotional Reactions and Emotional Tags*, we started our procedure by preparing the online platform in which the study would be conducted. This platform was prepared to allow users to view the images and perform their emotional evaluation. The emotional analysis was divided into three steps: i) identification of the most adequate polarity for the image; ii) selection of the emotion(s) felt during the image analysis; iii) selection of the representative tag of the content responsible for each of the emotions identified. Besides enabling this evaluation, the platform also had the ability to extract three additional types of data: i) coordinates of the user's gaze during the viewing of each image; ii) the user's facial expressions during the viewing period of the images; iii) a video of the user's face, if the user authorized it. To obtain the first two pieces of information, two different softwares were used: WebGazer, an eye tracking library that only depended on the user's browser and device webcam to work; and Face-api js, a machine learning library that takes advantage of Convolutional Neural Networks (CNNs) to analyse in real-time the facial expressions. After obtaining the results, we identified which were the three or more tags, considered as the most emotionally relevant for each of the images. This identification was done by checking which tags were the most voted for each image. Next, we verified the emotional evaluation performed to each image. This verification, started with the analysis of the most voted polarity for each image, followed by the verification analysis of the most voted emotion. In the next step, moving on to the analysis of the emotional content reported by the users, we started by verifying which tags were associated in more quantity to each one of the polarities, having also verified which ones were more associated to each one of the six emotions of Ekman and Neutral. Still taking the emotional content indicated, and identifying the tags as being globally the most relevant, we tried to understand what kind of tags they were, and to understand if the most associated polarity to each tag corresponded to the expected one, taking into account the emotion most linked to it. Finally, taking the data registered by the eye tracker, we also investigated which were the most looked at zones during the analysis of each image.

Contrary to what we had planned to do, we could not analyse the data resulting from Face-api js, since it could not analyse the facial expressions in the desired way, due to problems with the image captured by the users' webcam.

In *Study III: Tags Location*, we also started by preparing an online platform for the study. This platform allowed the user to analyse each of the images and mark for each of the emotionally relevant tags associated to them, the zone(s) that in their opinion best represented that content. After the end of the study, we started by verifying the quality of the results obtained, through the inter-rater agreement Fleiss Kappa measure. After this analysis, we then moved on to the two analyses that would allow us to identify the possible most emotionally relevant area. In the first analysis, we identified the most voted zones and the least voted zones. The analysis was divided into two phases: i) in which we analysed the results obtained at a global level; ii) in which we differentiated them, according to the emotions with which the images were associated during the second study. Complementarily to this analysis, we created a heatmap for each image, to make it easier for us to visualise not only the distribution of the emotional content throughout the various zones that composed the image, but also to identify more easily the zone(s) to which the greatest/least emotional charge was associated. In the last analysis, in order to fulfil the objective we had proposed, we compared the area that had been looked at for the longest time, and which could therefore be considered the area of greatest ocular interest in the image under analysis, with the area(s) to which most emotional charge was associated. This analysis allowed us to understand if there was a correspondence between these two types of zones. Having verified cases in which that did not happen, we also tried to understand why that happened, by checking which zone the zone of greater ocular interest would have coincided with, and which emotion the image was associated to. Complementary to this analysis, in order to be able to compare these two types of zones visually, we also built heatmaps with the average gaze coordinates registered by the several zones of each image, and compared them with the heatmaps created for the visualization of the emotional content distribution.

1.4 Contributions

Through this thesis, we contribute with:

- Procedure that allows classifying an image, according to the spontaneous emotional reactions of the observer at the moment of evaluation;
- Process for the identification of the emotional content responsible for the recorded emotional reaction, and the identification of the place where it was located at the moment of the evaluation. For location identification, the process designed by us allows crossing three types of information: i) eye tracker data at the moment of image

viewing; ii) tag of emotional content selected at the moment of image evaluation; iii) selection of the zones that best fit that content;

- A dataset composed of 252 images, which are accompanied by two types of heatmaps for each image: i) eye tracker heatmaps - where the average number of coordinates for each zone is registered; ii) heatmaps of the zones selected for emotional content - where the colours of each zone correspond to the total number of votes registered for the image, and in which the tags associated to it are also indicated.

1.5 Structure of the document

This document is organised as follows:

- Chapter 2 - Background and Related Work: In this chapter we present some of the progress that has been made in the various areas that make up this research, and briefly describe some concepts important to its development;
- Chapter 3 - Methodology: In this chapter, we present the main software tools used in the three studies that were part of our research, as well as a brief description of the experimental approach used and the reasons that led us to its implementation;
- Chapter 4 - Study I: In this chapter, we describe the first study developed within the scope of our research, in which we aimed to select the set of content labels that best represented the images assessed;
- Chapter 5 - Study II: Emotional Reactions and Emotional Tags: In chapter 5, we describe and analyse the second study of this research, which aimed to relate the emotional reactions of the volunteers when analysing the images given as visual stimuli, with the content present in each one and that was identified by the users through the selection of content tags;
- Chapter 6 - Study III: Tags Location: Chapter 6 - Study III: Localization of Labels: This chapter describes the execution and analysis of the last study carried out for this research, which aimed to identify the zone of the images that could be considered as the most emotionally relevant, whose verification was carried out by comparing the zone that was looked at the longest with the zone with the most emotional charge;
- Chapter 7 - General Discussion of the Research: In this chapter we make a general evaluation and discussion of all the results obtained in our research;
- Chapter 8 - Conclusion: In this last chapter, we will present the main conclusions resulting from the research carried out.

Chapter 2

Background and Related Work

In this chapter we briefly describe some of the works that have been done in the various areas related to our work. We start by describing what emotions are and how they can be classified. Next, we present some of the databases that have been used in the area of emotion analysis, presenting their advantages and disadvantages. We also present an explanation of what Tags are, how their notation can be performed and how they can be interpreted so as to enable us to better understand the multimedia content from which they were generated. After, we talk about Eye Tracking, where we describe what this technology consists of, how the data obtained from it can help us to verify which areas of interest of the multimedia content were analysed by the user, and the types of existing eye trackers. Finally, we present some Facial Recognition Systems, indicating what they are and the various types of systems that have been used in this area.

2.1 Emotions Representation and Emotional Polarity

Human beings possess the ability to categorize the stimuli to which they are exposed, into emotional categories. These categories can be based on basic emotions, on a dimensional approach or be based on evaluation criteria [9]. This capacity, made possible the prioritization of our attention, under emotional contents in detriment of non-emotional contents [17]. As a consequence, it enabled the processing of information according to its relevance, and a faster and more adapted response to each situation [9].

Polarity is a binary contextual factor for classifying stimuli that fall within a conceptual dimension, composed of two opposite poles, positive and negative [23]. In the specific case of emotional polarity, whose concept is based on the notion of linguistic polarity [20], it allows the classification of the target emotional stimuli as positive or negative, according to the emotional response provoked by them.

Emotions, on the other hand, are considered complex states of mind, which include physiological correlation, social roles, and cognitive factors, that together are responsible for an individual's reactive behavior [14]. Over the years, emotions have been evaluated

according to three perspectives: discrete (or categorical), dimensional and componential. While the discrete perspective bases its evaluation on the characteristics that distinguish one emotion from another, the dimensional perspective takes the approach of identifying emotions according to their position in relation to a defined number of dimensions [25]. The componential perspective, on the other hand, evaluates emotions according to various organizing systems, also known as components [8].

Discrete perspective is based on a set of basic and universal emotions, whose external manifestation is independent of the culture and personal experience of the individuals who manifest them. For this reason, these emotions can be revealed through facial expressions, without the need for verbal manifestation [14]. According to Ekman [16], these so-called basic emotions follow two basic ideas: 1) they differ from each other in several important aspects, which include not only expression but also evaluation, preceding events, behavioral responses, and physiology; 2) but, they share several common characteristics - rapid onset, short duration, non-imposed occurrence, automatic evaluation, and coherence among responses - that allow human beings to cope with their fundamental tasks without the need for very elaborate planning. This author defined 6 basic emotions: anger, disgust, fear, happiness, sadness and surprise. Other authors have also approach this subject, such as Plutchik[35], who based his perspective on evolution, just like Ekman had done. Plutchik proposed 8 basic emotions: anger, disgust, anticipation, joy, trust, fear, surprise and sadness.

From a dimensional perspective, emotions are evaluated according to three dimensions: valence, arousal and dominance. Valence is responsible for indicating which motivational system is activated [29], ranging from highly positive to highly negative [27]. Arousal, on the other hand, is responsible for evaluating the intensity of the activation of this system [29], ranging from exciting to relaxing. In the case of dominance, it represents the degree of control over the effective stimulus, thus evaluating the reaction as being under control to out of control [27]. Despite the existence of these 3 dimensions, in most cases only the first two dimensions are used. This is because the most used model is the Circumplex model of Affect (CMA), introduced by Russell et al. [36] This model proposes that all affective states are consequences of cognitive interpretations of central neural sensations, resulting from two independent neurological systems.

The componential perspective, proposed by Scherer [13], includes five organic systems: cognitive evaluations, physiological reactions, behavioral tendencies, motor expressions, and subjective feelings. According to this perspective, the individual's organism is continuously evaluating stimuli and responding to the most important ones, which according to the same theory are the emotions. This author proposes, the Semantic Emotional Space, which is based on Russel's CMA, and is divided into the following dimensions: pleasure, arousal, control and conductiveness [8].

Among the various perspectives presented above, dimensional and discrete are the

two most commonly used. In the dimensional perspective, the most widespread model is Russel's CMA, while in the discrete perspective, the six basic and universal emotions defined by Ekman are the most used.

2.2 Image Databases for Emotional Analysis

In order to address the problem of the limited number of images for specific themes, several datasets have appeared in recent years. From these datasets that have been emerging, we can highlight the following: Affective Picture System (IAPS) [24], Geneva Affective Picture Database (GAPED) [15], Nencki Affective Picture System (NAPS) [27], Mikels [29], NAPS Basic Emotions (NAPS-BE) [37] and EmotionROI [34].

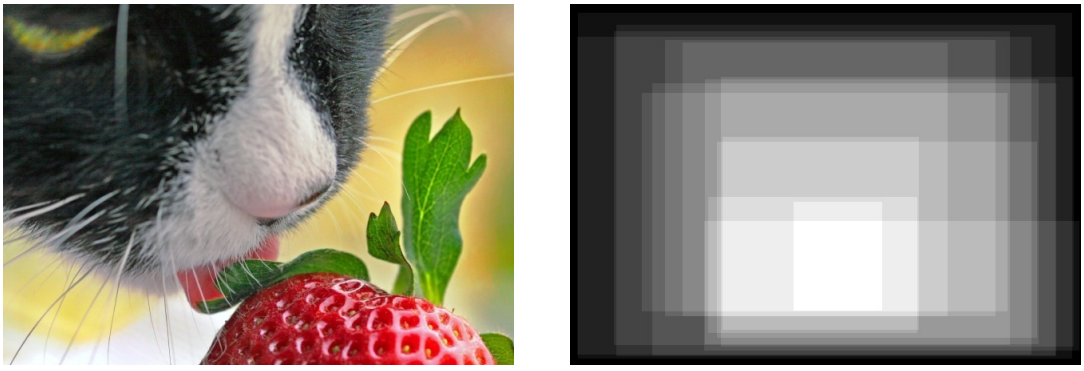
The IAPS, emerged in 1997, and was developed with the intention of bringing together in a single dataset, standardised, emotionally evocative and internationally accessible photographs that brought together a wide range of semantic categories, which would enable experimental research in the areas of emotion and attention. Despite its extensive use, this database has some limitations and problems: it only provides valence, arousal and dominance [29], low quality images. Moreover, the intensive use of its images means that there is a decrease in the impact they have on participants in studies using them [4]. In order to overcome these limitations, GAPED was created in 2011 with the aim of increasing the availability of emotional visual stimuli. This new database, despite having some information about the emotional polarity of the images (negative, neutral or positive), is not effective when it is necessary to use discrete emotions [4], (e.g. happiness), unlike the dataset mentioned above [29].This is due to the fact that discrete emotions are considered a subordinate division of the two-dimensional valence and arousal system, present in the first dataset. By being dependent on this system, the dataset in question allows through valence the identification of which motivational system was activated, and from the perspective of arousal to verify the intensity of this activation [29].

In 2014, the NAPS emerged, which, as in the case of the previous dataset, aimed at eliminating problems in other databases, such as a limited number of specific categories of stimuli or low quality of images used as visual stimuli[27]. With this database came the division of images into 5 categories: people, faces, animals, objects and landscapes. NAPS was mainly standardized for the affective dimensions valence, arousal, and approach-avoidance, not taking into account the discrete emotions expressed by the images. As such, in order to evaluate them, and to provide researchers with a set of discrete and reliable emotional norms for a set of images, the same authors created NAPS-BE in 2016. In this new database, the photographs were divided into three distinct sets, according to their value in positive, neutral and negative, which avoided the mixture of emotions [37].

Another database that contributed to reduce the lack of emotional information was the

database created by Mikels et al. in 2005. Taking advantage of the IAPS database, it collected descriptive emotional category data from IAPS subsets in order to identify images that most emphatically evoke a specific discrete emotion, and make these images suitable for investigating a particular discrete emotion [29]. Although innovative, as mentioned by Alarcão and Fonseca [4], both the NAPS and GAPED database present similar procedural problems: pre-selection of discrete emotions for each of the images analysed by the participants, causing a priori a restriction of the results. This type of restriction conditions the results obtained, since what for one participant can lead to a positive emotion, for another it can awaken a contrary emotion, since these are dependent on each person's experiences.

Based on this idea, and in order to better understand the regions that evoked a certain emotion, the Dataset EmotionROI emerged in 2016. This dataset, by Peng et al. [34], was based on a set of images from the Emotion6 database, which through a study with users, made it possible to identify the regions in the images that most influenced the evocation of one of the 6 emotions defined by Ekman (anger, disguste, happiness, fear, sadness and surprise) and consequently to make a forecast of the Emotional Stimulus Maps (ESM) (Figure 2.5).



(A) (B)

Figure 2.1 – (A)Example of a image for emotion Joy from dataset EmotionROI; (B) Groundtruth of the example image for Joy of the dataset EmotioR

The ESM results from the average among the selections coming from a user study, and the results of the observations that best capture the regions inducing a particular emotion. As such, this dataset is currently composed of a set of images, which are divided by the 6 emotions defined by Ekman, as well as a ground thruth ESMs corresponding to each of the images stored in it.

Reference	Name	Number of images	Short Description
[23]	IAPS	1178	Dataset that includes images of animals, catastrophes, objects, people, etc
[14]	GAPED	730	Dataset divided by the polarity of the images: negative (spiders, snakes and scenes appealing to moral violence or legal norms), neutral (inanimate objects) and positive (baby animals, human babies and nature scenes)
[27]	NAPS	1356	Dataset composed of photographs from 5 categories: people, faces, animals, objects and landscapes
[37]	NAPS-BE	510	Dataset composed of subset of images coming from NAPS, chosen for explicitly evoking discrete emotions
[27]	Mikels	330	Dataset composed of a subset of images selected from the IASP
[34]	EmotionROI	1980	Dataset composed of a set of photographs divided by each of the six emotions defined by Ekman, which are accompanied by the respective Emotional Stimulus Maps (ESM)

2.3 Tags

Tagging means the annotation of multimedia content through keywords, known as tags, which aims to optimize the search for information using these tags as a starting point [31].

There are two types of tagging: Explicit Tagging and Implicit Tagging. Explicit Tagging is a form of annotation where multimedia data is annotated by the users themselves through keywords, which allow an optimised search of its content based on the associated tags. Implicit Tagging, refers to the exploration of information provided by the users' non-verbal reactions (for example, a smile) when viewing a given multimedia content, and the formation of a tag based on that reaction. While in the case of Explicit tagging, the data is only annotated when the users decide to associate some kind of tag to the content, Implicit Tagging, by contrast, is a spontaneous action that allows the annotation of a data during the users interaction with the content, and which is based on the reactions expressed by them. By relying on non-verbal behaviour, Implicit Tagging, can provide valuable clues about the type of content the users' are viewing, and provide effective tags about it. For this reason, this type of tagging has been used as a complement or even a replacement of Explicit tagging, which has revealed some inaccuracy for being based only on the individuals' interpretation of the content presented, as well as on personal and social needs [31].

When generated by users, i.e. through Explicit tagging, tags can be evaluated according to the type of abstraction associated with them. There are three levels of abstraction: subordinate, basic and superordinate [38]. Subordinate, the lowest level of abstraction, includes terms that are typically specific to the content being evaluated [8], and therefore have characteristics that overlap with other categories of terms (e.g., kitchen chair, which shares most of its attributes with other types of chairs). The basic level includes categories of attributes that are common to all or at least most members of a given category (e.g., chair, car) [38]. Finally, superordinate, includes more generic terms [8], which share only a few characteristics in common [38], and may even have different meanings for each

person (e.g., animal, furniture, vehicle) [8].

Once generated, these tags allow describing the multimedia content associated to them, and consequently a better interpretation and evaluation of it. As such, it is therefore important to interpret these tags as well, since this interpretation values and allows a better assessment of the analyzed content. This interpretation, can be carried out based on two frameworks specialized in describing art content: the framework established by Panofsky in 1972 [16], and the Shatford framework in 1986 [40], which is a reformulation of the previous one [18].

In his framework, Panofsky first established three levels:

- pre-iconography (General): general description of the objects and actions represented in a given image;
- iconographical (Specific): specific identification of objects, people, ideas, themes or concepts in a given image [40]. This level requires some background and cultural knowledge of the individual, since it is necessary that he/she understands the context in which the image is inserted to be able to identify what and who is represented [18].
- iconology (Abstract): identification of the intrinsic meaning of the content of an image, which requires a synthesis of the information obtained through the two previous levels [40]. This is considered the least concrete level, in which mythical creatures, symbolic representations and emotions are included [18].

Based on the work developed by Panofsky, Shatford tried to develop a theory that would allow the classification of images according to their theme. With this theory, the author intended to be able to identify and classify the types of themes present in an image, taking advantage of the answer to the following questions:

- Who?: which beings or objects are found in an image;
- What?: what events, actions, conditions and emotions are represented in an image;
- Where?: what sites, or places are displayed in the image;
- When?: the date on which the evaluated image was produced as well as the time and period represented in it.

Note that when making the evaluation of an image, not all questions need to be answered, since they are only carried out in order to avoid neglected topics [17]. The four questions presented above were added perpendicularly to those previously defined by Panofsky, resulting in a matrix composed by 12 categories [18], which can be visualized in Table 2.1.

Questions	Pre-Iconography (General)	Iconographical (Specific)	Iconology (Abstract)
Who?	Types of beings, objects: Cat, Tree	Named beings or objects: Portuguese, Holy Family	Mythical beings: Ogre, Unicorn
What?	General events: Birth, Game	Specific events: Christmas, Discoveries	Emotions or abstraction: Sadness, Happiness
Where?	Type of location: Beach, Basement	Specific location: Lisbon, Porto	Place symbolized: Heaven, Hell
When?	Cyclical time: Spring, Daytime	Specific time period: Battle of Aljubarrota, Renaissance	Time symbolized: Youth, Oldness

Table 2.1 – Table Panofsky/Shatford matrix adapted from [18]

2.4 Eye Tracking

The term Eye-tracking refers to the technology that is able to measure and register the eye movements of an individual before a stimulus, whether the stimulus occurs in a real or controlled environment. It is possible to determine: i) the areas in which the user fixed its attention; ii) how long the gaze was fixed; iii) the order in which the visual exploration occurred [7].

A visual stimulus is considered to be any object (e.g., an image), which is necessary to perform a task and whose visual perception by the observer is responsible for triggering its cognitive processes and, ultimately, also some actions.

The eye gaze data is usually analysed taking into account stimulus areas called Areas of Ocular Interest (AOIs). These areas vary from observer to observer, since what is an area of interest for one user, may not be for another participant [39]. Eye gaze data can be evaluated taking into account three pieces of information:

- Fixations: these are considered spatially-stable eye movements, width a duration between 100-300ms, depending on the stimulus. During this movement, the visual attention is focused on a specific area, and it is at this moment that cognitive processes are triggered;

- Saccades: continuous and rapid eye movements, lasting between 40-50ms. This type of movement occurs between fixations, but does not have the capacity to provide great visual perception;

- Pupillometry: the pupil is the orifice of the eye through which light enters and whose dilating action is controlled by the iris muscle [39]. When exposed to light, it has a diameter of about 3 mm (this diameter may vary from 1.5 to 9 mm). The

analysis of this diameter as a function of the user's cognitive activity allows conclusions to be drawn regarding the emotional response. For example, if it is dilated, it may mean that the participant is exposed to a positive or negative stimulus, and not to a neutral image [41].

Taking the first two pieces of information, it is possible to form the path of the visual scan and order it chronologically. From it, it is possible to take information such as the fixed zones or even the AOIs present in the evaluated stimulus [39].

Some research in psychology has concluded that the acquisition and processing of information occurs mostly during fixations. Furthermore, they have also shown that a small set of fixations is sufficient for the visual stimulus to be processed [39]. As for the interpretation of the fixations obtained, this is dependent on the context in which they were formed. As indicated in [39], a high ratio of fixations in a given AOI, could mean a high interest of the user in the content that is in that area. However, on the other hand, this ratio could also be an indication that the user had a greater difficulty in interpreting the content in that area [39].

Based on this information, it is possible to conclude, as mentioned in [39], that the eye gaze data allow us to obtain information such as the areas that held the user's attention the most during the stimulus analysis, as well as the effort made by the user during this process, and consequently the duration of this analysis [39].

This data, as mentioned earlier, is obtained through eye-tracking systems. Currently, there are three types of eye-tracking systems: i) mechanical systems, which include systems based on contact lenses, and which have integrated mirrors; ii) electronic systems, which take advantage of the electrical powers measured from contact electrodes, which are positioned near the user's eyes; iii) video systems, non-intrusive systems generally used in fixed-eye observations [7]. These systems are integrated into devices called Eye Trackers.

There are several types of Eye Trackers on the market,which vary among themselves in shape and in the methods used for tracking the user's eyes. Of the various types available, we can highlight two in particular: i) intrusive ones, which need to be integrated in a physical device, which is then carried by the user in order to track its gaze; ii) non-intrusive, also known as remote eye-trackers, which record the user's gaze from a distance and which are usually integrated into a monitor[7]. The eye tracker used in this research is included in the latter category.

Despite the different ways in which the various types of devices act, they generally have in common the fact that they require prior calibration and explicit configuration, besides typically being dependent on an infrared camera, which is placed at a fixed distance from the user, and being associated with high costs, which can reach thousands of dollars [33]. As such, more economical alternatives have been sought, such as introducing these technologies in webcams and offline software. Although older studies indicate that these

devices are less effective in detecting the gaze, recent advances have allowed an improvement in the resolution and processing speed allowing of these solutions to work in real time [33].

2.5 Recognition of Facial Expressions

The facial expressions of a user, are considered the most important non-verbal channels to express the internal emotions and intentions of a user [19]. Their detection is performed through Facial Recognition Systems [26], in which are integrated models that have the ability to perform Facial Expression Recognition (FER). The Facial Recognition process implies the capture of the image of a face, from a digital image, a video frame or even a surveillance camera [26][28], and the comparison of the information obtained with previously existing data [28].

Despite the advances occurred in the last years in this area, this type of Face Recognition Systems, in general, have some limitations that can influence the correct collection of information, of which we leave some examples: pose of the head, change in the illumination, facial expressions, aging and occlusion due to the presence of accessories in the image of the user [28]. In addition to these limitations, the systems that have the ability to recognize facial expressions, also have as an additional limitation, the difficulty of detecting facial expressions in videos. This difficulty comes from the fact that facial expressions have a dynamic pattern, which is divided into three stages: onset (beginning of the expression), peak (maximum intensity of the expression) and offset (moment when the expression disappears). As in most cases this pattern occurs rapidly, and the identification of the expression associated with this pattern becomes extremely difficult and challenging [19].

Several models have been presented over the years to obtain FER. Most of them use fixed and independent images, and completely ignore the temporal relations between consecutive frames. This limitation made it impossible to build sequences that could recognise subtle changes in the appearance of the user's face image. More recently, new models emerged based on Deep Neural Networks (DNNs) [19]. Unlike traditional models, which use engineering features to train the classifiers that evaluate facial expressions, DNNs are able to extract more discriminative features, which allow a better interpretation of the human face [19].

2.6 Discussion

Through the analysis of all the information described in this chapter, we are able to better understand the areas that make up or relate in some way to our work, and to define the approach to be used in our research.

Starting our analysis with the Emotions Representation and Emotional Polarity section, we verified that the visual stimuli to which a user is exposed, can be classified according to emotional categories, which could be of several types: categorization according to basic emotions, using a dimensional approach or through evaluation criteria. Starting with dimensional classification, more specifically Emotional Polarity, the approach in question allows stimuli to be classified into negative or positive, taking into account the emotional response they provoke in the user. However, in our view, a third polarity should be introduced in this classification: the neutral polarity. The introduction of this polarity, would enable the user to indicate that the stimulus to which he had been exposed, had not provoked any type of emotional reaction. By introducing this possibility, we would be allowing the user to be as correct as possible in his evaluation, since we would not be forcing him to choose a polarity that was not coherent with the emotional reaction expressed. Moving on to the categorisation according to emotions, we found that this could be performed according to three perspectives: discrete (depends on the classification according to basic and universal emotions), dimensional (classifies emotions according to the dimensions of valence, arousal and dominance); and componentional (takes into account 5 organ systems and classifies emotions according to the dimensions of pleasure, arousal, control and conductiveness). Based on the information obtained, we concluded that the most appropriate perspective to our approach would be the discrete perspective. This decision was based on two reasons: i) it would allow for a real-time assessment of our users' emotional reactions to the stimulus to which they had been exposed; ii) it is solely and exclusively dependent on the interpretation/classification given by the user to the stimulus to which he/she was exposed.

Although the two types of classification mentioned above allow for a correct assessment of the stimuli, both of them have, in our view, the problem of not producing a complete assessment of the emotional stimulus. As such, we opted in this research approach, to use both together in order to complement each other and obtain a more complete evaluation.

Moving on to Tags, we found that their annotation would bring us many advantages, as it would allow the annotation of the evaluated multimedia content. As previously mentioned, its annotation can be performed in two ways: Explicit Tagging (annotation of the content by the user itself) and by Implicit Tagging (provides effective tags, which give clues about content viewed by the user). Although the latter has over the years been implemented as a complement to the former, in the approach used in our research, we chose not to implement the two together, as we wanted to obtain each of our users' personal interpretation of the stimulus we have exposed them to. However, both types of annotation will be used, but at separate points in our research.

We will take advantage of the first type of annotation, in two distinct moments: i) first study - indication of which content tag best describes the content of each image evaluated;

ii) second study - during the emotional evaluation of each image, to indicate which content may have been responsible for a given emotional reaction, at the moment of visualization of the image being evaluated. Despite the fact that throughout our research we did not find any studies previously developed in this area, which correlated the description of the content of an image through tags with the emotional evaluation of the same, we proceeded with this approach, as it will allow us to understand if the information responsible for a given emotional reaction is the context transmitted by the image, or if, on the other hand, it is a specific object represented in the image. As for the second type of annotation, it will be present during the second study, through the API responsible for the Recognition of Facial Expressions.

Turning now to Eye Tracking and correlating it with the study of emotions. This technology is considered very advantageous, as it allows researchers to track the user's gaze, and thus to trace with greater precision the path of the visual scan carried out by the user and to order it chronologically. By allowing this tracking, it makes it possible to identify the various areas looked at during the analysis of a stimulus, which may enable a better interpretation of the emotions expressed by users. As previously mentioned, the operation of this technology depends on its integration in a device, the Eye traker, for which there are two alternatives: i) intrusive - require a physical structure and transportation; ii) non-intrusive - track the gaze remotely. The first option, in our view has two major disadvantages: i) they need to be carried by the user; and ii) they may cause discomfort to the user, since this type of device stays in direct contact with the user in order to track his/her gaze. On the other hand, non-intrusive devices have as major advantages, which meet the previous disadvantages: i) they do not need to be carried; and ii) they do not need to be in direct contact with the user to perform the tracking, leading them to be more easily accepted by users. Within this last category, there are some Eye trackers that only rely on webcams and offline software. As indicated in this chapter, in the first years these devices were associated with lower efficiency. However, they have been undergoing improvements, which have allowed a greater dissemination of this technology, since for its use it is only necessary to implement it in the platform where it will work. In addition, they have lower costs compared to other alternatives, since it is not necessary to have an extra physical device.

As such, given these reasons, and also taking into account the pandemic situation of COVID-19 in which this research was conducted, we chose to use a non-intrusive Eye Tracker, whose technology only requires the use of a webcam device.

Finally, the Recognition of the User's Facial Expressions, is performed taking advantage of Facial Recognition Systems, in which are integrated models capable of performing Facial Expression Recognition (FER). Along the years, several models have been presented for its recognition. However, most of these models, as seen before, did not have the ability to detect subtle changes in the user's face. To overcome that, more recent models

using Convolutional Neural Networks (CNNs) have emerged. The use of these systems in studies in the areas of emotions has the advantage of recognising and evaluating the emotional reactions of users through their facial expressions in real time, which is why we decided to integrate this type of system as a complement to the emotional evaluation of stimuli carried out by the users themselves.

Chapter 3

Methodology

In this chapter, we present the main tools that allowed and helped us to develop this research. We also present the approach followed for carrying out our research, by explaining the common objectives of the three studies and the specific objectives of each one of them.

3.1 Research Approach

Through this research, we intended to verify whether an image, when evaluated, may be responsible for triggering an emotional reaction. Furthermore, we also intended to understand which areas of the image could be considered as Areas of Emotional Interest. For that purpose, we collected information related to the areas examined for the longest time by a set of individuals (using eye-tracking technology), and the identification of the content of each image responsible for a given emotional reaction, through the identification of the content tag representing the area responsible for that reaction.

This way, we intend to create an approach that may be used in other studies in this area, whose focus is to analyse the reactions of an individual based on the analysis of a set of images with emotional interest.

This research was carried out in a pandemic context, resulting from the appearance of the new coronavirus SARS-COV-2, responsible for the development of the disease COVID-19. As such, all data for this work were obtained remotely, through online platforms prepared by us for this purpose, without direct contact with the participants.

This research was divided into three user studies, which, despite having individual objectives, always contributed to the overall objective of this research. In the first study, the objective was the evaluation of a set of images, from which we intended to select the content tags that best represented the context conveyed by them. In the second study, we tried to correlate the emotional reactions experienced by the participants, while viewing the images, with the area that had been looked at for the longest time. Furthermore, we also aimed to understand if the emotional reactions that had been self-reported by the participants, were related to any specific content, identified through one of the content

tags resulting from the first study. Finally, in the third and final study, we aimed to take advantage of the emotionally relevant content tags, and the areas looked at for the longest time, both of which were derived from the second study, to identify which area might be the most emotionally salient.

None of the participants in the three studies, received any kind of monetary compensation or course credits. Also, they were able to leave at any time, without having to provide any justification, as their participation was entirely voluntary. The studies conducted, were in accordance with the Declaration of Helsinki and were approved by the Ethics Committee of the Faculty of Sciences. In all these studies, volunteers were asked for their consent to participate, and given a short briefing about the study, and also about all the tasks to be performed throughout the study.

3.2 Software Tools

To perform our research we used three software tools, which were essential for its sucess: Clarifai API, WebGazer and Face-api js.

Clarifai API[1] is a machine learning platform that allows developers to label, build and implement artificial intelligence models for unstructured data [12]. Of the various models already provided by this platform, we used the General model. This model, has the ability to analyse an image given as input, and predict the presence of elements from a list of 11 000 different concepts, which range from objects (e.g. building) to themes related to the analysed images (e.g. art). Along with the list of concepts, is also provided the probability of the presence of each concept in the image, on a scale from 0 to 1 [13]. Taking advantage of this model, we analysed our images to obtain the concepts that best represented them. These concepts were then filtered to remove the less relevant ones. The resulting set served as a basis for the development of the first study, and a subset of this was also used in the other two studies of our research.

WebGazer [2], is an open source eye tracking library, written entirely in JavaScript, that relies only on the webcam and the browser of the user's device, to infer the location of their gaze in real time, and which has no need to send data to a server [33][32]. It has the ability to self-calibrate, through user interactions with the web page and mouse cursor movements. Moreover, it can also train the mapping through regression models, which combine the location of the user's gaze, with the locations on the screen with which there was interaction [33]. It is composed by three libraries, js-objectdetect tracking.js and Clmtrackr, which are responsible for detecting the user's eyes and face, and MediaPipe Facemesh, a machine learning library with the ability to predict the user's face geometry [33].

As mentioned above, this software is only dependent of the user's device. As such, it

[1] https://www.clarifai.com/

is not intrusive, being better accepted by our volunteers. Furthermore, given the pandemic context in which our research was carried out, its use it use was invaluable importance, since it allowed us to obtain the data remotely, without the need of a special device (eye-tracker).

Regarding Face-api js4, this is an API, also an open source JavaScript library, which is used to train and implement machine learning models [2]. It has the ability to detect and recognize the user's face, using the core tensorflow.js [1]. This library implements a set of convolutional neural networks (CNNs) [10]. Among the several available models, we used the "Face Expression Recognition Model". This model, has two strands: i) analysis and detection of the facial expressions of all the faces present in an image; ii) detection of the facial expressions of a single face, option for which we opted, since the study was to be performed individually. The model in question, has the ability to locate, in the image/video provided as input, the user's face, whose signalling is performed through a bounding box, with which is also provided the probability of a given expression having been expressed [30]. In our study, since the user's image could not be available after the session (e.g. when users do not allow video recording), we process the facial expressions in real time and only store the expression evaluation data in Json format. In this data, the keys correspond the facial expressions corresponding to each of the six emotions defined by Paul Ekman, and the values corresponded to the probability of each emotion having been expressed [1]. This software was intended to complement the emotional evaluation performed by the users.

3.3 Summary

In this chapter we presented the main tools that allowed and helped the development of this work. We also briefly presented the general approach that was used to carry out this research.

We began by talking about the general approach of the work. During its presentation, we indicated that, in general, it would combine two types of data: i) the areas looked at for the longest time by users, which would be identified by eye tracking; ii) identification of the content responsible for an emotional reaction, which would be identified through a content tag. Besides, we also mentioned that all the data of this research would be obtained remotely, due to the pandemic situation resulting from the appearance of the new coronavirus SARS-COV-2, which coincided with the development period of this work. We also explained that the work is divided into three differentiated studies, each of which would have specific objectives: Study I - Selection of the most representative content of each image, through the selection of a content tag; Study II - correlation of the emotional

[2]https://webgazer.cs.brown.edu/
[2]https://justadudewhohacks.github.io/face-api.js/docs/index.html

reactions experienced by each participant, with the emotional content identified by the selection of content tags and zone looked at for a longer period; Study III - identification of the area with the highest emotional salience, taking advantage of emotionally relevant content tags and the area that is looked at for the longest time in a given image.

Finally, we talked about the main software used throughout the work: i) Clarifai API General model - model belonging to the machine learning platform Clarifai API, which has the ability to analyse an image given as input, and return a set of concepts representing its content, which come associated with the probability of their presence in the image; ii) WebGazer - open source eye tracking library, written entirely in Javascript, and which relies only on the browser and the webcam of the user's device, to track their gaze; iii) Face-api js Face Expression Recognition Model - open source library, also written in JavaScript, which implements machine learning models, with integrated Covolutional Neural Networks (CNNs), which allow to evaluate in real time the user's facial expressions, and return an object composed by the expressions corresponding to each of the six basic emotions of Ekman and Neutral emotion, accompanied by the probability of its detection.

Chapter 4

Study I: Content Tags

The aim of this first study, was to identify the five labels that best described the content of each of image. As such, we defined, the following research question for this study: Which content (tags) best describe the image?

4.1 Preparation of the Image Dataset for the Study

A person's emotional reactions depend on the stimulus that is presented to them. As such, we chose a dataset that not only met the objectives of this study, but that also allowed us to correctly evaluate the emotional reactions of the participants in the following studies. With this in mind, we selected the EmotionROI dataset [34].

EmotionROI is composed by a set of 1980 images, which are equally divided by the six universal emotions defined by Paul Ekman - anger, disgust, happiness, fear, sadness and surprise. Besides the photographs, the dataset also has the ground truth Emotion Stimuli Map (ESM) for each image. These emotional maps resulted from the average of the selections of the various areas of each image, in a study carried out by the authors of the dataset with a group of users, who were asked to mark the areas that had most influenced the emotions evoked by each image [34].

From the original dataset, we selected 252 images, 42 of each of Ekman's six basic emotions, which were selected in two distinct phases: a first phase where 240 images were selected (six sets of 40 images of each of the six emotions), and a second phase where the remaining 12 images were selected (six sets of 2 images of each of the six emotions). The purpose of adding the last 12 images was to be able, at a later stage, to annotate as many images as possible by as many users as possible. The selection of the photographs in both phases was carried out through a process composed of several stages, of which only two are common. In the first stage, we selected images that had a minimum width and height of 480px. In the second stage, the number of bytes of each file was evaluated, and only those with a higher value were selected, since they would be able to reproduce their content with greater quality.

After these first two steps, in the case of the first selection phase, taking advantage of the set of photographs that had been selected for each of the emotions, we selected 20 photographs with the landscape format and 20 photographs that were a mixture of images with the portrait format and of equal dimensions. These photographs were then equally distributed across 4 groups of images. In the case of the images selected in the second phase, we took the remaining images that would have resulted from the common selection process, and selected for each of the emotions 2 photographs - 1 with the landscape format and another with the portrait format. These photographs were then added to a single group, thus making up the group of 12 additional images.

4.2 Study Software Tools

For the development of this study, it was necessary to obtain a set of 15 tags representative of the content of each of the images. As previously mentioned, these tags were obtained prior to the study through the General model of the Clarifai API, which has the ability to evaluate an image and return a set of concepts representative of its content, along with the probability of their presence.

In order to obtain the set of desired tags, we created a process for their selection, which was composed of two stages. In the first stage, we asked the General model of the Clarifai API to analyse each image, and obtained a set of 30 concepts, representative of the content of each image. In the next stage, these concepts were filtered, and only the 15 that had the highest probability of being present in each image were selected.

4.3 Participants

Twenty volunteers participated in the study: 60% individuals identified themselves as female and 40% as male. The participants ranged in age from 17 to 55 (with an average age of 34.25 and a standard deviation of 11.64), and were from a wide variety of backgrounds: 40% from Computer Science; 25% from Engineering and Related Techniques; 10% from Mathematics and Statistics; 10% from the Arts; and 5% from Social and Behavioral Sciences. Regarding eye conditions, 40% had no eye conditions whatsoever, 30% had Astigmatism and Myopia, 15% had only Astigmatism, 10% had only Hyperopia, and 5% had Myopia. All of the individuals who reported having some eye condition indicated that they use glasses to correct it.

4.4 Study Methodology

To be able to participate in our study, the volunteers had to have access to their own device (mobile phone, tablet or computer) and internet access. In addition, they had to be aged

16 years old or over and had to perform the session individually, to minimise external influences that could impact the choice of tags for each of the images. The schedule for the realization of our study, in which each of the sessions had a duration of 25 to 30 minutes, was left to the discretion of each of the volunteers. The online platform developed to support the study was divided into three parts: i) consent form; ii) personal data form; iii) experimental phase.

In order to ensure that an equal number of evaluations were obtained for each of the images, and that there was an equal distribution of possible votes for all their tags, the images and their respective tags were evaluated in five groups of sessions, corresponding to each of the five image folders that were prepared previously.

4.5 Procedure

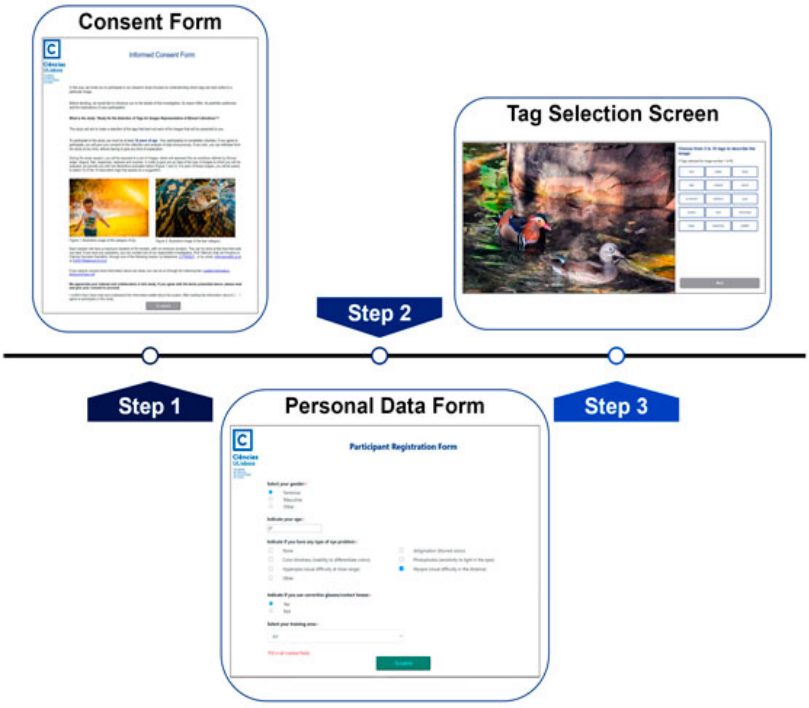


Figure 4.1 – Diagram representing the three steps taken by users in the first study

This study was conducted in an online platform developed for the context of this work. At the beginning of the study, a briefing was provided with all relevant information about the study, and two exemplary images were provided, similar to the ones that would be evaluated. Following the presentation of this information, we asked participants for their consent to participate in this study (Figure 4.1 - step 1). The Consent Form and Information Sheet were also available for download to ensure full access to all information

regarding this study and to clarify any doubts that may arise.

After accepting to participate in the study, the volunteers filled in a form with their personal information (Figure 4.1 - step 2). In this form, we asked them to provide information such as gender, age, whether they had any eye problems and their area of education. If the volunteers indicated that they had any eye condition, they were additionally asked to indicate whether they used glasses or corrective lenses to correct this condition. This last piece of information, was requested from the participants, as we would be asking them to view images and describe their content.

In the experimental phase, the volunteers were exposed to a set of images and respective 15 tags most likely to be present in their content, which had been obtained by the General model of the Clarifai API and selected by the process referred previously. During this phase, participants were asked to analyse the images and select between 3 and 10 tags that best represented the content of each photograph (Figure 4.1 - step 3).

4.6 The Five Tags per Image

With this analysis, we intended to identify the 5 tags that had the highest number of votes for each of the evaluated images. The purpose of this analysis was to understand which tags had been identified as the most relevant for each of the evaluated images, taking into account their content.

In order to identify them, we structured an identification process, which included four criteria to include the tags in the list of the most relevant ones. The first criterion allowed the identification and inclusion in this list only the tags that had been chosen by four people. If this criterion was not enough or was not satisfied at all, the tags which had been chosen by a minimum of three people were included in the list. If the required tags were still not all found, we moved on to the third criterion. The third criterion allowed the tags selected by at least two users to be included in the list of tags.Finally, if the five desired tags had not yet been identified, the last criterion came into effect, allowing the inclusion of tags that had been chosen by at least one person.

Although for most images this process allowed the identification of the five tags that best described their content, there were two exceptional cases. These two cases, occurred in images, in which each of its users had chosen only three tags. Due to the fact that there was a moderate agreement among the several evaluators, at the end of the identification process of the most relevant tags, only four tags were identified for each one of the images. As such, in order to reach the required number of tags, the selection of the fifth tag for each of these images was made by us from the list of 15 tags that would have been initially presented to the users.

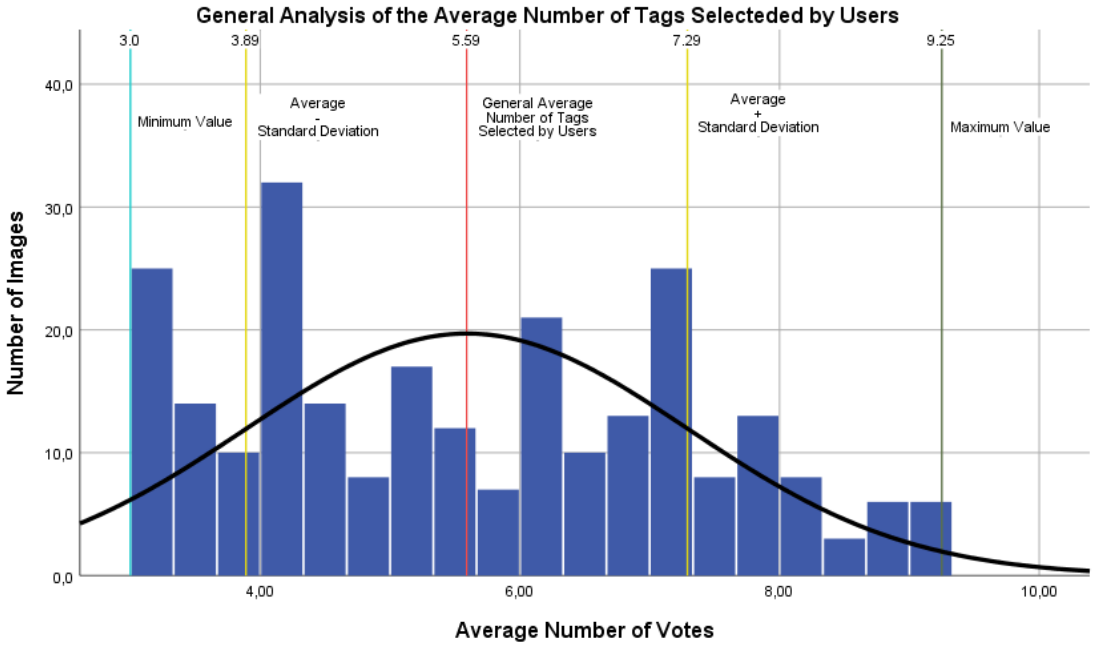


Figure 4.2 – Average, Standard deviation, Maximum and Minimum number of tags selected by users in Study I

4.7 Checking the Quality of Results

In order to ascertain the quality of the results obtained, for the analysis of the identification of the five most representative tags of the content of each image, we performed two analyses: i) verification of the average and standard deviation of the number of tags selected per user; ii) inter-rater agreement Fleiss' Kappa. In this section, we present the results of these two analyses.

4.7.1 Mean and standard deviation of the tags selected for the set of images

In order to understand the reason why, in some of the images, there would have been a greater difficulty in identifying the tags that best represented the content of the evaluated images, we investigated the average number of tags that would have been selected by each user for the total set of images.

The results of this analysis, showed that on average the number of tags selected per user during the evaluation of the images, would have been 5.59, with a standard deviation of 1.70 (Figure 4.2). These results, show not only that on average users selected about half of the maximum number allowed, and that in addition, as expected, there would have been some variability in the number of tags selected. Other conclusions that can be reached through the analysis of the graph in Figure 4.2 and taking into account these two

reference values and the normal curve, is that in 68% of the images each user selected on average between 3.89 and 7.29 tags.

Additionally, through this analysis it was also possible to verify that on average the minimum number of tags selected per user for each image would have been 3.00, as we expected, and the maximum number would have been on average 9.25 (Figure 4.2).

4.7.2 Inter-rater Agreement for the 5 tags

We verified the agreement between the several evaluators of an image, in relation to the five tags chosen for each. The analysis of this agreement was performed through the inter-rater agreement measure Fleiss' Kappa [3]. We selected this measure because: i) it allows verifying the level of agreement between two or more raters, when the evaluation method is categorical; ii) the targets to be evaluated (tags), were randomly selected from a population of interest, instead of being specifically chosen; iii) the raters, were not unique and were randomly selected from a larger population of raters [3]. Fleiss' Kappa and its 95% confidence level, were calculated using SPSS, package version 26 (SPSS Inc, Chicago, IL).

The average result for this analysis was 0.51 and its standard deviation was 0.23. There are no defined rules that can evaluate the Kappa value obtained through this analysis, as mentioned by Altman [5]. However, similar to what has been done in other works, so that we can make an analysis of these values, we will follow the guidelines defined by Altman in 1990 [5], which are based on Cohen's Kappa coefficient classification. From this, taking into account the average values obtained, we can state that in most cases the agreement between assessors was moderate. These values are justified by the fact that for our set of 252 images (Table 4.1): 6.35% had Kappa values < 0.2, indicating a poor agreement between the raters in relation to the choice of the 5 tags; 28.57% had a Kappa value of 0.21 - 0.40, indicating a poor agreement between the various raters; 31.35% had a Kappa value of 0.41 - 0.60, indicating a moderate agreement between raters; 23. 02% had a Kappa value of 0.61 - 0.80, concluding that for these images there was a good agreement between the various evaluators; 10.71% had an evaluation for this measure with values of 0.81 - 1.0, indicating a very good agreement between the evaluators of the images.

The values obtained for each of the images, can be consulted in Table A.1, which can be found in the appendix of this document.

4.8 Discussion of Study Results

Through this study, we intended to verify, for each image, which content could be considered more relevant, through the identification of a set of content tags. In order to achieve the desired set of tags, we started by identifying the five tags, which could be considered

Table 4.1 – Ranking table of the mean values obtained for the Fleiss Kappa measure in relation to the five tags chosen for each image

Kappa Value	Percentage of Images	Strength of agreement
$<$ **0.2**	6.35%	Poor
0.21 - 0.40	28.57%	Fair
0.41 - 0.60	31.35%	Moderate
0.61 - 0.80	23.02%	Good
0.80 - 1.00	10.71%	Very good

as the most relevant, through the identification of the five most voted tags by the evaluators of a given image. The results of this analysis showed us that: i) there was some variability in the choice of tags representative of the content of the evaluated images; ii) not all tags that in the end were considered as the most relevant for an image, had the vote of all its evaluators. As such, taking this into account, it was necessary for us to define four criteria for inclusion of the various tags selected, in the list of the five most relevant for each image under evaluation.

After obtaining these results, we went on to evaluate their quality through two analyses. In the first analysis, we tried to understand why there would have been some difficulty in selecting the five necessary tags, and so we calculated the average number of tags selected per user for each image. As can be seen from the values obtained, in most images the number of tags selected was around half of the maximum number of tags (5.59) that could be selected for each image. Also, as shown by the value of the standard deviation (1.70), as we already expected there would have been some variability in the number of tags selected. Besides as shown by the minimum value obtained for this analysis, there were image where their users only selected three tags representative of the content of the image under evaluation. Nevertheless, as we had verified during the identification process of the most relevant tags, most users selected about four or more tags per image.

As for the second analysis to check the quality of the results, which aimed to verify the level of agreement between the various evaluators regarding the tags selected as most relevant, we found that, on average, the agreement for the set of images was moderate. These results, once again, validated what had been verified during the identification process of the most relevant tags, which indicated that there was no total consensus regarding the choice of the five tags. However, looking at the results in more detail, we can state that for more than 50% of the images, the agreement between evaluators would have been moderate to very good. In general, these results lead us to conclude that the Clarifai API "General" model was effective in obtaining concepts representative of the content of a large number of images, although the concepts that were identified by this model as being the most likely to be present in an image were not totally consensual in some of the images.

4.9 Summary

In this chapter, we presented the first study of this work, whose main goal was to identify the most relevant content of each image, through the identification of the five content tags that best describe its content.

We started this study, with the preparation of the image dataset that would be used throughout the work. As mentioned, these images were selected from the EmotionROI dataset. Their selection followed a series of criteria defined by us (images with minimum dimensions of 480 px, identification of files with the highest number of bytes and the inclusion of images with varied formats), which led to the selection of a total of 252 images, 42 for each of the six basic emotions defined by Ekman (anger, disgust, fear, happiness, sadness and surprise).

Next, we proceeded to identify the concepts that best represented the content of each of the images. As previously mentioned, we initially obtained a set of 30 concepts for each image, which resulted from the analysis carried out by the General model of the Clarifai API. After being obtained, these concepts went through a filtering process, which resulted in the identification of the 15 concepts most likely to be present in each image. These concepts were then presented to the participants of this study, who selected between three and 10 tags that in their opinion best represented the content of each image evaluated.

After obtaining the results, we verified which were the five most voted tags for each image. During this analysis, we noticed that there was some variability in the selection of tags, so it was necessary to create several criteria to identify these tags. Although these criteria allowed the selection of the necessary tags in most images, there were two exceptional cases. In these two cases, we had to select a tag among the 15 that had initially been presented for those images.

Next, in order to check the quality of the results obtained during the identification process of the most relevant tags for each image, we performed two analyses. In the first analysis, we verified what would have been the average number of tags selected per user, for the whole dataset. The analysis showed that, on average, users would have selected 5.59 tags per image, with a standard deviation of 1.70. Moreover, it was also verified, an average minimum number of 2.40 and an average maximum of 9.25.

Finally, in order to assess inter-rater agreement regarding the choice of the five most voted tags, we calculated the Fleiss' Kappa inter-rater agreement measure. This analysis allowed us to verify that on average, Kappa was 0.51, with a standard deviation of 0.23. The values obtained, are indicative of the existence of a moderate agreement between users for the overall dataset. Moreover, when we look at the obtained results in more detail, we realize that for more than 50% of the images, this agreement was moderate to very good.

Chapter 5

Study II: Emotional Reactions and Emotional Tags

This second study had two objectives: i) to verify if there was any connection between the zones that were looked at for the longest time during the evaluation of the images and the emotional reactions experienced during that time; ii) to check the existence of a connection between the emotional reactions and the content of each image, represented by the tags selected in the previous study.

As such, in order to achieve our objectives in our study, we defined the following questions that guided us throughout its execution: i) *Which tags are most associated with each of the emotions?*; ii) *Which tags are most associated with each of the polarities?*; iii) *Do users associate more than one emotion with each image?*; iv) *Does the most selected emotion in an image correspond to the emotion in the original dataset?*.

5.1 Study Software Tools

In this study, two of the previously mentioned software were used: WebGazer and Face-api js.

WebGazer, the eye tracker library written in JavaScript, was used to obtain the coordinates of the user's gaze, during the viewing time of the evaluated images. As previously mentioned, this is a non-intrusive device, which is only dependent on the user's webcam and browser, to infer the location of their gaze.

As for Face-api js, like Webagzer, it is also an open-source library written in JavaScript, which was used to evaluate the facial expressions of our users during the viewing of the evaluated images. Its use intended to complement the emotional evaluation performed by the user himself, since this API provides the probability of a given emotional expression having been expressed at the moment of viewing the images.

5.2 Participants

In this study participated 83 users: 49.40% identified themselves as female and 50.60% as male. In relation to the nationality of the participants, the majority (77.11%) were Portuguese, 6,02% were American, 2.41% Norwegian, the same for German, 1.20% French, and a similar percentage for Polish, Peruvian, Romanian, Belgian, Bolivian, Singaporean, Malaysian, Angolan and Dutch. Regarding age, 63.86% of the participants were aged between 20 and 29 years, 19.28% between 30 and 39 years, 7.23% between 40 and 49 years and between 50 and 59 years, and finally 2.41% between 60 and 70 years. In relation to the level of education, the majority (84.34%) of individuals indicated that they had higher education, 13.25% up to the 12th year of education and 2.41% only the sixth year of education. In relation to their area of education, individuals from the most varied areas took part in our study, however, 43.37% indicated they belonged to the area of Computer Science, 15.66% to Engineering and Similar Techniques, 7.23% to Life Sciences and also 9.64% who indicated they had education in another area not specified in our options. In relation to the presence of some ophthalmological condition, 61.45% of the individuals revealed having some ophthalmological condition, of which 22.89% had astigmatism and myopia, 18.07% myopia and 8.43% astigmatism. All the individuals with some eye condition also indicated that they use corrective glasses/lenses to correct that condition. In addition to these personal data, the individuals were also asked about their emotional state, before starting the study. When assessing emotional polarity (Figure 5.1), 61.45% revealed to be neutral, 30.12% positive and only 8.43% negative. As for the emotions felt, the voting for the five levels, being 1 less intense and 5 most intense, can be seen in Figure 5.2. From this chart, we can highlight, the level one of the emotion disgust (25 votes). The indication of this level as one of the most voted, can be explained by the fact that one of the images that was given as an example on the consent page, was of a snake.

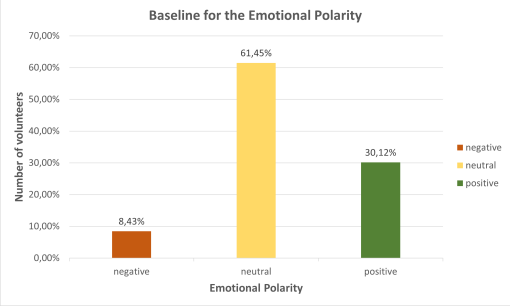


Figure 5.1 – Baseline for Emotional Polarity

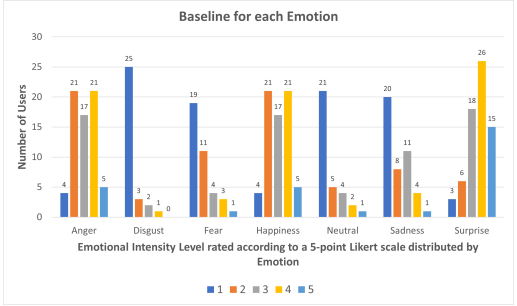


Figure 5.2 – Baseline for each Emotion

5.3 Study Methodology

To participate, our volunteers had to meet the following conditions: i) be aged 16 years or older and 100 years or younger; ii) have their own device (computer) with an integrated webcam and internet access; iii) have one of the following browsers installed on their device: Google Chrome, Firefox, Opera, Microsoft Edge and Brave; iv) carry out the sessions individually. The second condition was introduced in this study because we needed to make sure that in the experimental phase the device was fixed on a surface, so that the eye tracker could correctly track the coordinates of the user's gaze. As for the third condition, its introduction was due to limitations of the software used, which only allowed the study to be fully functional for the indicated browsers. Finally, the last condition was introduced in order to avoid external influences (e.g., comments from other participants or noise) that could influence the emotional experience of the participant and the correct evaluation of the images [6]. Because it was conducted online, there were no limitations of any kind in terms of time and place to carry it out. As had also happened in the previous study, we divided the selected photographs by several groups. Initially, the study was composed of seven groups, each of which had 36 images - six images for each of the six emotions. However, after receiving feedback that the sessions were too long and tiring, we divided them in half, giving a total of 14 groups, each with 18 images - three for each of the six emotions defined by Ekman. Each session of this last version, lasted between 10 and 15 minutes.

5.4 Pilot Tests

In order to validate and improve the online platform developed for this study, we conducted pilot tests with seven users (four of whom later participated in the main experiment, but using a different set of images). Six of these tests were carried out at the same time, to verify the capacity of the server to deal with several simultaneous participation's. The last one was carried out at a later date. All the tests were carried out on each user's own devices (computer), which had a built-in webcam and from different browsers, depending on the browsers that the volunteers had installed on their devices.

In these tests, in the image evaluation form, the five tags identified as the most relevant in the previous study were made available. However, the feedback obtained from the volunteers that participated in these tests, indicated that in some cases the suggested tags did not cover the content that had been responsible for a certain emotional reaction. As such, after obtaining this opinion from our users, we opted for adding a sixth tag, the tag "other-none", for the cases in which users did not agree with the suggested tags, or thought that they were not enough to express what they had felt while viewing a certain image.

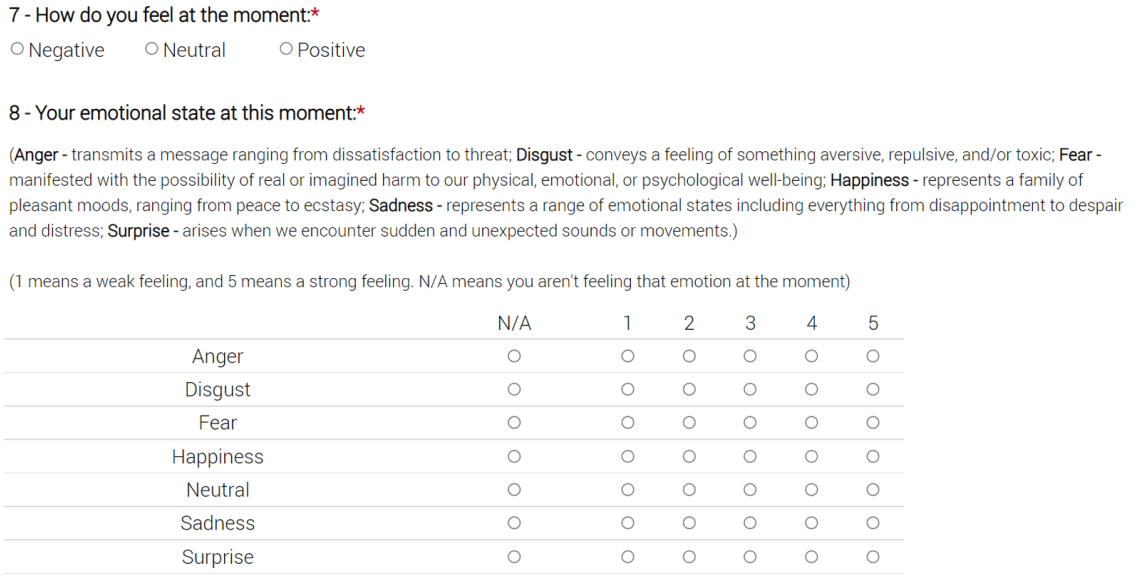
Besides this modification, it was also verified during these tests that the platform did

not work in two of the browsers tested: Firefox and Safari. As such, it was necessary to update the code of our platform, so that it could be supported in as many browsers as possible. However, despite this update, due to limitations in the software used in it, it was still necessary to limit access to only a few browsers (Google Chrome, Firefox, Opera, Brave and Microsoft Edge), leaving Safari out.

5.5 Procedure

At the beginning of the study, we provided a short briefing with all relevant information about the study, and requested consent for their participation in the study. In addition, we asked to videotape their participation, which was not compulsory. The Consent Form and Information Sheet were also made available for download in order to guarantee access to all the study information and to clarify any additional questions that may arise.

After accepting to participate, users filled in their demographic data on the personal data form provided, where they were asked to provide information such as age, gender, whether they had any ophthalmological conditions, academic qualifications, etc. In addition, each participant was also asked to indicate their emotional polarity at that moment and the intensity of each of their emotions, using a 5-point Likert scale. In order to make sure that users understood what each of the emotions meant, a short definition was given for each one (Figure 5.3 - step 1).


7 - How do you feel at the moment:*

○ Negative ○ Neutral ○ Positive

8 - Your emotional state at this moment:*

(**Anger** - transmits a message ranging from dissatisfaction to threat; **Disgust** - conveys a feeling of something aversive, repulsive, and/or toxic; **Fear** - manifested with the possibility of real or imagined harm to our physical, emotional, or psychological well-being; **Happiness** - represents a family of pleasant moods, ranging from peace to ecstasy; **Sadness** - represents a range of emotional states including everything from disappointment to despair and distress; **Surprise** - arises when we encounter sudden and unexpected sounds or movements.)

(1 means a weak feeling, and 5 means a strong feeling. N/A means you aren't feeling that emotion at the moment)

	N/A	1	2	3	4	5
Anger	○	○	○	○	○	○
Disgust	○	○	○	○	○	○
Fear	○	○	○	○	○	○
Happiness	○	○	○	○	○	○
Neutral	○	○	○	○	○	○
Sadness	○	○	○	○	○	○
Surprise	○	○	○	○	○	○

Figure 5.3 – Emotion evaluation of the user before the experimental part on the second part of the study

When moving to the experimental phase (Figure 5.4 - step 2), the participants' first task was to check whether WebGazer could correctly detect their faces. In order to work correctly, the users needed to have their face positioned inside the rectangle, in which the

eye tracker predicted the users' faces to be, and to be facing the webcam. If it was verified that the users' were respecting these conditions, the rectangle got a green border and it was possible to go on to the calibration phase. Otherwise, the border of the rectangle turned red, and it was not possible to move on to the next stage. To ensure that the participant continued to respect these conditions during the experimental phase, they were asked to remain as quiet as possible during the experiment, and to only use the mouse while filling in the evaluation survey for each of the images. Once this task was completed, they then moved on to the calibration of the eye tracker.

During the calibration (Figure 5.4 - step 3), the users were asked to look at the red circles appearing on the blank page in front of them. These circles, which had a white cross in the centre so as to concentrate the user's gaze on that location, would appear randomly when the user clicked on the previous circle. These buttons were strategically distributed all over the screen so that the eye tracker could detect the whole dimension of the user's screen. Once this stage was completed, users moved on to the image evaluation phase, which was divided in four steps (Figure 5.4 - step 4 to 7), which were repeated for each new image: i) appearance of a grey screen (step 4); ii) display of the image to be evaluated (step 5); iii) image evaluation questionnaire (step 6); iv) selection of content tags for each of the emotions felt during image evaluation (step 7). The first two steps of this sequence occurred in fixed intervals of five seconds and 10 seconds, respectively. The evaluation questionnaire and the choice of tags did not have a fixed evaluation time, so as not to pressure our participants or cause any type of stress during their participation in our study.

The grey screen, served as a neutral image. As mentioned in previous studies, the application of neutral stimuli has the ability to minimize and clear previous emotional influences [21]. As for the images, they were displayed in a random order on a screen with a black background, in order to highlight and minimize distractions during their evaluation. In the image evaluation phase, the user was asked to evaluate the images in terms of emotional polarity and emotions felt, for which it was necessary to select the intensity of each emotion, according to the Likert scale of five values. Once this task was completed, and if intensities between one and five were indicated for any of the emotions, the user was led to choose the content tag(s), which made mention of the image content that would have led to a certain emotional reaction.

Because we were dependent on the voluntary participation of remote users, who carried out their participation on their own devices, we encountered several problems. One of the problems was related to the users' video recording - in some cases the videos recorded during the sessions were unsuccessful in obtaining image, and therefore Face-api js data. As a result, the sessions in question resulted in sessions with incomplete data. Although the Face-api js data in these sessions could not be used to validate the emotional self-assessment performed by the user, we still decided that the data resulting

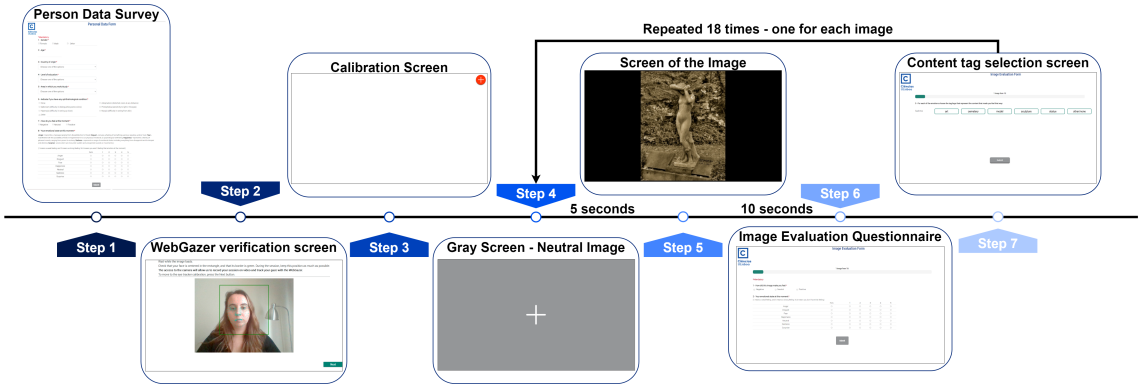


Figure 5.4 – Diagram representing the various steps of data collection in the second study

from the emotional self-assessment would be taken into account, since it would also contribute to verify which emotion and emotional content was more relevant to the images evaluated by these users. Another problem we faced was obtaining sessions that were incomplete - sessions that had been started, but whose user had not viewed all the images corresponding to that session.

All the data collected during the experimental phase, as well as the user's personal data, were stored in differentiated files in each participant's personal folder on the server where our study was located, to be later evaluated. In order to have consistent data that would allow us to analyse the data resulting from this study and achieve the proposed objectives, it was decided that an image would be considered as fully evaluated if we had obtained a complete evaluation of it (eye tracker data, users' emotional self-evaluation data and Face-api js data) from five users.

5.6 Results

In this section, we present the main results of the second study. We begin by presenting the identification process of the most selected tags for each image. Next, we present the analysis of the identification of the polarity and emotions most associated with each image. In the following stages, we present the analyses to verify which tags are most associated with each polarity and emotions. In addition, we also present the analysis to identify the type of content tags that were tagged, in which we also analyse the correlation between the type of tag tagged with the polarity and emotions associated with them. Finally, we verified which zones had a higher register of coordinates during the viewing of each image.

5.6.1 Most Selected Tags for Each of the Images

In order to verify which tags were considered most relevant by our users, and to prepare the third and last study of this research, we verified which tags had been the most voted

for each of the evaluated images.

To gather as many evaluations as possible for each image, we took into account the complete participations (users saw all the images), the incomplete participations (not all images of a given session were seen), and those in which there was some problem with the video recording of the user's image. In this first analysis, our purpose was only to verify at a global level which tags had the most votes, so we did not separate them by emotions.

To adequately perform this selection, we developed a small program in Python, in which two selection criteria were defined: i) first, the three tags with the largest number of votes would be selected, among all the tags selected by our users for a given image; ii) after this initial selection, if there was any tag that received the same number of votes as the tag chosen in the third place by the previous criterion, these would also appear in the list of the most relevant tags for that image. Since participants had the possibility of choosing the same content tag for more than one emotion, in our analysis, the tags that appeared more than once in the participation of a given user, were not eliminated and each repetition of them counted as one vote for this selection.

At the end of this analysis, from the 1260 tags that had been initially presented to the users, only 359 tags were considered emotionally relevant, for the total set of images. On average, this process led to the identification of 3.5 tags per image, with a standard deviation of 0.74. For the total dataset, there were 63.10% of images where three tags were identified, 25.40% where the process identified four tags, 9.92% of images had five tags identified as the most relevant and 1.59% where six relevant tags were identified. Furthermore, of all the tags chosen for the set of images evaluated, the 10 most chosen tags were other-none (146 images), nature (58 images), tree (16 images), water (14 images), animal (14 images), flower (13 images), colour (12 images), outdoors (12 images), ocean (12 images) and sky (11 images).

5.6.2 Images' Polarities

With this analysis we intended to verify which was the most voted polarity for each image. We identified, in the great majority of the cases, a great agreement between the several users, and that therefore there was only one polarity considered as the most correct for that image. As a result 30.95% of the images were considered as having mostly a neutral polarity; 29.37% of the images were mostly evaluated as having a negative polarity; and 26.98% of the images, the most voted polarity was the positive one. In the remaining images evaluated, there was a greater disagreement between the various evaluators and therefore we consider more than one polarity (Figure 5.5).

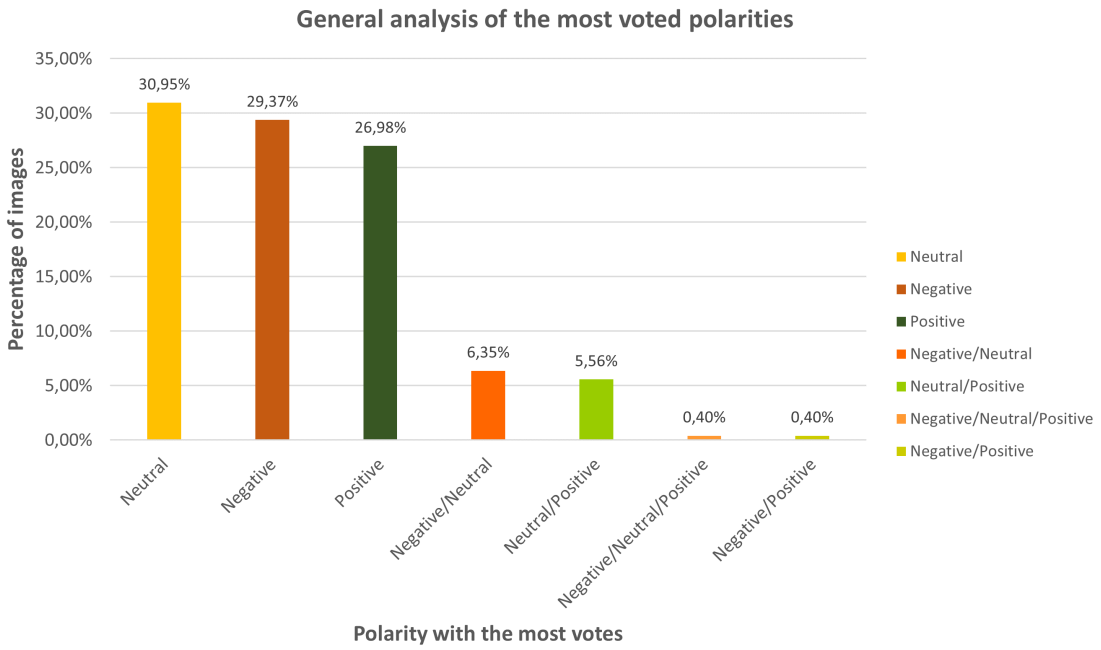


Figure 5.5 – Distribution of the percentage of images for each polarity or combination of polarities

5.6.3 Images' Emotions

With this analysis we aimed to understand which emotions users associated to each of the images evaluated in this study. This analysis was carried out by checking the relative frequency of each of the 6 emotions defined by Ekman and the neutral emotion in each of our images. Overall, we found that: 27.38% of our images had happiness as the emotion with the highest score; 14.68% were mostly associated with sadness; 13.10% were mostly associated with the neutral emotion; 7.54% were mostly found with the fear emotion; and only 3.17% of the images are associated with the surprise emotion. The remaining images, had more than one emotion as the one that had been most voted for.

Analyzing individually for each of the six original categories of the images: for anger, the emotion for which there was more votes in these images was happiness (16.67% of the images), having also a large number of votes for neutral (14.29% of the images) (Figure 5.6); in the case of disgust, in most cases, the emotion most voted for was disgust (42.86% of the images), there being therefore in this case a large number of images in which the original emotion corresponded to the one with the most votes (Figure 5.7); in the category of fear, as it happened in the case of disgust, the emotion most voted was the emotion originally associated with the image (26. 19% of the images) (Figure 5.8); in happiness, the same occurred as in the two previous cases, that is, in most images (69.05%) the emotion that received the most votes was happiness (Figure 5.9); in sadness, the same occurred again, since the emotion with the highest representation in these images

(50.00% of the images) was sadness (Figure 5.10); in surprise, this correspondence did not occur since the emotion with the highest representation among the images (66.67%) was happiness (Figure 5.11).

It should also be noted that the emotions that had the highest percentage of votes within each emotional category to which the images were originally associated were also represented in images in which more than one emotion was associated, as can be seen in the graphs of each emotional category.

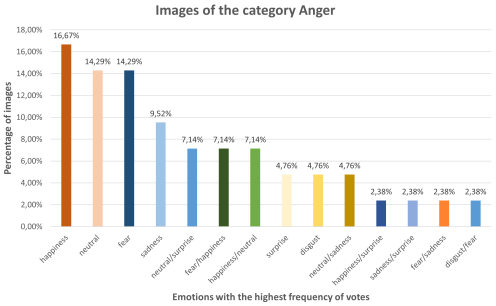


Figure 5.6 – Most voted emotions for the Anger images category

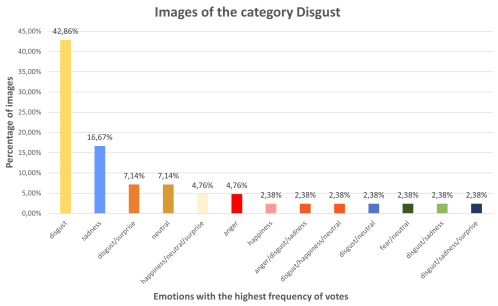


Figure 5.7 – Most voted emotions for the Disgust image category

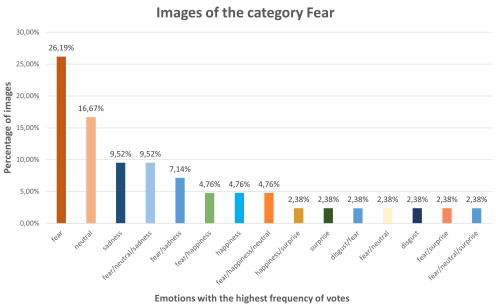


Figure 5.8 – Most voted emotions for the Fear images category

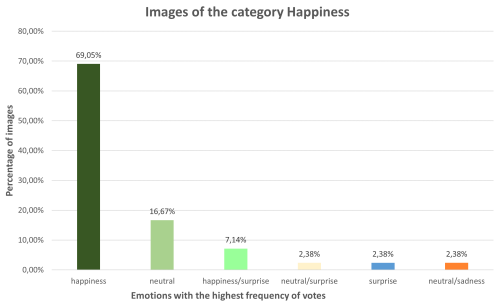


Figure 5.9 – Most voted emotions for the Happiness image category

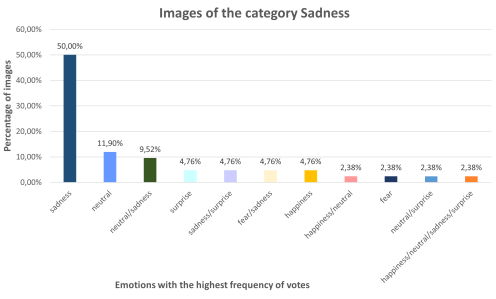


Figure 5.10 – Most voted emotions for the Sadness images category

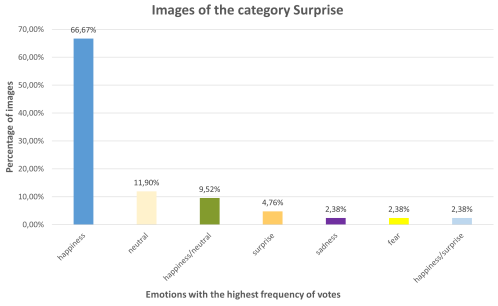


Figure 5.11 – Most voted emotions for the Surprise image category

5.6.4 Tags Most Associated with Each of the Polarities

In this study, a total of 1260 tags were made available to the users, from which they selected 359 has being emotionally relevant. This analysis aimed to verify which tags were more associated with each polarity. It allowed us to verify that from the set of tags considered relevant by the users, 35.38% were mostly associated to the Positive polarity, 32.87% to the Negative polarity and 19.50% to the Neutral polarity. In the remaining cases, there was more than one polarity with a large number of votes for the evaluated tag: 6.41% of the set of emotionally relevant tags had the Neutral/Positive polarity, 2.51% had the Negative/Neutral polarity, 2.23% had Negative/Positive and 1.11% had a tie for the 3 polarities. It is also important to note that there were cases in which our users considered that the tags provided were not the most appropriate to describe the emotional analysis they had performed to a particular image, and therefore marked the alternative "other-none". This alternative tag was mostly associated to the Neutral polarity (487 votes), but also to the other two polarities - Negative with 318 votes and Positive with 308 votes.

5.6.5 Tags Most Associated with Each of the Emotions

Taking the same 359 tags evaluated in the previous analysis, we checked which emotions had been considered as the most appropriate for each one of them, by our volunteers. This analysis, allowed us to verify that most of the tags (37.60%) considered relevant, were associated to the emotion Happiness, 11.98% to Sadness, 10.31% to Neutral, 9.47% to Disgust, 8.91% to Fear, 3.62% to Surprise and 1.11% to Anger. In the case of the remaining tags, there was more than one emotion considered as emotionally relevant for the evaluated tag, as can be seen in Table 5.1. Regarding the other-none tag, as previously mentioned, it was pointed out in some cases where the tags provided were not considered the most adequate for the emotional evaluation of a certain image. This tag, had as most voted emotion the neutral (210 votes).

5.6.6 Correlation between tags and polarities and tags and emotions:

With this analysis, we intended to analyse the type of content that had been considered emotionally relevant, taking advantage of the content tags selected during our study. Through its results, we want to understand if the emotional reactions experienced by our volunteers were the result of the general idea transmitted by the image under evaluation, or if, on the other hand, it resulted from the analysis of a specific object/being represented in it. We also intended to verify the distribution of the several types of emotions and polarities by the types of tags, and if the emotion most associated to each tag corresponded to the expected polarity.

Table 5.1 – Table resulting from the analysis of which tags are most associated with each emotion

Emotions	% of Tags
Happiness	37.60%
Sadness	11.98%
Neutral	10.31%
Disgust	9.47%
Fear	8,91%
Happiness/Neutral	4.74%
Surprise	3,62%
Happiness/Surprise	1,39%
Anger	1,11%
Fear/Sadness	0,84%
Neutral/Surprise	0,84%
Anger/Fear	0,84%
Fear/Happiness	0,84%
Neutral/Sadness	0,56%
Anger/Disgust	0,56%
Disgust/Fear	0,56%
Fear/Neutral/Sadness	0,56%
Disgust/Happiness/Surprise	0,28%
Happiness/Neutral/Sadness	0,28%
Anger/Fear/Happiness/Sadness	0,28%
Disgust/Surprise	0,28%
Anger/Sadness	0,28%
Disgust/Fear/Happiness	0,28%
Sadness/Surprise	0,28%
Disgust/Fear/Happiness/Neutral/Sadness/Surprise	0,28%
Anger/Fear/Neutral/Sadness	0,28%
Disgust/Fear/Happiness/Surprise	0,28%
Anger/Happiness/Sadness/Surprise	0,28%
Fear/Neutral/Surprise	0,28%
Disgust/Fear/Neutral/Sadness	0,28%
Disgust/Fear/Neutral	0,28%
Disgust/Neutral/Sadness	0,28%
Fear/Sadness/Surprise	0,28%
Anger/Disgust/Happiness/Sadness	0,28%
Fear/Surprise	0,28%
Disgust/Sadness	0,28%

In order to analyse only the most relevant tags, we selected those with 10 or more votes for any of the polarities. This resulted in a set of 83 tags for evaluation. To identify the type of content, we took into account not only the results of the two previous analyses,

but also the Panofsky/Shatford matrix. It was therefore necessary to verify the content of the images with which each of the analysed tags was associated. For example, taking the case of the tag ”decay”, which was mostly associated with the emotion Sadness and Negative polarity, the analysis led to it being classified as belonging to the *pre-iconography category (general)* and the question *What?* (Table 5.2). Its integration within this category and question was due to the fact that it performed a general description of objects, which in the images with which it was associated, were buildings or parts of buildings.

Table 5.2 shows the results obtained with this analysis. In general, the vast majority of the analysed tags belong to the pre-iconography (general) category (61 tags), being in second place the *Iconology (Abstract)* category with 13 tags and finally *Iconographical (Specific)* with only nine tags. Taking into account the emotions associated to each tag, one of the first observations that we were able to make was that none of the analyzed tags was mostly associated with the emotion Anger. In the case of the remaining six Ekman emotions, we noticed that the great majority or the totality of the tags associated to these emotions, as expected considering the general distribution of the tags, were associated to the *pre-iconography (general)* category (Disgust (D) - nine tags; Fear (F) - seven tags; Happiness (H) - 24 tags; Neutral - three tags; Sadness (S) - 14 tags; and Surprise (SU)- three tags). Moreover, in all 5 emotions analysed, in the vast majority or the totality, as was the case of the emotion Surprise, the tags associated with this category belonged to the *What?* (D - seven tags; F - five tags; H - 12 tags; N - two tags; S - eight tags; SU - three tags), the remaining tags were distributed by the question *Where?* (H - eight tags; and S - two tags) and *Who?* (H - four tags; S - four tags; F - two tags; D - two tags; N - one tag). In the case of the *Iconology (Abstract)* category (D - one tags; F - four tag; H - seven tags; and S - one tag) , similarly to what had happened with the previous category, the question with the highest tag representation was again *What?* (F and H with four tags; and D with one tag). The remaining tags were distributed over the remaining three questions: *Where?* (H - two tags) *When?* (H - one tag) and *Who?* (S - one tag). In the Iconographical (Specific) category (D - two tags; F - one tag; H - two tags; N - two tags and S - two tags), similarly to the two previous categories, the question with the highest tag representation was again *What?* (D - two tags; F, H, N and S with only one tag), the others being distributed among the questions *Where?* (N and S both with one tag) and *Who?* (H - one tag). Besides these tags, there was also an exceptional case of a tag which was mostly associated with more than one emotion - the monochrome tag which was associated with Neutral/Sadness. This tag, due to the content it was associated with and the fact that it described the general content of the images for which it had been selected, was integrated during this analysis in the pre-iconography (general) category and in the *What?*.

As previously mentioned, this analysis also aimed at verifying the correlation between emotions and polarities mostly associated with each of the evaluated tags. In general,

we can see that most of the emotions associated to tags were connected to the expected polarity, that is, in most cases, the tags associated with positive emotions like Happiness were associated to positive polarity (H - 19 tags), while emotions normally associated to something negative (D - eight tags; F - six tags; and S - 13 tags), were associated to negative polarity, and neutral as mostly associated to Neutral polarity (N - three tags). However, as can be seen in Table 5.2, there are some exceptional cases, of which we can highlight the case of the tags ”animal” and ”child”. These two tags, which are associated with the emotion Happiness, contrary to what would be expected, were mostly associated with Negative polarity and not with Positive polarity.

Questions	Pre-iconography (general)	Iconographical (specific)	Iconology (abstract)	Emotions	Polarity
Who?	animal child			**Happiness**	**Negative**
	girl man people woman		adult	**Sadness**	
	snake reptile			**Fear**	
	fish			**Disgust**	
	insect			**Disgust**	**Neutral**
	no person			**Neutral**	
	bird tree	seascape		**Happiness**	**Positive**
What?	abandoned broken decay grave old sculpture statue	black and white		**Sadness**	**Negative**
	dirty junk garbage pollution skull trash waste	blood bloody	calamity	**Disgust**	
	dark mask shadow	eye	horror mist scary pain	**Fear**	
	closeup texture			**Surprise**	
	portrait		light landscape	**Happiness**	
		grass		**Neutral**	
	monochrome			**Neutral/Sadness**	
	building			**Sadness**	**Negative/Neutral**
	abstract			**Surprise**	**Neutral**
	art reflection			**Neutral**	
	food			**Happiness**	
	splash storm			**Fear**	
	blooming cake color flora floral flower garden rock water wildlife	petal	beautiful sunset	**Happiness**	**Positive**
Where?	city street	cemetery		**Sadness**	**Negative**
		urban		**Neutral**	
	field			**Happiness**	**Neutral**
	beach lake mountain ocean sea sky waterfall		nature outdoors	**Happiness**	**Positive**
When?			sunset	**Happiness**	**Positive**

Table 5.2 – Table of analysis of the most relevant tags for our set of images according to the Panofsky/Shatford matrix

5.6.7 Areas of the Images Most Looked

In order to verify which areas of the images received more attention from our users, we checked the distribution of the gaze coordinates of the users collected with the WebGazer. As previously mentioned, some problems were detected in some of the sessions carried out by some of our volunteers, which included the impossibility of detecting the webcam of the device used, and consequently the register of the coordinates during the analysis of the images. Besides, it was also verified that, in some cases, the software used failed to register the dimensions and location of the image to be evaluated. As such, despite having reached the minimum number of participants for all the 14 groups of images evaluated, the users in which these two types of problems were detected, were not taken into account for this analysis.

We also verified, when analysing the data obtained, that the amount of data taken by the eye tracker, during the period in which each image was available for evaluation, varied from user to user, and from image to image. On average, WebGazer retrieved for 102 of the 252 images between 300 and 399 coordinates of the users' gaze, between 200 and 299 coordinates for 81 images, between 400 and 499 coordinates for 56 images, and between 100 and 199 coordinates for 10 images. The remaining three images, were the ones that obtained the highest amount of registered coordinates, verifying in these cases between 500 or more coordinates. Despite the high number of registered coordinates during this period, we verified that many of the coordinates had not been within the limits of the evaluated images and therefore were rejected. In table 5.3, we can consult the variation of registered coordinates within the limits of the dimensions of the various images, which allows us to verify that the discrepancy between the total of raw registers when compared with those that were analysed by this analysis, reaches a difference of 200 or more coordinates.

Table 5.3 – Table to verify the quantity of registered coordinates within the limits of the evaluated images and the quantity of images

Average number of coordinates registered within the image boundary	Number of Images
Gaze coordinates $\geq$ 200	5
20 $\leq$ Gaze coordinates $\leq$ 29	1
40 $\leq$ Gaze coordinates $\leq$ 49	5
50 $\leq$ Gaze coordinates $\leq$ 59	17
60 $\leq$ Gaze coordinates $\leq$ 69	19
70 $\leq$ Gaze coordinates $\leq$ 79	28
80 $\leq$ Gaze coordinates $\leq$ 89	40
90 $\leq$ Gaze coordinates $\leq$ 99	36
100 $\leq$ Gaze coordinates $\leq$ 199	101

To analyse the data, we divided the images in nine zones of equal dimensions, whose layout and organisation can be seen in the explanatory example in Figure 5.12. As the

dimensions of each zone and image varied from user to user, we took into account during the course of this analysis, the size with which the image was presented on the user's device, during the period of its visualization. Starting from these dimensions, and from the location of the image on the screen of the device where it was evaluated, we computed the coordinates detected from the gaze of the several evaluators. After, we identified the zone where the user would have concentrated his attention for a longer time, which could have been considered as the most relevant in emotional terms.

Figure 5.12 – Explanatory scheme with a image of the category Happiness, divided in the 9 zones in which our images were divided

In order to facilitate this analysis, the results obtained in this analysis were separated according to Ekman's six emotions with which the images were originally associated in the EmotionROI dataset, from which they originated. These results, presented in Figures 5.13, 5.15, 5.17, 5.19, 5.21 and 5.23, allow us to verify that for all the emotions, the zone with the greatest amount of coordinates is always the fifth zone (zone_E5), that is, the centre of the images (Figure 5.14, 5.16, 5.18, 5.20, 5.22 and 5.24). These results, lead us to hypothesize, that in most of the images associated with any of the emotions, the content that is more relevant in emotional terms is present in the centre of the images, or that WebGazer cannot properly track the edges of the screen.

Moreover, as it is also possible to verify, in four of the six emotions (Anger, Disgust, Sadness and Surprise), the second zone with the highest number of coordinates was the sixth zone (zone_E6), while in the remaining two emotions (Fear and Happiness), in the first one it was verified that the second most voted zone was the eighth zone, and there

was a tie in the second one between the sixth and eighth zones.

Finally, we could also verify, that there were areas that were not considered very relevant by our users. The first (zone_E1) and seventh (zone_E7) zones did not obtain the highest number of coordinate records in three of the emotions (Anger, Disgust and Fear) and third and fourth zones in two of the emotions (Disgust and Sadness). We also verified that in the case of the third zone (zone_E3), it is also absent in the emotion Fear, the first zone (zone_E1) in the emotion Surprise and the second zone (zone_E2) in the emotion Happiness. The absence of the previously mentioned zones leads us to think that their content would not be the most relevant in the various images evaluated in each of the emotions mentioned.

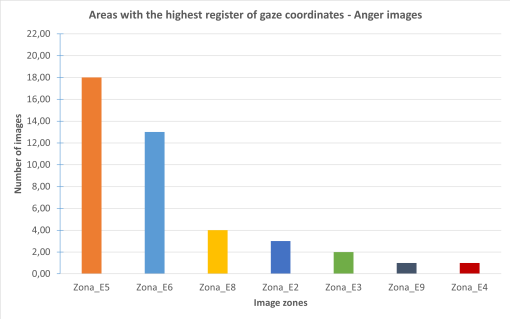


Figure 5.13 – Zones with the most gaze coordinates for Anger images

0,0	3,0	2,0
1,0	18,0	13,0
0,0	4,0	1,0

Figure 5.14 – Heatmap for the results obtained for the images of the Anger category

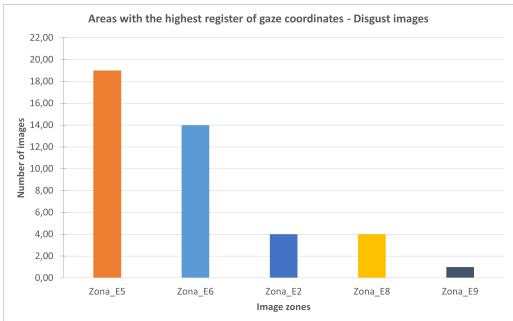


Figure 5.15 – Zones with the most gaze coordinates for Disgust images

0,0	4,0	0,0
0,0	19,0	14,0
0,0	4,0	1,0

Figure 5.16 – Heatmap of the results obtained for the images of the Disgust category images

5.7 Discussion of Study Results

This second study had two main goals: i) to verify if there was any relation between the emotional reactions experienced by the users and the content of the images they evaluated, which would be identified through the tags resulting from the previous study; ii) to verify

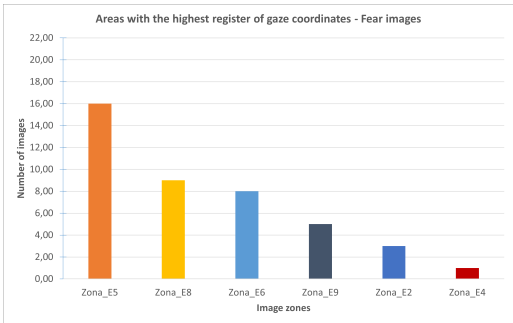


Figure 5.17 – Zones with the most gaze coordinates for Fear images

0,0	4,0	3,0
1,0	16,0	8,0
0,0	9,0	5,0

Figure 5.18 – Heatmap of the results obtained for the images of the Fear category images

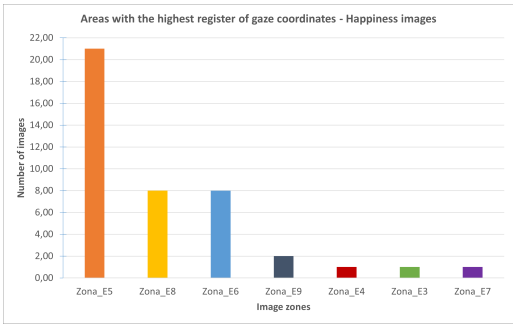


Figure 5.19 – Zones with the most gaze coordinates for happiness images

0,0	0,0	1,0
1,0	21,0	8,0
1,0	8,0	2,0

Figure 5.20 – Heatmap of the results obtained for the images of the Happiness category images

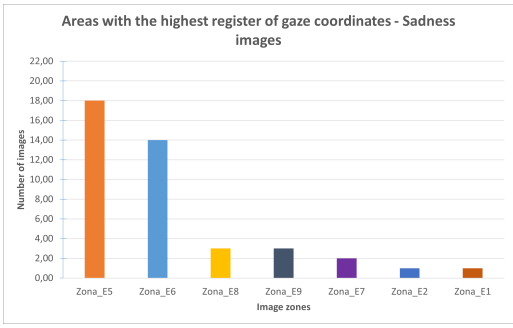


Figure 5.21 – Zones with the most gaze coordinates for Sadness images

1,0	1,0	0,0
0,0	18,0	14,0
2,0	3,0	3,0

Figure 5.22 – Heatmap of the results obtained for the images of the Sadness category images

if the areas looked at for a longer period would be related in some way with the emotional reactions experienced during the analysis of the images.

In order to achieve our goals, we started by checking which tags were most selected by our users, when evaluating each image. This analysis allowed us to realize: i) there was some variability in the tags chosen by the several evaluators, as it had occurred in the previous study; ii) there was a good agreement among the evaluators about the most significant content in each evaluated images, which was reflected by the ease with which

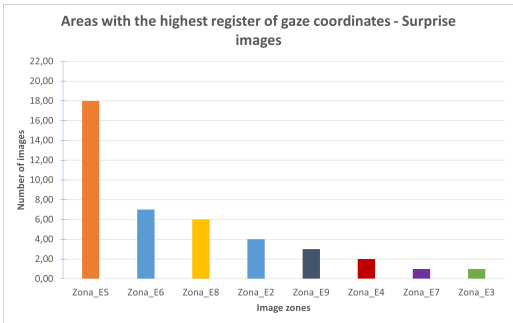


Figure 5.23 – Zones with the most gaze coordinates for Surprise images

Figure 5.24 – Heatmap of the results obtained for the images of the Surprise category images

it was possible to choose the three tags that we had initially defined that would be chosen, but also because there were some cases in which it was possible to select more than three tags. However, this last point also leads us to think that this may have been influenced not only by the agreement between users, but also by the fact that we allowed users to choose the same tag, for more than one emotion.

Next we checked which polarity was more associated with each image. As mentioned before, there was a great agreement among the several evaluators of an image, causing most images to have only one polarity as the most voted. However, as could be seen from the data obtained, there were cases where there was a tie in the number of votes for more than one polarity, which was reflected in images with more than one polarity as the most likely to be associated with it. This lack of agreement between the evaluators leads us to hypothesise that these images were not able to reflect a single polarity through their content, and the choice of the image, at the time of evaluation, was influenced by the tastes and life experiences of each of the evaluators.

Next, we checked which emotion was most associated with each of the images. One of the first conclusions that this analysis allowed us to reach, was that in most cases, the percentage of images with the highest value, were those whose votes corresponded to the emotion originally associated with them. However, there were two exceptions: images from the Anger and Surprise categories. In both cases, the emotion most voted for a large number of images, was Happiness. As previously mentioned, the images chosen for this research, came from the EmotionROI dataset. The images present in this dataset resulted from a search for keywords that corresponded to Ekman's six emotions and their synonyms, on the Flickr website. After this search, the images obtained also went through a human moderation process, which allowed the elimination of images whose authors did not consider adequate. Crossing this information, with the data obtained in our research, and focusing our analysis on the two exceptional cases mentioned above, it is possible to hypothesize that the type of annotation performed by the authors of the chosen images, differs from the type of annotation performed by our volunteers. As mentioned in [22], the

annotation originally carried out, may be a reflection of the emotion that the author of the evaluated images wanted to convey through them, that is, the annotation associated with these images may be of the intended type. However, our volunteers, when evaluating these images, based their evaluation on the emotional effect that the images provoked, performing, as expected, an induced annotation of the effective state of the evaluated image. By crossing these two types of annotation, we noticed that there was a difference between what they were supposed to convey and what they actually conveyed.

Contrary to what we had initially planned, we were not able to analyse the data obtained through Face-apijs. This would allowed us not only to validate the most voted emotions for each of the images, but also to check whether it was the area looked at for the longest time that was responsible for a particular emotional reaction. It was not possible to perform the analyses because the video from most of the participants did not allow to identify the expressions resulting from the emotions felt.

Moving on now to the analysis of the verification of which polarity is most associated with each of the tags selected for several images, we verified, as expected, that the amount of tags associated with positive and negative polarity were very similar. These values, as expected, reflected what had been verified during the analysis of the most voted polarity for the images. As for the analysis of the tags, taking into account the emotions, once again, the results obtained ended up being a reflection of the results that had been obtained in the analysis that checked which emotion had been most associated with each of the images.

In the next step, taking the results of these last two analyses we wanted to verify which type of tags were associated with the various emotions and polarities, and additionally check if the emotion most associated with each of the tags corresponded to the expected polarity. As reflected in Table 5.2, in the vast majority of cases the content that was considered as most relevant, in the various emotional categories, was not a specific content, but the general context transmitted by the image. This leads us to hypothesize that in most of the images evaluated, the content that was responsible for a given emotional reaction was not a specific object/being present in its content, but the general idea/theme transmitted by the image. Another conclusion that we could reach with this analysis is that, as expected, in most of the tags evaluated by this analysis, the emotion associated to it had the expected polarity as its correspondent, that is, tags with emotions considered negative had the negative polarity associated to them, the same happening in the case of emotions generally designated as positive. However, as mentioned above, there were two exceptional cases: the tags animal and child. Contrary to what would be expected, the most voted polarity associated to the emotion (Happiness emotion) for these two tags was not positive, but negative. The tag animal, appeared in a set of 20 images which in its majority, the most voted polarity was the negative one. However, contrary to what would be expected, the emotion with the highest number of votes was happiness for two reasons:

i) most selections of this tag occurred associated with the emotion happiness in images whose polarity was associated with the positive and neutral polarities; ii) in some of the images associated with the negative polarity, in which this tag was chosen, users not only associated this tag with emotions like sadness, disgust and fear, but also, in some cases, with the emotion happiness. In the case of the tag child, it appeared in a set of eight images. However, contrary to what happened in the case of the previous tag, the votes that led to the association of this tag to its polarity came from only two images: one associated to the negative polarity, from which came most of the selections of this tag (13 votes) and another with the Negative/Neutral polarity, which associated through two more selections, this tag to the negative polarity. As for the happiness emotion, similarly to the previous case, it was also found that most votes for this tag came from images with positive and neutral polarity.

Finally, we finalise the analysis of this study, with the verification of which zones registered the most attention from our users during the analysis of the images. As can be seen from the results, the zone of the image with the highest amount of registered coordinates, for most of the images in all emotional categories, was zone_E5, the central zone of the images. As mentioned in [39], this could mean that the users showed a great interest in the indicated area, which leads us to hypothesize that in most of our images, the content that could have been responsible for the emotional reactions experienced by our users, and which were registered in the evaluation of these images, could be present in this zone. However, although the data obtained allowed us to identify which zones may have been responsible for the emotional reactions registered by our users, the same data also allowed us to verify the existence of a large amount of coordinates, outside the limits of our images. These registers can be explained by two hypotheses: i) low level of attention of our users during the moment of image evaluation; ii) involuntary interference with the calibration of the eye tracker, at the moment of visualization through an unthought interaction with the device screen.

5.8 Summary

In this chapter, we presented the second study carried out within the scope of this research, which had the main purpose of correlating the emotional reactions experienced by the volunteers who participated in it, with the content of the evaluated images.

Through the results obtained in it, we began by checking which tags had been considered as the most relevant for each of the images, and consequently those that could be considered as the ones that best reflected the content present in the images analysed. From this analysis resulted sets of tags, which were composed of three or more tags.

Next we checked the associated polarity for each of the images. The results of this analysis allowed us to verify the existence of images associated with each one of the 3

polarities, and some exceptional cases in which there was no agreement on which polarity was more adequate for the image analysed. In the next analysis, we checked which emotion was most associated to each image, which allowed us to conclude, as expected, that in most of the emotion categories (Disguts, Fear, Happiness and Sadness), the emotion most associated to most of the images belonging to these categories, corresponded to those originally associated to them. However, that was not true for Anger and Surprise, where the emotion most voted was Happiness.

In the following steps, we checked which polarity and emotions were most associated with each of the tags selected by our users. These analyses, as expected, reflected in most cases, the results that had been obtained in the analyses that had been carried out to verify which emotion and polarity was most associated with each of the images. Taking the results of these two analyses, we also analysed the tags according to their type, where we concluded that in most cases, the tags selected were generalist tags. Additionally, this analysis also made it possible to verify if the polarity associated to each tag was the expected one considering the emotion it was associated to. As expected, in most cases, negative emotions were associated with negative polarities, and positive emotions with negative polarities. However, there were two exceptional cases where this was not the case, and where the positive emotion (Happiness) was associated with the negative polarity. Finally, we also analysed the data obtained by the eye tracker, in order to verify what could be considered the possible area of our images that could be considered as the most emotionally relevant. The data resulting from this analysis, showed that in most cases of images from all emotional categories, the zone with the highest number of registered coordinates was the fifth zone (the center).

Chapter 6

Study III: Tags Location

For this study,we intend to identify the area of each image that is emotionally most relevant, taking advantage of the tags selected in the previous study. To guide our research, we defined one question: i) Does the most looked at area have the greatest emotional charge?

6.1 Participants

157 users participated in this study: 57,96% identified themselves as female and 42,04% as male. Regarding age, most of our volunteers (47.77%) were between 16 and 25 years old. The remaining volunteers consist of 25.48% aged between 26 and 35 years, 15.29% aged between 46 and 55 years and, finally, 5.73% aged between 36 and 45 years, with the same value for those aged between 56 and 65 years. In relation to nationality, similarly to the previous study, the Portuguese nationality was the most representative (63.69%). Furthermore, it was also verified that the second country with the greatest participation (8.92%) was once again the United States of America, followed by the United Kingdom (4.46%). The remaining volunteers came from the most varied countries, among which we can highlight: Angola, Australia, Belgium, Cape Verde, France, Germany, India, Italy, Mozambique, Poland, Switzerland and Vietnam.

Regarding the level of education, the majority (80.89%) of the individuals indicated that they had studied up to University, 11.46% up to the 12th year of schooling, 0.64% had even the fourth year of schooling, while the rest had between the fourth and ninth years of School education. In relation to the area of education, individuals from the most varied areas participated in our study, however, we can highlight: 25.48% as belonging to the area of Informatics, 15.29% of Engineering and Related Techniques, 8.28% of Social and Behavioural Sciences, 7.01% of Arts and also 16.56% who indicated having education in another area not specified in our options. As for the presence of some eye condition, 61.15% of the individuals revealed having some eye condition, of which: 18.47% had myopia, 17.83% astigmatism combined with myopia, 10.83% only astigmatism, 5.10%

hyperopia, 2.55% photophobia, and 1.27% astigmatism combined with color blindness and myopia, a percentage which was also verified for the case of myopia combined with hyperopia. The remaining percentage of the population that indicated having some eye condition had astigmatism combined with some other disease, or only daltism. All the individuals with some ophthalmological condition also indicated that they use corrective glasses/lenses to correct this condition.

6.2 Study Methodology

As in the two previous studies, this one was also conducted online. To participate, the volunteers had to meet the following conditions: i) be between 16 and 100 years old; ii) have their own device (computer, tablet or mobile phone) with internet connection; iii) perform the study individually, in order to minimize external influences that could influence the choice of the image zones for each of the tags. As in the previous study, this was also disseminated via email (institutional and personal) by the study authors, as well as also through social networks such as Facebook, Instagram, Twitter, Reddit and Linkedin. In order to facilitate obtaining the necessary data, the photographs were divided in sessions. This time we prepared 28 sessions, each of which consisted of 9 photographs - 6 photographs of each of the universal emotions defined by Paul Ekman and 3 other photographs, which were also representative of these emotions, but whose selection was made randomly during the process of preparing the folders - and their respective tags. Each session took 5 minutes and could be held at a time that suited each participant.

6.3 Pilot Tests

In order to validate and improve the online platform that we developed for this study, we conducted pilot tests with two users, one of whom later participated in the main experiment. These tests were carried out at different times, through the devices of the volunteers, a computer and a tablet respectively, and using different browsers.

During these pilots tests, some bugs were detected in the platform code, which prevented the correct display and layout of the image and the areas to be selected for the emotional content. As such, after their detection, they were corrected, and the study began.

6.4 Procedure

Before starting the experimental phase, the volunteers were presented with an informative text in which the aim of our study and the tasks to be performed by it were mentioned. In addition, similarly to what had occurred in the two previous phases, two examples of

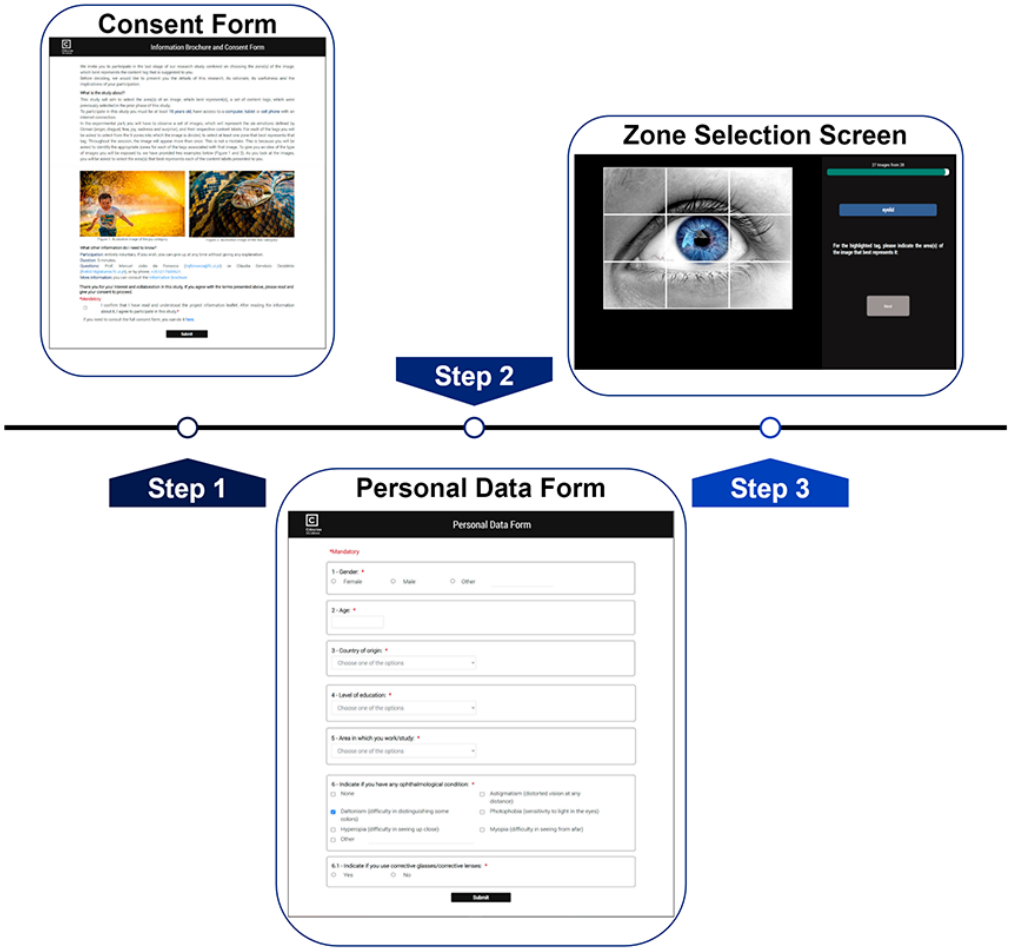


Figure 6.1 – Diagram representing the various steps of data collection in the third study

images were also given, exemplifying those to which the participants would be exposed during the experimental part of the study. In addition to this informative text, consent to participate was also requested and it was assured that the data obtained would be fully anonymised at the time of publication and presentation (Figure 6.1 - step 1). Once the consent for participation was given, users went on to fill in their personal data, where they were asked for some information such as gender, age, academic qualifications, area of work/study, if they had any ophthalmological condition and if they used corrective glasses or contact lenses, if the previous question had been different from none (Figure 6.1 - step 2).

After these two first stages the users moved to the experimental phase (Figure 6.1 - step 3). In this phase, we asked them to evaluate each image by indicating at least one zone that represented the content tag that was presented to them. In order to facilitate this task, several Image-tag pairs were created using the previously evaluated images and their respective most voted tags. The tag ”other-none” was not consider in this study. The images and their respective tags appeared to the participants in a random order. As the

number of tags to be evaluated in this phase varied from image to image, the number of pairs per session also varied.

Each photograph, was divided into 9 zones and presented with a black background in order to minimise distractions during their evaluation (Figure 6.1). When selecting the zone(s) that best represented one of the tags associated to a given image, the zone(s) was highlighted with a green colour and a check icon, so that the user could have a visual signal that the zone(s) was already selected.

In order to have consistency in relation to the number of participations and respective evaluations of our images and their respective tags, we considered that a session was fully evaluated after being viewed entirely by a minimum of 5 users. However, both complete and incomplete sessions - sessions in which not all image tags or all images were evaluated - were taken into account for the analysis of the results.

All the data collected during this last phase, was stored in an individual folder for each user, and in different files for each evaluated image. This folder, and its respective files, were stored in the server on which our study was being carried out, in order to allow later access and analysis of them.

6.5 Results

In this section, we present the main results of the third and last study of this work. We begin by checking the level of agreement between the various evaluators of an image, regarding the zones chosen for each of the tags considered as the most relevant for that image. In the next analysis, we performed the identification of the zones of the images to which emotional content was associated, where we tried to understand which zones had more emotional content associated, and those to which no content was associated. Finally, in the last analysis, we compared the zone of each image to which the user had looked for more time, with the zones where emotional content was detected.

6.5.1 Inter-Rater Agreement for the Zones for Each Tag

We verified the agreement between the several raters of an image, regarding the choice of the areas that best represented each tag, that had previously been considered as the most emotionally relevant. The analysis of this agreement was performed through the Fleiss Kappa inter-rater agreement measure. We chose this measure because: i) it allows checking the agreement between two or more raters, when the response variable is categorical; ii) the targets to be assessed by it (zones chosen for the tags), were chosen randomly from a population of interest, instead of being chosen specifically for this assessment; iii) the raters, were not unique and were also chosen randomly, from a larger population.

This evaluation was carried out separately for each tag, since for different tags it was possible to choose the same area of the same image. Then for each image, we computed

the agreement between the raters by considering the selected zones for the tag being evaluated. This process, was repeated for two (cases in which the third most chosen tag was other-none) or more tags representative of the emotional content of each image. This implied a total of 737 evaluations carried out.

The Fleiss Kappa was calculated through the statsmodels.stats.inter_rater.fleiss_kappa model, present in the Python package Statsmodels, which allows performing statistical tests. To be able to perform this analysis, taking advantage of this model, we coded each tag, for which we checked the agreement for the choice of the selected zones.

After obtaining the results of this analysis for each tag, we calculated the mean value of the Fleiss Kappa for each image. This allowed us to have an overall assessment of the level of agreement between the various users regarding the areas that best represented the emotional content associated with them. The average of the Fleiss Kappa values from each image was 0.03 and the standard deviation 0.14. This shows a large variability in the selection of the most adequate zones for each emotional content evaluated. Similarly to what we did previously in the first study, we classified the values obtained as being good or bad. As mentioned by Altman [5], there are no defined rules for assessing a kappa value as good/bad. However, similarly to what has been done in other works that take advantage of this statistical analysis, and to what we did before, we classified the values obtained according to the guidelines defined by Altman in 1990 [5], these are an adaptation of the guidelines used for the classification of Cohen's kappa coefficient. Following these guidelines we found that (Table 6.1): 87.70% of the images have a poor average agreement value between their evaluators, which means a lot of variability between the answers given; 11.11% had a fair average agreement between the evaluators, showing that in these cases there was a high variability in the choice of the zones, although this variability was smaller than in the previous images; and only 1.19% of the images had a moderate average agreement. The individual values obtained for this analysis may be consulted in Table A.2, which is annexed to this document.

Table 6.1 – Ranking table of the mean values obtained for the Fleiss Kappa measure for each image

Average number of Fleiss Kappa values for each images	Pecentage of Images	Strength of agreement
$<$ **0.2**	87.70%	Poor
0.21 - 0.40	11.11%	Fair
0.41 - 0.60	1.19%	Moderate

6.5.2 Zones with Emotional Content

In this section, we first describe our overall analysis of all images and then we do a more detailed analysis by each emotion.

Overall Analysis

From the data colleted in this study, we analysed the distribution of emotional content by the zones of each image. We did that by counting the number of times each tag was associated to each zone of the image. Then, with these values we created categorical heat maps, divided in the nine zones (see Figure 6.2). These heat maps, were designed using the submodule Pytplot, belonging to the Matplotlib software library, made for and from the Python programming language. All generated heatmaps, had a fixed scale for measuring votes per zone from zero to 27 (highest value of votes recorded for the set of all evaluated images), and a sequential colormap equal for all evaluated images. With the intention of verifying the distribution of emotional content across the image, when creating these heatmaps, we annotated each of the zones with the tags and their respective votes. The colour associated to each zone resulted from the sum of the votes of all tags associated to it. As we can see in Figure 6.2, the zones with lighter colours are the ones with fewer votes, or in some cases none, and therefore considered to have less emotionally relevant content, and the zones with darker colours, are the zones with more votes, and therefore a greater association of emotional content.

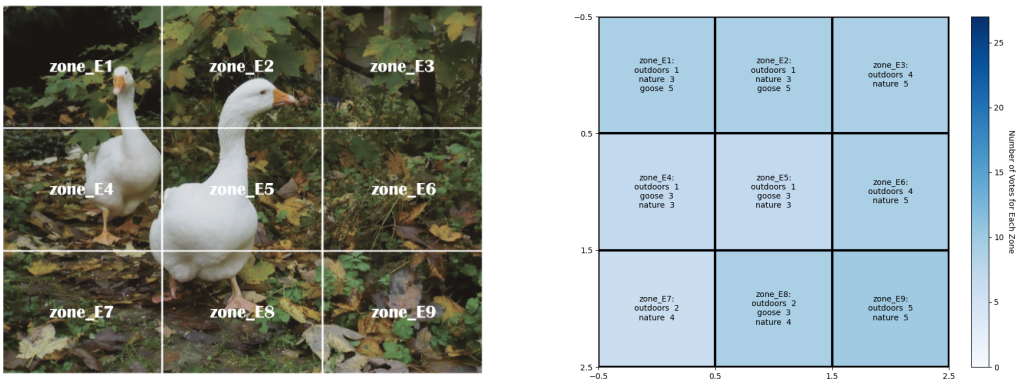


Figure 6.2 – Example of an image of the emotional category Happiness and corresponding heatmap, resulting from the selection of the various zones for the emotional content evaluated, where the most voted zone highlights darker blue (zone_E9)

Through this analysis, we noticed that in general: (i) in most of the images, as we expected, there was a great variability in the choice of the zones that best suited the emotional content evaluated; (ii) in some cases, we verified the existence of more than one zone with a large number of votes, i.e., with more emotional charge; iii) although the votes indicated that a certain zone(s) could be considered as having the highest emotional charge, in some cases, we found that the maximum value of votes attributed to a content was not in the zone identified as having the most emotional charge, but in the zone(s) with the second highest number of votes. For example in Figure 6.2, an example of an image associated with the emotion Happiness, we can see that the goose tag is not even

associated with the zone that was marked as having the highest emotional charge, but rather in zone_E1 and zone_E2, two of the zones with the second highest number of votes.

Analysis by Emotions

Separating now the results obtained, according to the six Ekman emotions associated to the image by the users, in the previous study, we found that: the majority of the images associated to the emotions Disgust, Fear, Happiness, Neutral and Sadness (Figures 6.5, 6.7, 6.9, 6.11,6.13 respectively), had as the most voted zone, and as such, potentially the most relevant for those images, the centre zone (zone_E5); in the case of the images associated with the emotion Anger (Figure 6.3), an emotion with which only two images are associated, we found that one had as the most voted zone the central zone of the image (zone_E5) and the other had two most voted zones, which were zone_E7 and zone_E8. In the case of Surprise (Figure 6.15), we verified in two images that the most voted zone was the centre of the image (zone_E5) and other two cases the zone_E6, and four other images, where the most voted zones were, respectively, zone_E8, zone_E2, zone_E1 and zone_E3. As for the images that had been associated with more than one emotion (Table A.3 which is in annex), we can highlight that in general the central zone of the image is once again the most voted one for the various categories (25 images), with the exception of the image from the anger/disgust/sadness category, in which the most voted zone was zone_E4. Besides this, we can also highlight the zones zone_E6 and zone_E8, which are also quite present among the most voted zones for the generality of the images categories.

Finally, in order to understand which zones contributed less or not at all to the emotional analysis of the evaluated images, we checked which zones received the least number of votes. To do so, we checked the least voted zone, in the images with only one least voted zone, and also the list of least voted zones, of those images in which there was more than one. Starting with the emotion Anger (Figure 6.4), we can verify that, looking at the data obtained from this analysis for the two images associated with this emotion, the zone with the fewest votes was zone_E3. In Disgust (Figure 6.6), the analysis showed, that the least voted zone was again zone_E3, having also highlighted zone_E7 and zone_E1. Moreover, the analysis also showed that in 13 of the photographs of this emotion, with the least number of votes, corresponded to zones with no emotional content assigned. As for Fear (Figure 6.8), it was again verified that zone_E3 was the least voted, and also zone_E1, zone_E8 and zone_E7. Besides, it was also verified that there were seven of the photographs where there were zones with no emotional charge attributed. In the case of Happiness (Figure 6.10), zone_E3 was again the least voted for the majority of the cases, with zone_E1 and zone_E2 also being highlighted. Similarly to the two previous emotions, there were again cases with zones without emotional content, having this situation occurred in eight images.

As for the images attributed to Neutral (Figure 6.12), the least voted zone for the

majority of the images was zone_E7, and zone_E1 can also be highlighted. The same situation was registered for the Sadness images (Figure 6.14). As for the images with zones without emotional charge, in the case of Neutral this situation occurred in five photographs, and in Sadness in 13 images, being this last emotional category the one that has registered the biggest number of images in which that happened. Finally, in relation to Surprise (Figure 6.16), there was an equal number of photographs in which zone E3 and zone E9 are in the list of the less voted zones. Of the photographs of this emotion, only in four there were zones with no emotional charge attributed.

As for the images that had more than one emotion attributed to them, as it can be seen in table A.4 (in annex), what was verified was that in general for these images, the zones that appeared more often identified as having the lowest number of votes were zone_E3 (27 images), zone_E9 (25 images) and zone_E1 (21 images). Additionally, we also verified for these pictures, how many of them the least voted zones corresponded to zero, what we found was there were 2 photographs in which this situation existed for the emotions neutral/surprise, happiness/neutral, sadness/surprise, fear/neutral/sadness and fear/sadness; and that such cases existed in 1 photograph in the emotions fear/neutral, happiness/neutral/surprise, fear/neutral/surprise, disgust/sadness, neutral/sadness, fear/surprise, disgust/surprise, fear/happiness and happiness/neutral/sadness/surprise.

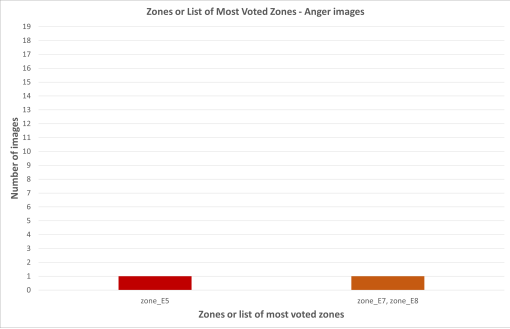


Figure 6.3 – Zones or list of most voted zones for Anger images

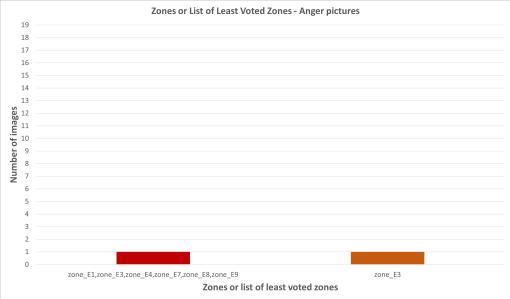


Figure 6.4 – Zones or list of least voted zones for Anger images

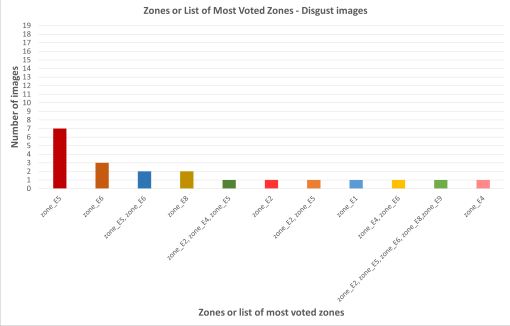


Figure 6.5 – Zones or list of most voted zones for Disgust images

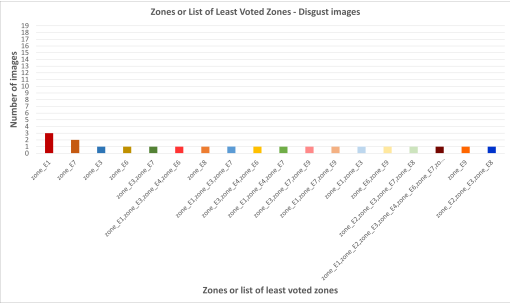


Figure 6.6 – Zones or list of least voted zones for Disgust images

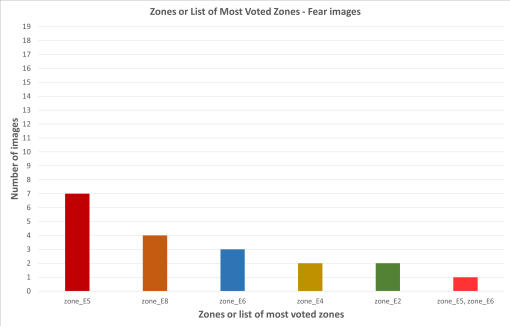


Figure 6.7 – Zones or list of most voted zones for Fear images

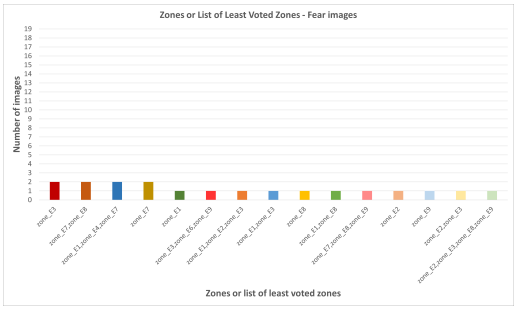


Figure 6.8 – Zones or list of least voted zones for Fear images

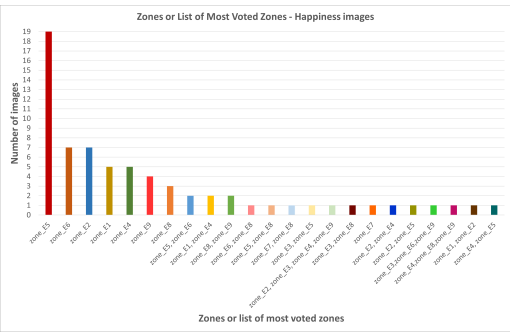


Figure 6.9 – Zones or list of most voted zones for Happiness images

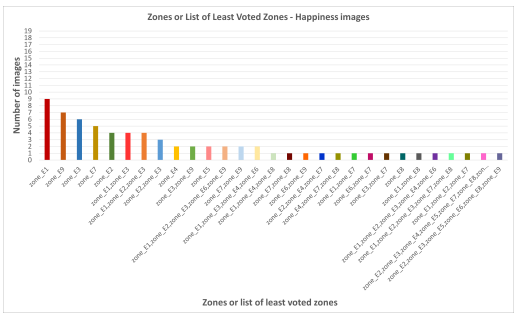


Figure 6.10 – Zones or list of least voted zones for Happiness images

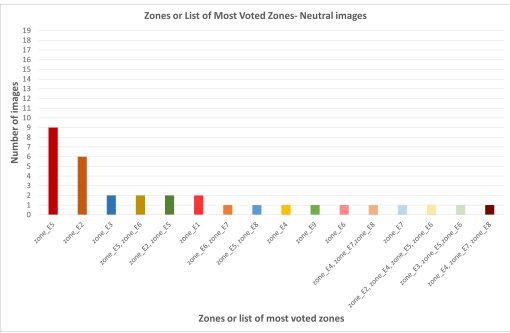


Figure 6.11 – Zones or list of most voted zones for Neutral images

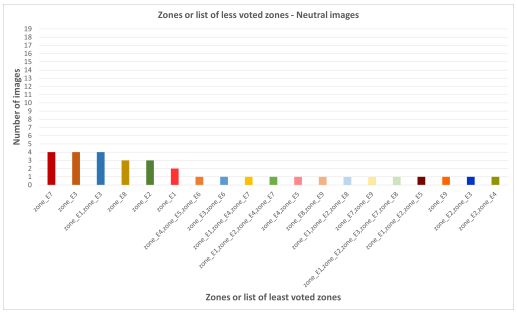


Figure 6.12 – Zones or list of least voted zones for Neutral images

6.5.3 Zones with Emotional Content vs. Most Looked Zones:

In order to check if the areas of the images, which had previously been recorded as having the most ocular interest (looked at longer), coincided with areas in which there was emotional content, we compared the data obtained previously in the analysis of the eye tracker records (previous study) with the data obtained in the analysis of the areas selected for each of the images.

From this comparison, we verified that in only 64 of the 252 images evaluated, we have a coincidence between the most voted zone(s) (candidates for the most emotionally

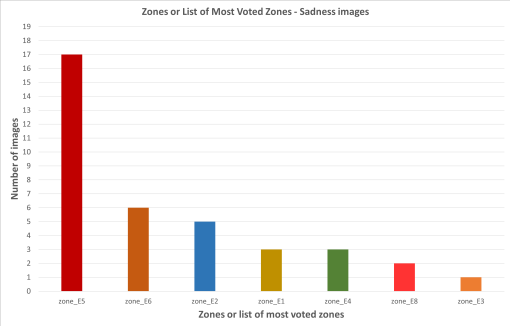


Figure 6.13 – Zones or list of most voted zones for Sadness images

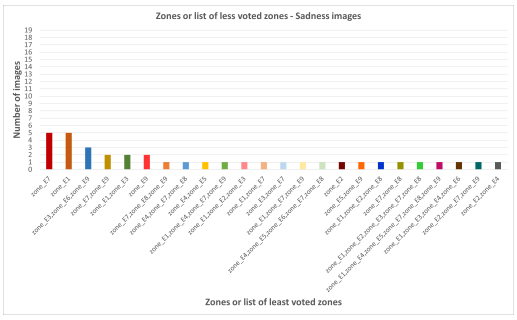


Figure 6.14 – Zones or list of least voted zones for Sadness images

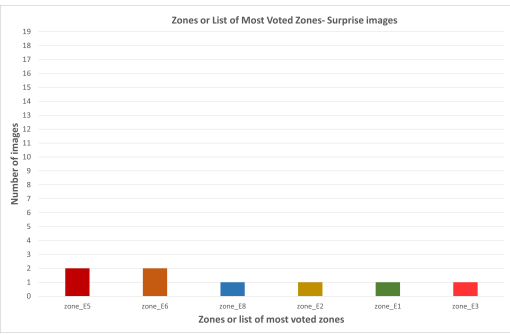


Figure 6.15 – Zones or list of most voted zones for Surprise images

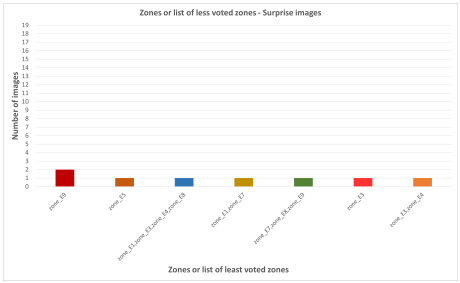
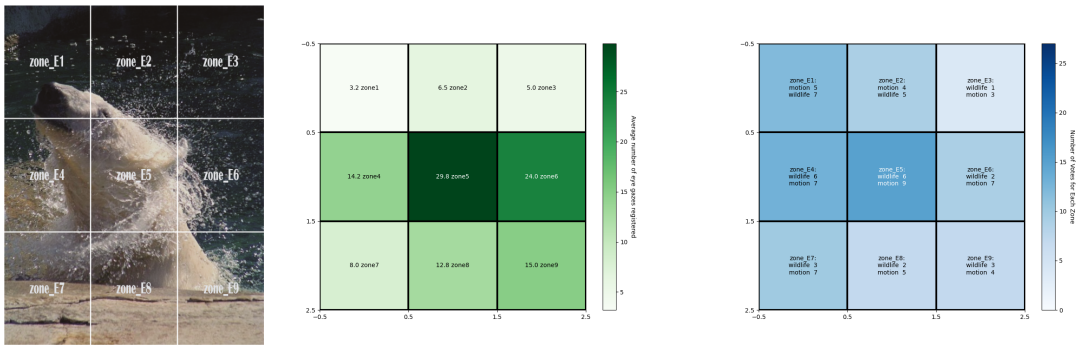


Figure 6.16 – Zones or list of least voted zones for Surprise images

relevant zones) and the zones with most ocular interest. As such, only in these cases we can say that the zones of ocular interest corresponded to the zones of greater emotional interest (see Figure 6.17). As we can see in Table 6.2, the images where most of this situation occurred were associated with the emotions Happiness (12 images), Neutral (11 images) and Sadness (10 images). In order to understand what had happened with the remaining 188 images, in which this correspondence was not verified, we checked the remaining selected zones of these images.

Through this verification we found that in six images (example in Figure 6.18), the zone of ocular interest coincided with zones in which no type of emotional content was registered. Of these images, four belonged to Sadness, one to Sadness/Surprise and another to Neutral/Surprise. In another 18 images, the area of ocular interest coincided with the less voted zones for the evaluated images. Within this set of images, four corresponded to images related during the second study to Happiness, three to Disgust, the same number of images for Neutral and Sadness, one to Disgust/Fear, the same for Fear, Sadness/Surprise, Fear/Neutral and Fear/Sadness.

As for the remaining images, we found that: (i) 58 images (example in Figure 6.19) had the area with most ocular interest, to coincide with the second most voted zones(s), of which mostly (20 images) corresponded to Happiness (Table 6.3); (ii) 47 images, the area with most ocular interest corresponded to the third most voted zones (Table 6.4),where



(A) (B) (C)

Figure 6.17 – Example of an image where the area of ocular interest coincided with the area of greatest emotional interest: (A) image from the Surprise category that was assessed; (B) heatmap resulting from the data recorded by the Eye Tracker, during image analysis and where the area with most ocular interest is highlighted in darker green (zone_E5); (C) heatmap resulting from the selection of the most appropriate areas to the several emotional contents associated to the image, where the Zone of Emotional Interest is highlighted in darker blue (zone_E5)

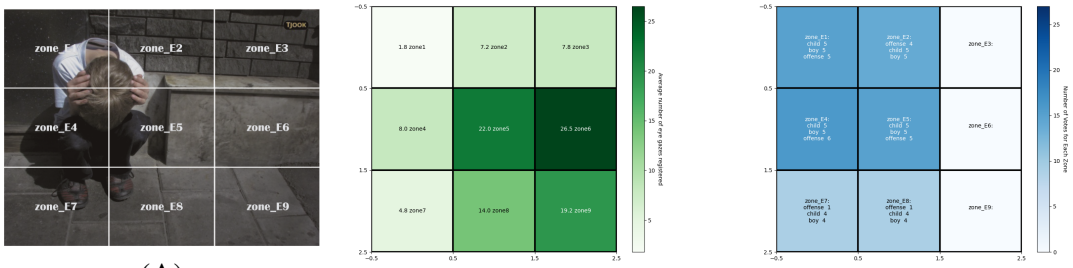
Table 6.2 – Cases in which there was a correspondence between the areas with most ocular interest and the most voted area for emotional content

Emotion	Number of images
happiness	12
neutral	11
sadness	10
disgust	6
fear	5
surprise	3
fear/sadness	3
neutral/sadness	2
happiness/neutral/surprise	2
neutral/surprise	1
disgust/fear	1
fear/happiness/neutral	1
disgust/sadness/surprise	1
fear/neutral/sadness	1
happiness/neutral	1
fear/happiness	1
fear/neutral	1
disgust/neutral	1
happiness/surprise	1

we can highlight that in this group there are images both associated with positive and neutral feelings, such as Happiness (11 images) and Neutral (7 images), and also linked to negative thoughts, such as Sadness (6 images), Disgust (5 images) and Fear (4 images);

(iii) the remaining 59 images, the area with most ocular interest coincided with zones in which there was emotional content, but which, according to the users' evaluation, did not correspond to zones especially relevant for the emotional content evaluated. In these last images, as it is possible to see in Table 6.5, the photographs associated with Happiness, are again the ones found in greater quantity.

Making a general overview of the correspondence between the zone of greater emotional interest and the three most voted zones (see Table 6.6), we can state that in 169 (67%) of the 252 images evaluated, there was a correspondence between the zone on which the user focused more attention during the analysis of these images, and one of the three zones with greater emotional charge, and therefore of greater emotional relevance. Moreover, as can be seen in Table 6.6, which compares these three types of correspondence, the emotion in which more correspondent cases were registered was Happiness.



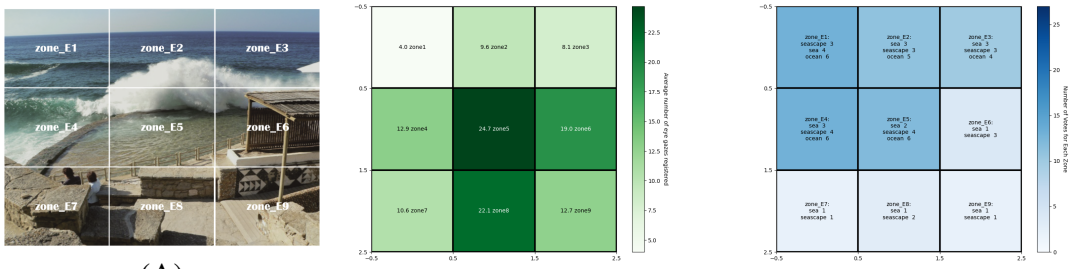
(A) (B) (C)

Figure 6.18 – Image from the Sadness category in which the area of ocular interest coincided with an area where no emotional content was registered: (A) analysed image; (B) heatmap resulting from the Eye Tracker data where the area with most ocular interest is highlighted in darker green (zone_E6); (C) heatmap resulting from the selection of the most suitable areas for the various emotional contents, where the Zone of Emotional Interest is highlighted in darker blue (zone_E4)

6.6 Discussion of Study Results

With this study, we intended to understand which area of an image was emotionally more relevant, evaluating for that purpose, if the area that was looked at for the longest time corresponded to the area with the greatest emotional charge, that is, the one that had the greatest number of registered votes.

We started the analysis of the results of this study by checking the agreement among the several users for the selection of the zones that best represented the emotional content associated to each one of the evaluated images. As previously mentioned, this verification was performed separately for each of the content tags associated with the image, since for the several tags it was possible to select the same zone of the image. From the results, we verified that in most of the cases there was a great variability in the choice of the



(A) (B) (C)

Figure 6.19 – Image from the Happiness category in which the area of ocular interest coincided with an area where was registered the second most voted zone: (A) analysed image; (B) heatmap resulting from the Eye Tracker data where the area with most ocular interest is highlighted in darker green (zone_E5); (C) heatmap resulting from the selection of the most suitable areas for the various emotional contents, where the potencial Zone of Emotional Interest is highlighted in darker blue (zone_E1 and zone_E4)

Table 6.3 – Cases in which there was a correspondence between the area with most ocular interest and the second most voted area(s) of emotional content

Emotion	Number of images
happiness	20
sadness	9
neutral	5
disgust	5
fear	5
neutral/surprise	2
happiness/surprise	2
surprise	2
disgust/happiness/neutral	1
disgust/surprise	1
disgust/sadness	1
fear/neutral/surprise	1
sadness/surprise	1
fear/happiness	1
anger	1
fear/happiness/neutral	1

most adequate zones for the tag under evaluation, and as such there was a poor or even fair agreement among the several evaluators of the images. These results for most of the images may be explained by the fact that some of the image evaluators, when selecting the zones that best represented each of the content tags, may have chosen to select only the zones in which the content under evaluation appeared in great prominence, while other users may have chosen to select all the zones in which that content appeared, even if it appeared in little or no prominence.

Table 6.4 – Cases in which there was a correspondence between the area with most ocular interest and the third most voted area(s) of emotional content

Emotion	Number of images
happiness	11
neutral	7
sadness	6
disgust	5
fear	4
disgust/surprise	2
neutral/sadness	2
happiness/neutral	2
surprise	2
fear/neutral/sadness	1
anger	1
fear/happiness	1
happiness/surprise	1
fear/sadness	1
fear/surprise	1

Table 6.5 – Cases in which there was a correspondence between the area with most ocular interest and the remaining zones where emotional content was present

Emotion	Number of images
happiness	22
neutral	7
sadness	5
happiness/neutral	5
fear	4
neutral/sadness	3
fear/happiness	2
disgust	2
fear/neutral/sadness	2
happiness/surprise	2
fear/sadness	1
anger/disgust/sadness	1
surprise	1
happiness/neutral/sadness/surprise	1
neutral/surprise	1

In the next analysis, we took the data obtained in this study and built heatmaps for each image, in order to visualize the distribution of the several emotional contents associated with each image, and identify which could be the zones with more emotional interest in each of the evaluated images. These heat maps, were built taking into account the total number of votes associated to each zone, and annotated according to the tags that were

Table 6.6 – Summary table of the images where there was a correspondence between the area of greatest ocular interest and one of the three most voted zones

Emotions	Most voted zone	Second most voted zone	Third most voted zone
anger		1	1
disgust	6	5	5
disgust/fear	1		
disgust/happiness/neutral		1	
disgust/neutral	1		
disgust/sadness		1	
disgust/sadness/surprise	1		
disgust/surprise		1	2
fear	5	5	4
fear/happiness	1	1	1
fear/happiness/neutral	1	1	
fear/neutral	1		
fear/neutral/sadness	1		1
fear/neutral/surprise		1	
fear/sadness	3		1
fear/surprise			1
happiness	12	20	11
happiness/neutral	1		2
happiness/neutral/surprise	2		
happiness/surprise	1	2	1
neutral	11	5	7
neutral/sadness	2		2
neutral/surprise	1	2	
sadness	10	9	6
sadness/surprise		1	
surprise	3	2	2

associated with the nine zones, in which the images were divided. The results obtained showed that, for all the evaluated zones, there were cases in which the zones with the highest number of votes were not only one, but several. These results can be explained by the fact that in some cases, the emotional content under evaluation was not in a specific zone, but occupied more then one zone of the image, or even almost all of it. Besides, it was also possible to verify that, in some cases, there were tags that did not reach their maximum number of votes in the zones that were considered by their evaluators as having the highest emotional interest, but in zones where their number of votes was the second highest. This fact, once again, can be explained by the possibility that users, in some cases, chose not to select all the zones where a certain content was present, but to select only those zones where the content under evaluation was more prominent.

Through this analysis, we also checked which zones had received the most number of votes (the most emotionally charged and therefore potentially the most relevant zone) and which had received the least number of votes, for each image evaluated. Through this verification, we saw that for a large number of images associated to the emotions Disgust, Fear, Happiness, Sadness, the most relevant zone was the centre of the image (zone_E5).

In the case of the images belonging to the Anger category, which has only two associated images, it was found that in one of the images the most voted zone was the zone in the centre (zone_E5) and the other had two zones as the most voted: zone_E7 and zone_E8. In the case of Surprise, as previously mentioned, the zones that could potentially be considered as the most relevant were in the centre of the image in two cases, in the zone_E6 in other two cases and in the remaining four images of this emotion the most voted zones were distributed by the zone_E8, zone_E2, zone_E1, and zone_E3. Finally, for the images associated to more than one emotion, the central zone of the image was once again the most voted zone, with the exception of the image associated to anger/disgust/sadness.

Additionally, also through this analysis, we tried to understand which zones had been considered less relevant for the emotional content of each of the evaluated images. As it was possible to notice by the data presented, for the emotions Anger, Disgust, Fear and Happiness, the least voted zone was zone_E3. In the case of the emotions Neutral and Sadness, the zone with the fewest votes was zone_E7, and for Surprise there was a tie between zone_E3 and zone_E9. As for the images that have more than one emotion associated, the less voted zones, for the majority of the cases were zone_E3, zone_E9 and zone_E1. These results demonstrate that in the majority of the associated cases, the relevant content that was to be evaluated was found in the middle line of the image.

Finally, in order to achieve the goals we had set ourselves with this study, we compared the results obtained in the Eye Tracker evaluation at the end of the previous study with the results obtained from the selection of the most adequate zones for each emotional content associated to each image. The analysis carried out allowed us to verify that in 169 of the 252 images evaluated, there was a correspondence between the zone looked at for the longest time and one of the three zones with the highest number of votes. Of this set of images, 64 had a correspondence between the area of greater ocular interest and the area with the most votes (area with a greater emotional charge). As such, we can conclude, as indicated by Sharafi et al. [39], that the users showed a great interest in the indicated area, and that consequently this will be the area responsible for the emotional reactions experienced during the moment of viewing the image. As for the remaining images in this group, in 58 the correspondence occurred with the zone(s) with the second highest number of votes, and in another 47 with the third highest number of votes. Additionally, we also verified the existence of 59 images in which the correspondence had occurred with zones where there was emotional content, but not as relevant as the photographs belonging to the group of images mentioned above.

Starting by analysing the cases where there was a correspondence with the second most voted zone(s), we could verify that most of the images in which this occurred corresponded to images associated with the emotion Happiness. The results obtained lead us to formulate the hypothesis that the emotional content responsible for the emotional reaction experience was found in these zones and not in those where there was a greater

emotional charge. This hypothesis, can also be put for the case of the images where the correspondence occurred with the zones where there was emotional content registered, but that were not part of the top 3 of the number of votes. The hypothesis in question is placed because, similarly to what happened with the second most voted zone(s), most of the images where that happened, were also associated to Happiness. Moreover, still on the cases where correspondence occurred with the second most voted zone, we can also put the hypothesis of some kind of unintentional interference with the device screen, which may have changed the calibration of the eye tracker. If this interference had existed, the registered zone could have been different from the expected one.

As for the cases in which the correspondence occurred with the third most voted zone(s), the majority of the cases in which this occurred corresponded to images associated with some kind of negative emotion (Disgust, Fear, Sadness) or images where these were one of the emotions associated with the evaluated images. Considering this, it is possible that the areas with higher emotional charge, had provoked in their evaluators some kind of discomfort, and that consequently led the user to look away to areas where that content was less present.

In general, for the cases where the correspondence occurred with the second and third most voted zones, or even with the other zones where there was registered content, but which did not belong to the group of the least voted zones, we can say that the choice of the zones with higher emotional charge came from the opinion of the users who participated in this last study. These users, as mentioned above, may have chosen not to select all the areas where the content to be evaluated appeared, but only the one(s) that they thought the content under evaluation was more visible, which may have influenced the determination of this zone.

It was also verified that, there were also some cases in which the area with most ocular interest had been registered in areas where there was no emotional content. In these cases, the results lead us to suspect that the content present in the area(s) that may be considered as the most relevant may have caused a great discomfort in the evaluator, leading the users to choose to look away to areas where there was no emotional content. This hypothesis is based on the fact that the images, where the area with most ocular interest was registered in areas where there was no emotional content associated, belong mostly to the Sadness category or were somehow associated to Sadness or Surprise.

Finally, there were also cases in which this correspondence occurred with the least voted zones. Similarly to what had occurred with the images where the correspondence was made with the third most voted zone, most of the images belong to negative emotions or are partly associated to it. As such, taking into account the emotions to which most of these images are associated, there is in this case, also the possibility that the content represented in the zones with higher emotional charge, have had some negative effect on the user, leading the user to look away to these zones where it was not especially relevant.

6.7 Summary

In this chapter, we presented the last study of our research, which had as its main objective to understand if the zones of the images that had been previously identified as the ones that would have obtained more attention from the user (looked at for longer period of time), corresponded to the one of the zones with greater emotional charge, or if on the other hand this was not the case.

We began our analysis by verifying the agreement between several users in relation to the zones that best represented the emotional content of each evaluated image. We realized that in most images there was a great variability in the choice of zones, producing a poor or weak agreement between the evaluators.

Next, in order to visualize how the emotional content had been distributed by the several zones, and try to understand which could be the most interesting zone of an image, we verified which zones had been selected for each emotional content and created categorical heatmaps, in which the respective content tags were pointed out in each zone. This analysis allowed us to verify that for most of the photographs belonging to the categories Disgust, Fear, Happiness, Neutral, Sadness and for most of the images associated to more than one emotion, the zone identified as being the one that could potentially be of more emotional interest was the centre of the image (zone_E5). In the case of the images of the Anger categories, it was found that in one of the images it was zone_E5 and in the other image associated to this emotion it was zones_E7 and zone_E8. In the case of the images of the Surprise category, in two of the images the zone identified as the most relevant was zone_E5, verifying for another two the zone_E6, and in the remaining four images of this emotion the zones_E8, zone_E2, zone_E1 and zone_E3. Regarding the image that was associated with anger/disgust/sadness, the zone with the most votes was zone_E4. Besides checking the most voted areas, we also analysed which ones would have received the least votes and, as such, the least relevant ones. This last part of this analysis allowed us to verify that for Anger, Disgust, Fear and Happiness, the zone that was in the list of the least voted was zone_E3. In the case of Neutral and Sadness, it was zone_E7, and for Surprise there was a tie between zone_E3 and zone_E9. As for the images with more than one associated emotion, for the generality of these images, the less voted zones were zone_E3, zone_E9 and zone_E1.

Finally, in order to achieve the objective that we had proposed for this study, we compared the results of the Eye Tracker analysis (previous study) with the results of the previous analysis. Through this, we verified that, contrary to what we had thought, only some of the areas which had been registered as area with most ocular interest corresponded to areas of emotional interest, and as such, they would in fact have been responsible for the emotional reactions experienced by the users. However, although in most cases, the areas of most ocular interest did not correspond to the areas with the highest recorded emotional charge, they did correspond to areas where emotional content was recorded, which may

indicate: i) the content that was in these zones attracted more attention from users than the zone(s) that received the most votes; ii) the zone of greater emotional interest may have led to a great or slight discomfort to the evaluator, causing the user to look away from that zone; iii) the existence of an unintentional interaction with the device screen, which caused a change in the eye tracker calibration; iv) as some users did not select all the zones where the content under evaluation was present, it may mean that the most voted zone, and consequently the one considered the most emotionally relevant, may not be the expected one, taking into account the results obtained in the previous study.

Chapter 7

General Research Discussion

With this work, we proposed not only to understand if the images presented could provoke some kind of emotional reaction in their observers, but also to identify what type of specific content would be responsible for provoking that reaction. To achieve our objectives, we developed three studies, each with different objectives.

In the first study, we set out to identify the content that best described the images, by identifying the content tags that best described them. In order to make this choice possible, we presented users with a set of 15 content tags that were most likely to be represented in each of the images, and which had resulted from the evaluation by the General model of the Clarifai API. Of these 15 tags, we asked users to select between three and 10 tags which, according to their opinion, best represented the content present in the images. The results showed that, although there was not total agreement between the various evaluators, in general, the agreement between them regarding the choice of tags was mostly favourable, with more than 50% of the evaluations having an moderate to very good agreement between the various evaluators. However, since we had defined the goal of selecting the five tags that best described the images, and since, in some of the evaluated images, the agreement between the evaluators was weak or even poor, it was necessary to define several criteria that would allow us to select these tags. Overall, this selection process allowed us to obtain the necessary tags for most of the images.

In the second study, we had two main goals: i) to verify if there was a connection between the zones looked at for the longest time and the emotional reactions experienced by the users; ii) to verify if there was a connection between the emotional reactions registered and the content of each image, represented by the tags selected during the evaluation. In order to achieve these objectives, we defined secondary objectives, which were intended to help us reach the necessary results: i) which tags were most associated with each emotion; ii) which tags were most associated with each polarity; iii) understand if users associated more than one emotion to each image; iv) if the associated emotions corresponded to the emotions to which the images were originally associated. In order to achieve the intended results, we colected data in three ways: WebGazer (eye tracker) - to

track the user's gaze; face-api js - system for recognition of facial expressions; emotional self evaluation by the users - with selection of the emotional self polarity, emotions and content tags for each emotion felt while viewing the images.

We started the analysis of our results by checking which tags were most associated to each image and identifying the three tags considered as the most relevant for each image. The results of this analysis, showed some variability in the choice of tags among the various evaluators. Despite the existence of this variability, there was no difficulty in identifying the necessary tags. In some cases it was even possible to identify more than three emotionally most relevant tags. However, during this process, we realized that, contrary to what would be expected, the third most voted tag for some of the images was other-none. This meant that in these cases, contrary to what we wanted, the content responsible for or part of an emotional reaction remained unidentified, since there was no tag that was favourable to this identification. Besides, there were still other cases, where there were more than three tags representing emotionally relevant content.

In the next phase, we verified which polarities were associated to each image. From the results obtained, we noticed that the values for the various polarities were similar. Furthermore, the analysis also revealed the existence of cases in which there were a tie in the votes between various polarities, which would end up being reflected in the association of more than one appropriate polarity for these images. As for the most appropriate emotions for each image, the analysis of their verification revealed, that in most images within a given emotional category, the users' votes, indicated a concordance between the voted emotion and the original emotion. However, there were two exceptions: Anger and Surprise. In Anger, the results revealed that none of its images was associated by the evaluators with the original emotion. In the case of Surprise, only a small percentage of these images was associated with Surprise. As for the existence of more than one emotion associated with the evaluated images, the results showed that in most cases, users associated only one emotion. However, there were in all emotional categories, cases in which there was not a total agreement among the several evaluators, which led to images with more than one associated emotion.

Regarding the tags, as we expected, the results both in terms of polarities and emotions ended up being a reflex of what had been verified for the images, and therefore it was possible to make a connection between the content and the emotional reactions registered by the users. Additionally, we also checked what type of content had been considered emotionally relevant. The results of this analysis showed that, contrary to what was thought, most of the content identified was generalist and not specific. This may mean that what led to the emotional reactions registered in most cases may not be something specific in the image, but the general idea transmitted. Finally, we also verified the coordinates registered by the eye tracker during the period of visualization of the images, and understood which area was looked at for more time in each one of them. In most cases, the zone

where more coordinates were registered was the centre of the image (zone_E5). However, contrary to what we had originally planned, it was not feasible for us to perform the analysis of the Face-api js data. As mentioned before, the participants' videos did not allow Face-api js to correctly identify the users' facial expressions, and therefore it ended up being impossible to compare its results, either with the determination of the most adequate emotions to describe the images and their content, or with the eye tracker data, in order to confirm if, in fact, the content that was looked at for the longest time was responsible, or not, for the registered emotional reaction. As such, of all the objectives we set out to achieve during this study, only the first main objective was not achieved.

Finally, in the third and last study, our main goal was to identify which area could be considered more relevant emotionally, taking advantage not only of the emotional content tags recorded in the previous study, but also of the areas looked at for longer in each of the images. In order to achieve this goal, we defined one secondary goal: i) Verify if the most looked at area have the greatest emotional charge.

We started the analysis of the results by verifying the agreement between the several evaluators, regarding the choice of the most adequate zones for each emotional tag. As it was possible to see by the results presented, there was a great variability in the choice of the zones, which led to a poor or weak agreement between the evaluators in most images. As already mentioned, these results may have been influenced by the fact that some of the evaluators chose to select only the zones in which the content was in great predominance, while others chose to select all the zones in which it appeared, even if it was not in great prominence in that zone.

In the next phase, in order to understand the most emotionally relevant zones, we checked which zones had the most votes for each of the images. The results obtained were separated according to the emotions with which the images had been associated during the previous study. Besides the most voted zones, we also tried to identify the less relevant ones, and so we also checked which zones were voted the least for each one of the images. The analysis allowed us to realize two things in relation to the most voted zones: i) some of the images had more than one zone as the most voted, and as such more than one candidate zone as the most relevant; ii) some of the emotional content under evaluation, did not reach its maximum value of votes in the most voted zones, but rather in the second most voted zones; iii) for five of the emotions to which the images were associated (Disgust, Fear, Happiness, Neutral and Sadness), and for most cases where there was more than one emotion associated to the image, the zone in the centre (zone_E5), revealed itself as the most relevant; iv) in the case of Anger and Surprise, we also identified the centre of the image as one of the most relevant, but there were still other zones as the most voted, as for example zone_E8, also in a similar number of images. As for the less voted zones, in general, the zone that appeared more times, as the less voted or one of the less voted, was zone_E3 (zone in the centre of the first row), and we can also highlight

zone_E9 (zone in the lower right corner of the image).

Moving now to the final analysis of this study, we compared the results of the voted zones for the emotional content of each image, with the results of the eye tracker analysis. For that, we created heatmaps for the voted zones and compared them with the heatmaps of the eye tracker records. As it was possible to verify by the results presented, in most of the images the area of greater ocular interest coincided with zones where there was emotional content registered. However, in six of the evaluated images, the area that was looked at longer coincided with an area in which no emotional content was registered. The results for these images, may indicate that there was some kind of discomfort of the users, regarding the content that was in the most emotionally relevant area.

As for the remaining images, in 169 of them, the correspondence occurred with one of the three most voted zones. Of this set of images, in only 64 were have a correspondence with the most emotionally charged zone. As such, the most probable hypothesis for these cases is that the zone that was looked at the longest probably corresponded (considering the associated charge) to the zone responsible for the emotional reactions registered for these images. As to the remaining images of this group, in 58 of them there was a correspondence with the one or one of the second most voted zones, while in the remaining 47 images, the correspondence occurred with the one or one of the third most voted zones.

Starting with the images in which there was a correspondence with the second most voted zone, we may hypothesize that the content responsible for the emotional reactions was found in this zone and not in the one(s) where there was a registered higher emotional charge. This possibility is raised, having in mind, as previously mentioned, the emotions with which these images are associated, which in their majority is Happiness. However, it is not possible for us to confirm this hypothesis with certainty, and it should be confirmed through future studies. Another possible hypothesis for these cases is that there was some kind of unintentional interference with the initial calibration of the eye tracker, for instance, an unintentional interaction with the user's screen. This interaction may have led the eye tracker to misidentify the image area. As for the images in which the correspondence occurred with the third most voted zone, considering the fact that most images are associated with a negative emotion (Disgust, Fear and Sadness), we hypothesize that the content registered in the zones with higher emotional charge may have been responsible for provoking some kind of discomfort in the evaluators. This situation may have led users to divert their gaze to areas where it was in less quantity.

As for the remaining images, those where the correspondence occurred with zones where emotional content was registered, but which were not part of the three most voted (59 images), similarly to the images associated with the category of the second most voted zones, these images are mostly associated with the emotion Happiness. As such, in these cases, we also hypothesize that the content responsible for the emotional reactions is found in these zones. However, it will also be necessary to confirm these results in future

work. Finally, in relation to the remaining 18 images, where the correspondence occurred with the less voted zones, we think that the results obtained may be explained by the fact that most of the images are associated to negative emotions. Taking that into account, and similarly to the images belonging to the category of the third most voted zones, users may have focused on zones where the content responsible for their discomfort was less present.

Making an overall assessment of the work, and taking into account the results obtained for the three studies, the first objective, which was to understand if an image could or could not provoke an emotional reaction, was achieved, although it was not possible, as previously mentioned, to compare the emotions registered by the users with the Face-api data, as initially planned. As for the second objective, it was certainly achievable for some of the images, remaining identified for the remaining cases, but with the need to implement a confirmation through another method in a future work.

Chapter 8

Conclusions

In this chapter, we present the dissertation summary, as well as the final contributions of our work and some limitations we went through during its development. We also present some ideas to explore in the future that will allow further development of the research we have carried out.

8.1 Summary of the Dissertation

In this work, we tried to understand if an image, used as a visual emotional stimulus, could provoke some kind of emotional reaction in its observer. Moreover, we also tried to identify what content responsible for the emotional reaction experienced by the user, when viewing each of the images.

In chapter 2, we briefly described some of the most important concepts of the various areas that make up this work, and some of the progress that has been made. We began by indicating that the classification of a visual stimulus could be done in three ways: categorization according to basic emotions, through a dimensional approach, and using evaluation criteria. Of these three forms of classification, we focus on describing the first two, which were used in this research. Emotional polarity, classifies a stimulus into negative or positive. The categorization by emotions varies according to the type of categorization done, which can be discrete, dimensional or componential. The discrete categorization takes into account basic and universal emotions, such as the six defined by Ekman (anger, disgust, fear, happiness, sadness and surprise). In the case of dimenisonal categorization, an evaluation of the stimulus is made according to the dimensions of valence, arousal and dominance. The componential categorization classifies visual stimuli according to the dimensions of pleasure, arousal, control and conductiviness.

As for emotional polarity, we also proposed a third polarity, neutral, for stimuli that had not provoked any type of emotional reaction. Still in relation to emotions, we also made a brief review about the datasets that used.

We also talked about Tags and the two ways in which this annotation can be per-

formed: explicit tagging - annotation made by the user himself; and implicit tagging - annotation taking into account the user's reactions to the evaluated content. We also mentioned the three levels of abstraction used to evaluate explicit tagging: subordinate (specific terms), basic (attributes common to all or almost all members of a category) and superordinate (generic terms). We also talk about the Panofsky-Shatford matrix for interpreting the type of annotated content, which combines Panofsky's framework, which takes into account three levels of evaluation (pre-iconography (general), inconographical (specific) and iconology (abstract)), with Shatford's framework, an adaptation of the previous one, whose evaluation is based on the answer to the questions who?, what?, where? and when?

Next, we talked about eye tracking technology, which is responsible for measuring and recording the eye movements of an individual in the presence of a stimulus, and which has the ability to understand which areas the user has fixed their gaze on, for how long the fixation took place and the order in which the visual exploration occurred. Furthermore, we also mention that this technology needs to be integrated in devices called eye trackers, for which there are two types: intrusive (they require a physical structure and transportation) and non-intrusive (they track the gaze remotely). Finally, we also talk about Facial Expression Recognition Systems, a type of face recognition system, where models that perform Facial Expression Recognition (FER) are integrated. As mentioned, the most recent ones have integrated Convolutional Neural Networks (CNNs), which recognise and evaluate emotional reactions in real time.

In chapter 3, we present the main tools for the development of this work, as well as the general approach for its development.We began by indicating the approach we would follow, where we indicated that our strategy would involve combining two types of information: the area looked at for the longest time, identified by eye tracking, and the emotional content, identified by content tags related to the content responsible for an emotional reaction. Furthermore, we informed that all data would be obtained remotely, due to the pandemic situation we were in, and that our work would be divided into three studies, each of which with its own specific objective.In the final section of the chapter, we present the three main software tools used: the General model of the Clarifai API (a machine learning model that has the ability to analyse an image and return a set of concepts representative of its content), WebGazer (an eye tracking library that uniquely uses the user's browser and webcam device to track their gaze), and Face-api js (a machine learning library that has built-in Convolutional Neural Networks (CNNs) to evaluate a user's facial expressions in real time).

In chapter 4, we presented the first study carried out within the scope of this work, whose objective was the identification of the most relevant content for each image. We began by describing the process for the selection of the set of images, which would later be used throughout the work. This process, which included several criteria, resulted in

a set of 252 images - six sets of 42 images, each one representing one of the six basic universal emotions defined by Ekman. Next, we proceeded to identify the concepts that best described each image, whose identification was performed by the General model of the Clarifai API, which analysed each image, and returned for each one 30 content tags, followed by the probability of their presence in the image. After that, the tags were filtered by us, resulting in the end in only 15 most representative ones.

After presenting these concepts to the users, who selected between three and 10 tags, we proceeded to the identification of the five most voted tags. During the identification process, we faced some variability in the choice of tags and, therefore, it was necessary to create several criteria. Although in most cases this process allowed for the selection of the necessary tags, there were two exceptional cases where it was necessary to choose the fifth tag from the list of 15 initial tags. Next, we checked the quality of the results obtained. We start by verifying what would have been the average number of tags selected per user for each image, where we verify that it would have been 5.59, with a standard deviation of 1.70. Besides, through this analysis we can also verify that the minimum average value registered would have been three and the maximum 9.25. Lastly, we also checked the level of agreement between users. To measure this agreement, we used the inter-rater agreement measure Fleiss' Kappa, whose average of the values obtained for the dataset was 0.51, indicating that the overall agreement was moderate, with a standard deviation of 0.23. Looking more closely at the results, we realised that this value was a reflection of the fact that more than 50% of the images had moderate to very good agreement.

In chapter 5, we presented the second study of this work, which had as its main objective, to correlate the emotional reactions experienced by users with the area that was looked at for the longest time, and also to understand if these reactions were linked to the content present in the image. Taking advantage of the results of the study, we started by trying to understand which tags were considered as the most relevant for each image. This analysis resulted in sets of three or more tags. Next, we tried to understand which polarity was the most associated to each image. The results of the analysis indicated the existence of images associated to each of the three polarities, but also some cases in which there was no agreement, resulting in images with more than one polarity. In the following analysis, we went to check which emotion was the most associated with each of the images. As expected, in most of the images, there was a correspondence between the voted emotion and the original one. However, in the case of Anger and Surprise, the most voted emotion for most of the images of these emotions, was Happiness. Moreover, through this analysis, we also verified the existence of images where there was no concordance regarding the most appropriate emotion, resulting in images with more than one associated emotion.

In the next step, we verified which polarity and emotion was more associated to each of the content tags. We verified that, in most cases, the observed results were a reflection of the ones obtained for the images. Additionally, we also tried to understand what kind of

tags were selected and if the polarity associated to a tag corresponded to the expected one, taking into account the emotion associated to it. In general, the results showed that most of the selected tags were generalist and that the polarity corresponded to the expected one. However, two exceptions were detected, where the positive emotion Happiness was associated with a negative polarity. Finally, we also analysed the results obtained by the eye tracker, in order to understand which could be considered the most relevant area of each image. The analysis results showed that in most cases, this zone was the centre of the image (zone_E5).

In chapter 6, we presented the last study of this work. There we intended to understand if the area which was looked at for the longest time corresponded to the area with the highest emotional charge. We began the analysis of the results by checking whether there was agreement among users regarding the choice of the most appropriate areas for each content evaluated. The results showed that the mean value obtained for each image, for the Fleiss' kappa measure, used to evaluate this agreement, was indicative, of a poor or weak agreement for most images. Next, we tried to understand which could be considered the zones with higher and lower emotional charge, by identifying the zones with higher and lower number of votes for tags, through the creation of heatmaps and the counting of the number of votes for each zone. The results obtained for this analysis, showed that for most images, of most emotions or with more than one emotion, this zone corresponded to the centre of the image (zone_E5). However, there were exceptions, of which we highlight: Anger - one image in which it was zone_E5, and another at zone_E7 and zone_E8 (tie in the number of votes); Surprise - two images at zone_E5, two at zone_E6, and the remaining four at zone_E8, zones_E2, zone_E1 and zone_E3, respectively; anger/disgust/sadness - in zone_E4. As for the least voted, in most cases this corresponded to zone_E3. However, there were also exceptions: Neutral and Sadness - zone_E7, Surprise - zone_E3 and zone_E9; more than one associated emotion - zone_E3, zone_E9 and zone_E7.

Finally, we went to compare the results of the analysis that had identified the zones looked at the longest, with the results of the previous analysis. Of the images analysed, in only 169 of these images was there a match with one of the three most voted zones, of which only 64 images had a correspondence between the most voted zones and the zones looked at the longest. Besides, it was also found that in six images, the zone that was looked at the longest corresponded to zones where no type of emotional content was registered, which may indicate discomfort in relation to the content presented. As for the remaining images, the correspondence always occurred with areas where emotional content was found, but which were not the most emotionally charged areas, which may mean: i) more interest in the content of those areas; ii) some discomfort regarding the content in the area with higher emotional charge; iii) unintentional interaction with the eye tracker calibration.

In chapter 7, we state the general discussion of the entire research, and evaluate whether or not we had achieved our intended objectives. We started the analysis, by discussing the results of the first study, in which we intended to identify the content tags that best represented each image. As indicated in this chapter and also through the results already presented, this objective was achieved for most cases, with the exception of the two cases in which it was necessary for us to select the fifth tag ourselves, among those that had been initially presented.

Next, we analyse the results of the second study, where we intended not only to relate the emotional reactions to the area looked at the longest but also to the content present in the images. As previously indicated, in most cases it was possible to identify the whole content responsible for a given emotional reaction. However, there were images in which the third tag identified as the most relevant was other-none. As such, in these cases, part of the emotional content remained unidentified. As for the emotional reactions experienced and areas looked at for the longest time, the analyses of the results obtained allowed us to obtain these two types of information. Besides, it was also possible to correlate the emotional reactions with the content of the images. On the other hand, as it was not possible to analyse the data recorded by Face-api js, we were not able to compare this information neither with the emotional reactions recorded nor with the area looked at for the longest time. As such, it was neither possible to confirm if the indicated reaction was in fact the most suitable, nor to understand if the area looked at the longest was responsible for the reactions experienced. In this way, only one of the main objectives ended up being fully accomplished.

Finally, we analysed the results of the third and last study, which aimed to understand if the area looked at the longest was the one with the greatest emotional charge. As we can see from the results obtained, this comparison was possible. However, in only 64 of the 252 images we identified a correspondence between these two areas. Analyzing in a general way all the results, as indicated in this chapter, it was possible to achieve both general objectives for most of the images, although it is necessary in future works to implement of other methods and improve the already existing ones.

8.2 Contributions and Limitations

Over the years, several studies have been carried out in the area of emotion recognition, which use images to perform this recognition.However, so far we are not aware of any other work that has tried to do this recognition by combining the following two types of information: content annotation and the areas looked at during the image analysis. As such, by doing this work, we have tried to combine these two types of information in order to carry out this identification, and to contribute with:

- Process for annotating the most relevant content for each of the images, through the

selection of the content tags that best represent it;

- Procedure for the emotional categorization of an image, taking into account the spontaneous emotional reactions resulting from the evaluation of an image;
- Procedure for the identification of the emotional content responsible for an emotional reaction, whose identification results from the conjugation of three information: i) eye tracker data, recorded during the viewing of the image; ii) content tag marked at the moment of the emotional evaluation of the image; iii) selection of the image zones that best fit each content tag selected;
- Dataset composed of 252 images, coming from the EmotionROI dataset, annotated with two types of information: i) eye tracker data - average number of coordinates registered in each zone; ii) registration of the zones for each emotional content tag - number of votes registered in each of the zones considered as relevant for that content. These annotations are in both Json file and heatmap format.

As mentioned throughout this study, it had some limitations to its development. One of the first limitations was related to the fact that all studies were conducted online, which meant that we did not have the full control over the conditions of their realization. This limitation ended up affecting the collection of data by Face-api js, preventing from using this data. The data collected with the WebGazer, was also affected, since a large amount of data was registered outside the image boundaries in some of the users. Besides that, the eye tracker also had the problem of not being the most stable one, and of not being able to track the screen corners very well, although the calibration created tried to fix that problem. Finally, there was also the issue that we had to limit the second study to only some browsers and devices, due to the limitations of the software used.

8.3 Future Work

Looking at the results obtained in this work, and the limitations pointed out in the previous section, we can establish some guidelines for future work. In relation to the emotional content that remained unidentified in some of the images, because the third most voted tag was other-none, we consider important that in a future work the users should have the possibility of adding new tags. This content could then be integrated into categories of content tags, which would be composed of synonymous tags.

In relation to the eye tracker, we think that a solution for the problems encountered would be the use of another eye tracker, more stable and with the ability to track the entire screen in a more consistent and precise way. It would also be important to compare this data with the data obtained in this work, in order to confirm our conclusions. Moreover, in order to better control the conditions for obtaining the necessary data, and additionally

to make sure that the conditions necessary for obtaining the data from Face-api js are respected, it would be important that this data be obtained in person from a single device. Additionally, similarly to what would be done with the data obtained with the new eye tracker, it would also be important to compare the Face-api data with the data from the emotional analysis performed to the images in this work, and from the eye tracker, in order to confirm if in fact the identified zones correspond to the ones responsible for the recorded emotional reactions.

Bibliography

[1] face-api.js. "https://justadudewhohacks.github.io/face-api.js/docs/index.html". (accessed: 01.06.2021).

[2] Tensorflow.js. "https://github.com/tensorflow/tfjs". (accessed: 01.06.2021).

[3] Laerd Statistics (2019). Fleiss' kappa using spss statistics. "https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php", September 2021. Statistical tutorials and software guides.

[4] Soraia M Alarcao and Manuel J Fonseca. Enriching iaps and gaped image datasets with unrestrained emotional data. In *DMSVIVA*, pages 57–64, 2018.

[5] Douglas G Altman. *Practical statistics for medical research*. CRC press, 1990.

[6] Patrícia Arriaga and Gisela Almeida. Fábrica de emoções: A eficácia da exposição a excertos de filmes na indução de emoções. *Laboratório de Psicologia*, 8(1):63–80, 2010.

[7] Ana M Barreto. Eye tracking como método de investigação aplicado às ciências da comunicação. *Revista Comunicando*, 1(1):168–186, 2012.

[8] Renata G Bianchi, Vânia PA Neris, and Anderson L Ara. Tags vs. observers–a study on emotions tagged and emotions felt with flickr pictures. *Multimedia Tools and Applications*, 78(15):21805–21826, 2019.

[9] Tobias Brosch, Gilles Pourtois, and David Sander. The perception and categorisation of emotional stimuli: A review. *Cognition and emotion*, 24(3):377–400, 2010.

[10] Diogo Carleto. Face-api.js: Javascript face recognition leveraging tensorflow.js. "https://www.infoq.com/news/2018/11/faces-api-js/". (accessed: 01.06.2021).

[11] Moran Cerf, E Paxon Frady, and Christof Koch. Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of vision*, 9(12):10–10, 2009.

[12] Clarifai. Everything you need to label, build, and deploy ai models for unstructured data. "https://www.clarifai.com/products/platform". (accessed: 20.05.2021).

[13] Clarifai. Image recognition ai model. "https://www.clarifai.com/models/image-recognition-ai". (accessed: 20.05.2021).

[14] Osvaldo Da Pos and Paul Green-Armytage. Facial expressions, colours and basic emotions. *Colour: design & creativity*, 1(1):2, 2007.

[15] Elise S Dan-Glauser and Klaus R Scherer. The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, 43(2):468–477, 2011.

[16] Paul Ekman. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200, 1992.

[17] Shaojing Fan, Zhiqi Shen, Ming Jiang, Bryan L Koenig, Juan Xu, Mohan S Kankanhalli, and Qi Zhao. Emotional attention: A study of image sentiment and visual attention. In *Proceedings of the IEEE Conference on computer vision and pattern recognition*, pages 7521–7531, 2018.

[18] Jennifer Golbeck, Jes Koepfler, and Beth Emmerling. An experimental study of social tagging behavior and image content. *Journal of the American Society for Information Science and Technology*, 62(9):1750–1760, 2011.

[19] Behzad Hasani and Mohammad H Mahoor. Facial expression recognition using enhanced deep 3d convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 30–40, 2017.

[20] Zhong-qing Jiang, Wen-hui Li, Ying Liu, Yue-jia Luo, Phan Luu, and Don M Tucker. When affective word valence meets linguistic polarity: behavioral and erp evidence. *Journal of Neurolinguistics*, 28:19–30, 2014.

[21] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017.

[22] Christina Katsimerou, Joris Albeda, Alina Huldtgren, Ingrid Heynderickx, and Judith A Redi. Crowdsourcing empathetic intelligence: the case of the annotation of

emma database for emotion and mood recognition. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 7(4):1–27, 2016.

[23] Claudia Kawai, Gáspár Lukács, and Ulrich Ansorge. Polarities influence implicit associations between colour and emotion. *Acta Psychologica*, 209:103143, 2020.

[24] Peter J Lang, Margaret M Bradley, Bruce N Cuthbert, et al. International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 1(39-58):3, 1997.

[25] Petri Laukka, Patrik Juslin, and Roberto Bresin. A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, 19(5):633–653, 2005.

[26] Chenyang Li and Chunfang Li. Web front-end realtime face recognition based on tfjs. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–5. IEEE, 2019.

[27] Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. The nencki affective picture system (naps): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior research methods*, 46(2):596–610, 2014.

[28] V Mekala, Vibin Mammen Vinod, M Manimegalai, and K Nandhini. Face recognition based attendance system. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(12):520–525, 2019.

[29] Joseph A Mikels, Barbara L Fredrickson, Gregory R Larkin, Casey M Lindberg, Sam J Maglio, and Patricia A Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4):626–630, 2005.

[30] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.

[31] Maja Pantic and Alessandro Vinciarelli. Implicit human-centered tagging [social sciences]. *IEEE Signal Processing Magazine*, 26(6):173–180, 2009.

[32] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer.js democratizing webcam eye tracking on the browser. "`https://webgazer.cs.brown.edu/`". (accessed: 01.06.2021).

[33] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence - IJCAI 2016*,

2016-January:3839–3845, 2016. International Joint Conferences on Artificial Intelligence.

[34] Kuan-Chuan Peng, Amir Sadovnik, Andrew Gallagher, and Tsuhan Chen. Where do emotions come from? predicting the emotion stimuli map. pages 614–618, 2016.

[35] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350, 2001.

[36] Jonathan Posner, James A Russell, and Bradley S Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, 17(3):715–734, 2005.

[37] Monika Riegel, Łukasz Żurawski, Małgorzata Wierzba, Abnoss Moslehi, Łukasz Klocek, Marko Horvat, Anna Grabowska, Jarosław Michałowski, Katarzyna Jednoróg, and Artur Marchewka. Characterization of the nencki affective picture system by discrete emotional categories (naps be). *Behavior research methods*, 48(2):600–612, 2016.

[38] Stefanie Schmidt and Wolfgang G Stock. Collective indexing of emotions in images. a study in emotional information retrieval. *Journal of the American Society for Information Science and Technology*, 60(5):863–876, 2009.

[39] Zohreh Sharafi, Bonita Sharif, Yann-Gaël Guéhéneuc, Andrew Begel, Roman Bednarik, and Martha Crosby. A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering*, 25(5):3128–3174, 2020.

[40] Sara Shatford. Analyzing the subject of a picture: a theoretical approach. *Cataloging & classification quarterly*, 6(3):39–62, 1986.

[41] Sylvain Sirois and Julie Brisson. Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(6):679–692, 2014.

Appendix A

Appendixes

Table A.1 – Fleiss' Kappa analysis table for each of the 5 ideal tags

Table 1: Fleiss' Kappa measurement for the 5 ideal tags	
Image	**Value of K**
anger_105	0,3631
anger_115	0,4715
anger_118	0,5210
anger_119	0,5960
anger_12	0,6768
anger_131	0,3229
anger_135	0,5181
anger_144	0,2982
anger_147	0,7576
anger_154	0,5501
anger_175	0,0331
anger_178	0,8773
anger_184	0,5181
anger_198	0,6386
anger_203	0,8773
anger_207	0,8773
anger_21	0,5960
anger_220	0,5526
anger_222	0,6386
anger_226	0,5181
anger_232	0,5000
anger_254	0,3631
anger_258	0,3819
anger_262	0,2982
anger_267	0,5960
anger_281	0,2237
anger_287	0,1383
anger_291	0,1367
anger_302	0,3631
anger_35	0,6386
anger_37	0,8773
anger_39	0,1282
anger_41	0,5501
anger_42	1,0000
anger_67	0,7576
anger_79	0,7576
anger_81	0,4203
anger_82	0,2982
anger_86	0,6386
anger_88	0,7576
anger_96	0,7576

Continued on next page

Table A.1 – continued from previous page

Image	Value of K
anger_97	0,6748
disgust_111	0,4203
disgust_12	0,2394
disgust_124	0,4444
disgust_143	0,2157
disgust_144	0,3103
disgust_154	0,5960
disgust_164	0,2827
disgust_168	0,3631
disgust_173	0,8773
disgust_18	0,7576
disgust_180	0,3416
disgust_181	0,8773
disgust_19	0,3819
disgust_194	0,3416
disgust_195	0,5501
disgust_197	0,4715
disgust_202	0,5960
disgust_203	0,2827
disgust_205	0,2271
disgust_220	0,8773
disgust_221	0,4118
disgust_243	0,5181
disgust_249	0,4203
disgust_254	0,4715
disgust_263	0,2982
disgust_284	0,5501
disgust_291	0,4203
disgust_298	0,3631
disgust_303	0,2157
disgust_307	0,4203
disgust_309	0,1367
disgust_316	0,5960
disgust_319	0,2453
disgust_39	0,3187
disgust_5	1,0000
disgust_73	0,6386
disgust_75	0,3631
disgust_8	0,6768
disgust_84	0,6386
disgust_85	0,3416
disgust_86	0,4203
disgust_93	0,2237

Continued on next page

Table A.1 – continued from previous page

Image	Value of K
fear_10	0,5501
fear_105	0,8773
fear_112	0,7576
fear_113	0,1367
fear_120	0,1282
fear_125	0,4343
fear_126	0,5807
fear_130	0,8773
fear_136	0,5960
fear_143	0,7576
fear_144	0,2982
fear_150	1,0000
fear_170	0,3819
fear_211	0,5181
fear_222	0,5960
fear_243	0,3981
fear_248	0,6386
fear_251	0,2827
fear_266	0,0476
fear_275	0,2237
fear_276	0,6386
fear_283	0,6868
fear_286	0,6386
fear_287	0,5960
fear_288	0,2982
fear_292	0,5501
fear_294	0,4203
fear_296	0,7576
fear_299	0,0331
fear_300	0,2827
fear_301	0,2827
fear_31	0,3631
fear_313	0,4667
fear_328	0,2827
fear_44	0,0476
fear_45	0,5501
fear_49	0,5501
fear_52	0,5960
fear_55	0,3939
fear_62	0,5269
fear_71	0,4444
fear_79	0,2827
joy_107	0,6386

Continued on next page

Table A.1 – continued from previous page

Image	Value of K
joy_108	0,3819
joy_11	1,0000
joy_110	0,4715
joy_111	0,5181
joy_115	0,6386
joy_121	0,7576
joy_122	0,7172
joy_13	0,7576
joy_14	0,1383
joy_148	0,3939
joy_15	0,5181
joy_153	0,6386
joy_156	0,8773
joy_159	1,0000
joy_16	0,8773
joy_206	0,5501
joy_214	0,8773
joy_236	0,5960
joy_237	1,0000
joy_251	0,6386
joy_270	0,6386
joy_275	0,2157
joy_292	0,5181
joy_302	1,0000
joy_310	0,1383
joy_314	0,2982
joy_325	0,5501
joy_33	0,1282
joy_35	0,7576
joy_36	0,1282
joy_38	0,5181
joy_4	0,7576
joy_46	0,6386
joy_48	0,1367
joy_55	0,4715
joy_66	0,0476
joy_70	0,6386
joy_72	0,6774
joy_73	0,7576
joy_98	0,8773
joy_99	0,3631
sadness_101	0,2827
sadness_107	0,8773

Continued on next page

Table A.1 – continued from previous page

Image	Value of K
sadness_115	0,4715
sadness_124	0,7576
sadness_125	0,3939
sadness_147	0,2982
sadness_15	0,3819
sadness_151	0,8773
sadness_152	0,2237
sadness_160	0,3819
sadness_173	0,6386
sadness_18	0,4715
sadness_185	0,3103
sadness_19	0,6386
sadness_190	0,3103
sadness_196	0,4203
sadness_199	0,3819
sadness_20	0,6296
sadness_218	0,5181
sadness_221	0,3939
sadness_232	0,4715
sadness_238	0,5807
sadness_254	0,7576
sadness_27	0,2271
sadness_277	0,5501
sadness_296	0,4203
sadness_300	0,2982
sadness_306	0,4203
sadness_327	0,5960
sadness_327	0,4203
sadness_33	0,4444
sadness_330	0,4715
sadness_44	0,6386
sadness_45	0,7576
sadness_48	0,2271
sadness_51	0,4203
sadness_6	0,3819
sadness_63	0,8773
sadness_72	0,5501
sadness_8	0,4444
sadness_80	0,5181
sadness_99	0,4203
surprise_110	0,5181
surprise_114	0,8374
surprise_115	0,2157

Continued on next page

Table A.1 – continued from previous page

Image	Value of K
surprise_13	0,2827
surprise_135	0,7576
surprise_138	0,5501
surprise_139	0,7576
surprise_15	0,3416
surprise_168	0,5960
surprise_187	0,7576
surprise_190	0,6386
surprise_204	0,7576
surprise_22	0,2237
surprise_220	0,4715
surprise_221	0,6386
surprise_222	0,7576
surprise_225	0,6386
surprise_238	1,0000
surprise_245	0,5960
surprise_25	0,4444
surprise_250	0,7576
surprise_273	0,7172
surprise_29	1,0000
surprise_303	0,2827
surprise_305	0,2157
surprise_309	0,2827
surprise_313	0,3478
surprise_32	0,7576
surprise_323	0,3187
surprise_324	0,8773
surprise_35	0,6386
surprise_38	0,3416
surprise_48	0,3103
surprise_53	0,6386
surprise_6	0,5181
surprise_65	0,3416
surprise_66	0,5269
surprise_69	0,7576
surprise_83	0,2157
surprise_85	0,3631
surprise_93	0,7576
surprise_96	0,4715

Table A.2 – Fleiss' Kappa analysis table for each tag associated to each image

Table 2: Fleiss' Kappa measurement each tag of each image			
Image	**tag**	**Value of K**	**Average Value of K for image**
anger_105	house	-0,042	-0,04
anger_105	tree	-0,042	
anger_115	branch	0,221	0,08
anger_115	cold	0,076	
anger_115	fall	0,091	
anger_115	nature	0,018	
anger_115	tree	-0,009	
anger_118	architecture	-0,184	-0,11
anger_118	bridge	-0,121	
anger_118	city	-0,164	
anger_118	river	-0,170	
anger_118	sky	0,065	
anger_119	light	0,027	0,08
anger_119	people	0,239	
anger_119	police	0,026	
anger_119	street	0,023	
anger_12	art	0,205	-0,06
anger_12	cola	-0,220	
anger_12	soda	-0,193	
anger_12	symbol	-0,014	
anger_131	barbed wire	-0,169	0,04
anger_131	danger	0,231	
anger_131	horror	0,083	
anger_131	security	0,022	
anger_135	city	-0,144	-0,06
anger_135	flood	0,031	
anger_135	storm	-0,078	
anger_144	rapids	0,073	0,36
anger_144	rock	0,650	
anger_147	ocean	-0,125	0,08
anger_147	power	0,413	
anger_147	water	-0,063	
anger_154	black and white	-0,232	-0,14
anger_154	girl	0,083	
anger_154	monochrome	-0,175	
anger_154	ocean	-0,250	
anger_175	cereal	-0,124	-0,13
anger_175	sky	-0,138	
anger_178	sea	-0,040	0,00
anger_178	wave	0,045	
anger_184	ocean	0,124	0,24

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
anger_184	sea	0,354	
anger_198	ocean	-0,076	0,03
anger_198	rock	0,143	
anger_203	abandoned	0,112	0,05
anger_203	building	-0,030	
anger_203	decay	0,073	
anger_207	ocean	-0,095	-0,04
anger_207	sea	0,094	
anger_207	water	-0,117	
anger_21	field	0,539	0,16
anger_21	girl	0,167	
anger_21	people	-0,232	
anger_21	woman	0,167	
anger_220	girl	-0,139	-0,11
anger_220	light	-0,091	
anger_220	woman	-0,100	
anger_222	sky	0,002	0,09
anger_222	snow	0,078	
anger_222	street	0,169	
anger_222	winter	0,110	
anger_226	field	0,554	0,31
anger_226	sky	0,166	
anger_226	storm	0,197	
anger_232	H2O	-0,212	-0,03
anger_232	liquid	-0,196	
anger_232	splash	0,318	
anger_254	art	-0,218	-0,23
anger_254	smoke	-0,250	
anger_258	ocean	0,167	0,03
anger_258	sky	-0,250	
anger_258	sunset	0,178	
anger_262	evening	0,360	0,35
anger_262	sea	0,286	
anger_262	seascape	0,615	
anger_262	sunset	0,135	
anger_267	nature	0,229	0,05
anger_267	tree	-0,123	
anger_281	river	-0,170	-0,08
anger_281	rock	-0,081	
anger_281	scenic	0,002	
anger_287	monochrome	-0,211	-0,09
anger_287	portrait	-0,141	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
anger_287	woman	0,078	
anger_291	abstract	-0,162	-0,10
anger_291	art	-0,169	
anger_291	line	-0,171	
anger_291	texture	0,111	
anger_302	bloody	-0,250	0,12
anger_302	horror	0,168	
anger_302	scary	0,435	
anger_35	action	-0,015	-0,04
anger_35	athlete	-0,136	
anger_35	exercise	-0,017	
anger_35	motion	0,024	
anger_37	ocean	-0,212	-0,02
anger_37	wave	0,171	
anger_39	art	0,526	0,53
anger_39	H2O	0,531	
anger_39	sooty	0,533	
anger_41	building	0,107	-0,02
anger_41	demolition	-0,044	
anger_41	waste	-0,135	
anger_42	ocean	-0,184	0,05
anger_42	sea	0,298	
anger_42	seascape	0,043	
anger_67	beach	0,447	0,06
anger_67	ocean	0,060	
anger_67	sea	-0,125	
anger_67	water	-0,140	
anger_79	nature	-0,144	0,12
anger_79	waterfall	0,375	
anger_81	surf	0,150	0,08
anger_81	surfboarding	0,215	
anger_81	water	-0,140	
anger_82	closeup	0,598	0,19
anger_82	color	-0,072	
anger_82	heart	0,038	
anger_86	black and white	0,187	-0,09
anger_86	ocean	-0,232	
anger_86	storm	-0,233	
anger_88	beach	0,372	0,12
anger_88	ocean	-0,016	
anger_88	sea	0,003	
anger_96	landscape	0,108	0,08
anger_96	mountain	0,262	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
anger_96	nature	-0,120	
anger_97	eye	0,167	0,00
anger_97	face	-0,139	
anger_97	skin	-0,035	
disgust_111	animal	-0,221	-0,23
disgust_111	slimy	-0,221	
disgust_111	wet	-0,239	
disgust_12	hole	-0,250	-0,04
disgust_12	nest	0,167	
disgust_124	bird	-0,176	0,09
disgust_124	broken	0,364	
disgust_143	dirty	0,050	-0,02
disgust_143	nature	0,062	
disgust_143	water	-0,171	
disgust_144	beach	-0,150	-0,15
disgust_144	fish	-0,232	
disgust_144	nature	-0,200	
disgust_144	sand	0,000	
disgust_154	fish	0,722	0,49
disgust_154	food	0,265	
disgust_164	animal	-0,181	0,07
disgust_164	skull	0,327	
disgust_168	outdoors	0,195	0,24
disgust_168	pollution	0,288	
disgust_173	ash	0,244	0,06
disgust_173	cigar	-0,235	
disgust_173	trash	0,176	
disgust_18	animal	0,149	-0,03
disgust_18	fish	0,004	
disgust_18	head	-0,250	
disgust_180	abandoned	0,017	0,04
disgust_180	garbage	0,088	
disgust_180	waste	0,019	
disgust_181	mud	0,176	0,21
disgust_181	soil	0,252	
disgust_19	cadaver	-0,235	-0,20
disgust_19	carp	-0,206	
disgust_19	fish	-0,164	
disgust_194	color	-0,167	-0,12
disgust_194	texture	-0,098	
disgust_194	wall	-0,094	
disgust_195	animal	-0,094	0,22
disgust_195	environment	0,368	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
disgust_195	pollution	0,250	
disgust_195	waste	0,354	
disgust_197	building	-0,120	0,24
disgust_197	roof	0,608	
disgust_202	garbage	0,180	0,20
disgust_202	trash	0,217	
disgust_203	closeup	0,230	0,08
disgust_203	wildlife	-0,069	
disgust_205	closeup	0,212	-0,04
disgust_205	pain	-0,149	
disgust_205	skin	-0,172	
disgust_220	cooking	0,289	0,32
disgust_220	food	0,201	
disgust_220	lunch	0,348	
disgust_220	meal	0,434	
disgust_221	dirty	0,125	0,25
disgust_221	old	0,306	
disgust_221	rusty	0,306	
disgust_243	building	0,190	0,17
disgust_243	people	0,148	
disgust_249	blood	0,011	-0,04
disgust_249	bloody	-0,093	
disgust_254	garbage	-0,230	-0,05
disgust_254	street	0,302	
disgust_254	trash	-0,207	
disgust_263	insect	0,233	0,23
disgust_263	invertebrate	0,233	
disgust_284	abandoned	0,060	0,09
disgust_284	adult	0,306	
disgust_284	waste	-0,102	
disgust_291	adult	-0,165	-0,09
disgust_291	barbecue	-0,052	
disgust_291	festival	-0,139	
disgust_291	food	0,083	
disgust_291	man	-0,168	
disgust_298	fungus	0,098	0,25
disgust_298	nature	0,401	
disgust_303	animal	-0,067	-0,02
disgust_303	river	-0,076	
disgust_303	water	-0,099	
disgust_303	wildlife	0,148	
disgust_307	abandoned	0,274	0,32

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
disgust_307	decay	0,375	
disgust_309	hanging	0,106	0,00
disgust_309	outdoors	-0,104	
disgust_316	invertebrate	-0,231	-0,23
disgust_316	worm	-0,231	
disgust_319	animal	-0,082	-0,08
disgust_319	fish	-0,081	
disgust_39	color	-0,141	-0,11
disgust_39	nature	-0,144	
disgust_39	water	-0,046	
disgust_5	garbage	0,058	0,05
disgust_5	junk	0,065	
disgust_5	pollution	0,024	
disgust_73	garbage	-0,109	-0,11
disgust_73	pollution	-0,080	
disgust_73	trash	-0,126	
disgust_75	abandoned	0,057	0,43
disgust_75	calamity	0,806	
disgust_8	closeup	0,167	0,06
disgust_8	food	-0,054	
disgust_84	garbage	-0,108	-0,04
disgust_84	litter	-0,076	
disgust_84	pollution	0,066	
disgust_85	interior design	0,053	0,01
disgust_85	no person	-0,156	
disgust_85	table	0,131	
disgust_86	dark	0,419	0,10
disgust_86	decay	-0,137	
disgust_86	dirty	0,019	
disgust_93	abstract	-0,169	-0,17
disgust_93	texture	-0,169	
fear_10	city	0,124	0,25
fear_10	dark	0,551	
fear_10	street	0,084	
fear_105	animal	-0,178	-0,21
fear_105	mouth	-0,250	
fear_112	lion	-0,150	-0,16
fear_112	predator	-0,132	
fear_112	wildlife	-0,196	
fear_113	abandoned	-0,139	-0,01
fear_113	old	0,113	
fear_120	black and white	0,117	0,24

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
fear_120	portrait	0,360	
fear_125	abandoned	0,186	0,17
fear_125	black and white	0,159	
fear_126	city	-0,166	-0,03
fear_126	people	-0,017	
fear_126	urban	0,082	
fear_130	arachnid	-0,142	-0,06
fear_130	closeup	0,100	
fear_130	spider	-0,142	
fear_136	dark	0,382	0,19
fear_136	nature	-0,004	
fear_143	fog	0,214	0,03
fear_143	mist	-0,147	
fear_144	insect	0,117	-0,01
fear_144	nature	-0,099	
fear_144	wildlife	-0,054	
fear_150	cornea	0,318	0,38
fear_150	eyeball	0,741	
fear_150	eyelash	0,400	
fear_150	eyelid	0,615	
fear_150	vision	-0,184	
fear_170	horror	0,156	0,10
fear_170	mask	0,012	
fear_170	pain	0,133	
fear_211	eye	-0,196	-0,13
fear_211	portrait	-0,061	
fear_222	black and white	0,083	0,21
fear_222	dark	0,388	
fear_222	mist	0,235	
fear_222	mystery	0,118	
fear_243	field	0,288	0,20
fear_243	grass	0,117	
fear_248	reptile	-0,192	-0,15
fear_248	snake	-0,059	
fear_248	viper	-0,197	
fear_251	abandoned	-0,148	0,05
fear_251	decay	-0,181	
fear_251	shadow	0,471	
fear_266	black and white	-0,075	0,03
fear_266	cold	0,215	
fear_266	winter	-0,038	
fear_275	field	0,111	-0,08
fear_275	nature	-0,128	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
fear_275	rural	-0,153	
fear_275	sepia	-0,130	
fear_276	dark	0,207	0,21
fear_276	mist	0,410	
fear_276	tree	0,022	
fear_283	architecture	-0,018	0,03
fear_283	building	0,236	
fear_283	castle	-0,121	
fear_286	danger	0,242	-0,06
fear_286	deadly	-0,167	
fear_286	reptile	-0,167	
fear_286	snake	-0,167	
fear_287	crowd	0,441	0,20
fear_287	people	-0,045	
fear_288	animal	-0,161	-0,11
fear_288	nature	-0,051	
fear_292	landscape	0,252	-0,05
fear_292	nature	-0,130	
fear_292	ocean	-0,197	
fear_292	sea	-0,138	
fear_294	sea	0,254	-0,05
fear_294	storm	-0,203	
fear_294	water	-0,202	
fear_296	blood	0,092	0,10
fear_296	bloody	0,070	
fear_296	messy	0,124	
fear_299	light	-0,160	-0,12
fear_299	nature	-0,226	
fear_299	tree	0,026	
fear_300	color	-0,134	-0,12
fear_300	nature	-0,224	
fear_300	outdoors	-0,148	
fear_300	tree	-0,056	
fear_300	wood	-0,056	
fear_301	adult	0,224	0,32
fear_301	girl	0,436	
fear_301	woman	0,286	
fear_31	monochromatic	-0,007	0,02
fear_31	shadow	0,041	
fear_313	dark	-0,085	-0,08
fear_313	texture	-0,073	
fear_328	city	0,143	0,11

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
fear_328	portrait	0,071	
fear_44	pattern	-0,006	0,10
fear_44	reflection	0,318	
fear_44	texture	-0,023	
fear_45	balloon	0,167	0,16
fear_45	child	-0,121	
fear_45	grass	0,318	
fear_45	wear	0,292	
fear_49	art	0,203	0,03
fear_49	shadow	-0,149	
fear_52	animal	0,037	0,03
fear_52	lizard	0,097	
fear_52	reptile	-0,033	
fear_55	color	-0,185	0,03
fear_55	leaf	-0,072	
fear_55	light	0,143	
fear_55	old	0,387	
fear_55	wall	-0,123	
fear_62	bathtub	0,170	0,10
fear_62	people	0,023	
fear_71	illuminated	-0,026	0,08
fear_71	light	0,193	
fear_79	sepia	-0,019	0,09
fear_79	silhouette	0,192	
joy_107	baking	-0,146	0,22
joy_107	cake	0,583	
joy_108	flower	-0,097	0,06
joy_108	nature	0,188	
joy_108	outdoors	0,075	
joy_11	cold	-0,102	0,07
joy_11	nature	0,246	
joy_11	snow	0,051	
joy_110	man	-0,013	0,18
joy_110	nude	0,196	
joy_110	outdoors	0,371	
joy_111	beautiful	0,063	0,08
joy_111	flora	0,303	
joy_111	nature	-0,113	
joy_115	beach	-0,119	0,11
joy_115	couple	0,423	
joy_115	silhouette	0,012	
joy_121	blooming	-0,129	-0,13
joy_121	flower	-0,137	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
joy_121	garden	-0,124	
joy_121	nature	-0,135	
joy_122	flower	-0,222	-0,19
joy_122	garden	-0,122	
joy_122	nature	-0,228	
joy_13	beach	0,130	0,16
joy_13	summer	0,230	
joy_13	water	0,107	
joy_14	nature	0,152	0,28
joy_14	reflection	0,408	
joy_148	field	0,512	0,19
joy_148	flower	0,146	
joy_148	nature	0,103	
joy_148	rural	0,011	
joy_15	color	-0,198	-0,22
joy_15	flower	-0,242	
joy_153	calf	0,414	0,11
joy_153	cow	-0,202	
joy_156	people	-0,161	0,01
joy_156	platform	0,100	
joy_156	train	0,219	
joy_156	train station	-0,228	
joy_156	transportation system	0,105	
joy_159	flower	-0,230	-0,18
joy_159	petal	-0,219	
joy_159	rain	-0,102	
joy_16	nature	-0,233	-0,03
joy_16	winter	0,173	
joy_206	duck	-0,250	-0,17
joy_206	lake	-0,146	
joy_206	nature	-0,125	
joy_214	education	0,198	0,16
joy_214	literature	0,131	
joy_236	color	0,265	0,15
joy_236	nature	0,028	
joy_237	craft	-0,073	-0,03
joy_237	handicraft	0,000	
joy_237	handmade	0,019	
joy_237	textile	0,028	
joy_237	wool	-0,117	
joy_251	bee	-0,250	-0,15
joy_251	blooming	0,011	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
joy_251	flower	-0,205	
joy_270	cappuccino	-0,230	-0,22
joy_270	coffee	-0,196	
joy_270	drink	-0,230	
joy_275	bike	-0,061	-0,02
joy_275	recreation	0,026	
joy_292	blue sky	-0,172	-0,11
joy_292	lake	-0,200	
joy_292	landscape	-0,196	
joy_292	mountain	-0,132	
joy_292	nature	0,139	
joy_302	fireplace	0,351	0,17
joy_302	furniture	0,117	
joy_302	home	0,056	
joy_310	commerce	-0,005	-0,01
joy_310	street	0,283	
joy_310	tourist	-0,104	
joy_310	travel	-0,039	
joy_310	urban	-0,182	
joy_314	flower	-0,120	0,06
joy_314	garden	0,242	
joy_325	bird	-0,130	-0,15
joy_325	polar	-0,153	
joy_325	wildlife	-0,157	
joy_33	art	-0,195	-0,09
joy_33	color	-0,149	
joy_33	tree	0,076	
joy_33	watercolor	-0,108	
joy_35	child	-0,200	-0,12
joy_35	leisure	-0,205	
joy_35	nature	0,067	
joy_35	waterfall	-0,132	
joy_36	color	0,130	0,00
joy_36	field	-0,049	
joy_36	flora	0,074	
joy_36	flowerbed	-0,153	
joy_38	artistic	0,052	0,08
joy_38	bright	0,155	
joy_38	gold	0,037	
joy_4	nature	0,005	0,13
joy_4	swan	0,250	
joy_46	beach	-0,176	-0,14
joy_46	sand	-0,084	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
joy_46	sea	-0,158	
joy_48	art	-0,117	0,09
joy_48	recreation	0,298	
joy_55	beautiful	0,030	-0,12
joy_55	flora	-0,222	
joy_55	nature	-0,167	
joy_66	building	-0,147	0,01
joy_66	family	0,167	
joy_70	black and white	0,217	0,11
joy_70	child	0,004	
joy_72	goose	-0,197	-0,07
joy_72	nature	-0,198	
joy_72	outdoors	0,186	
joy_73	candy	-0,140	-0,01
joy_73	celebration	-0,138	
joy_73	delicious	0,302	
joy_73	sweet	-0,065	
joy_98	botanical	-0,120	0,00
joy_98	flower	0,021	
joy_98	nature	0,109	
joy_99	flower	-0,176	-0,17
joy_99	nature	-0,170	
sadness_101	landscape	-0,232	-0,06
sadness_101	sky	0,110	
sadness_107	sculpture	-0,033	0,04
sadness_107	statue	0,109	
sadness_115	eye	0,079	-0,02
sadness_115	eyeball	-0,091	
sadness_115	face	0,139	
sadness_115	vision	-0,200	
sadness_124	dark	-0,123	0,00
sadness_124	shadow	0,325	
sadness_124	woman	-0,207	
sadness_125	outdoors	-0,132	0,13
sadness_125	people	0,396	
sadness_147	animal	-0,123	-0,13
sadness_147	street	-0,120	
sadness_147	wall	-0,143	
sadness_15	cemetery	0,130	-0,01
sadness_15	sculpture	-0,146	
sadness_151	sculpture	0,175	0,07
sadness_151	snow	-0,038	
sadness_152	people	-0,142	0,06

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
sadness_152	portrait	0,256	
sadness_160	dress	-0,191	-0,13
sadness_160	home	-0,191	
sadness_160	loneliness	0,113	
sadness_160	woman	-0,250	
sadness_173	man	0,153	0,19
sadness_173	monochrome	-0,085	
sadness_173	wait	0,514	
sadness_18	old	0,461	0,29
sadness_18	sculpture	0,118	
sadness_185	adult	0,153	0,31
sadness_185	monochrome	0,464	
sadness_185	shadow	0,306	
sadness_19	rain	-0,144	-0,15
sadness_19	raindrop	-0,150	
sadness_19	rainy	-0,149	
sadness_190	boy	-0,233	-0,12
sadness_190	child	-0,233	
sadness_190	offense	0,112	
sadness_196	face	0,081	0,09
sadness_196	girl	0,105	
sadness_196	portrait	0,079	
sadness_199	boy	0,155	0,04
sadness_199	child	-0,068	
sadness_20	field	0,032	0,07
sadness_20	sport	0,188	
sadness_20	stadium	-0,011	
sadness_218	abandoned	-0,087	-0,05
sadness_218	old	0,065	
sadness_218	rusty	-0,176	
sadness_218	tree	-0,019	
sadness_221	grief	0,184	0,08
sadness_221	old	-0,136	
sadness_221	sadness	0,189	
sadness_232	flower	0,027	0,06
sadness_232	rose	0,089	
sadness_238	grave	0,628	0,29
sadness_238	old	0,310	
sadness_238	sculpture	-0,065	
sadness_254	reflection	0,212	0,00
sadness_254	river	-0,009	
sadness_254	water	-0,212	
sadness_27	art	-0,059	-0,11

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
sadness_27	painting	-0,169	
sadness_277	one	-0,250	-0,14
sadness_277	water	-0,202	
sadness_277	wood	0,030	
sadness_296	flora	0,292	0,17
sadness_296	rose	-0,042	
sadness_296	still life	0,274	
sadness_300	abstract	-0,226	-0,08
sadness_300	reflection	0,065	
sadness_306	empty	0,188	0,02
sadness_306	water	-0,148	
sadness_32	girl	0,284	0,07
sadness_32	monochrome	0,105	
sadness_32	sepia	-0,178	
sadness_327	adult	-0,239	-0,14
sadness_327	man	-0,250	
sadness_327	portrait	0,067	
sadness_33	child	-0,208	-0,22
sadness_33	teddy	-0,235	
sadness_33	toy	-0,202	
sadness_330	bicycle	-0,064	-0,05
sadness_330	wheel	-0,031	
sadness_44	sculpture	-0,197	-0,06
sadness_44	statue	-0,051	
sadness_44	tombstone	0,061	
sadness_45	adult	-0,128	0,14
sadness_45	alone	0,351	
sadness_45	street	0,207	
sadness_48	city	0,099	-0,07
sadness_48	woman	-0,235	
sadness_51	animal	0,053	0,22
sadness_51	farm	0,390	
sadness_6	cemetery	0,044	-0,12
sadness_6	sculpture	-0,205	
sadness_6	statue	-0,196	
sadness_63	child	0,018	0,10
sadness_63	color	0,156	
sadness_63	rain	0,113	
sadness_72	decay	0,123	0,06
sadness_72	flower	0,091	
sadness_72	old	-0,049	
sadness_8	container	0,279	0,28
sadness_8	decoration	0,125	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
sadness_8	furniture	0,448	
sadness_80	black and white	-0,242	-0,20
sadness_80	countryside	-0,142	
sadness_80	road	-0,216	
sadness_80	rural	-0,205	
sadness_99	cemetery	-0,080	0,04
sadness_99	grave	0,153	
surprise_110	nature	-0,018	-0,10
surprise_110	river	-0,137	
surprise_110	waterfall	-0,133	
surprise_114	fungus	0,075	0,09
surprise_114	mushroom	0,034	
surprise_114	nature	0,157	
surprise_115	outdoors	0,266	0,02
surprise_115	park	-0,239	
surprise_115	tree	0,038	
surprise_13	luxury	0,188	0,24
surprise_13	outdoors	0,293	
surprise_135	animal	0,096	0,07
surprise_135	nature	-0,144	
surprise_135	wild	0,144	
surprise_135	wildlife	0,172	
surprise_138	evening	0,152	0,04
surprise_138	nature	-0,070	
surprise_138	sky	0,098	
surprise_138	sunset	-0,037	
surprise_139	beach	-0,162	-0,07
surprise_139	nature	0,030	
surprise_15	landscape	-0,148	-0,07
surprise_15	sky	-0,032	
surprise_15	sunset	-0,024	
surprise_168	color	0,556	0,26
surprise_168	nature	0,101	
surprise_168	tree	0,122	
surprise_187	animal	-0,051	0,09
surprise_187	bird	-0,080	
surprise_187	nature	-0,023	
surprise_187	tree	0,416	
surprise_187	wildlife	0,182	
surprise_190	butterfly	-0,149	-0,10
surprise_190	nature	-0,046	
surprise_204	landscape	-0,162	-0,15
surprise_204	nature	-0,066	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
surprise_204	water	-0,250	
surprise_204	waterfall	-0,141	
surprise_22	color	-0,051	0,02
surprise_22	fall	-0,059	
surprise_22	illustration	-0,021	
surprise_22	leaf	0,296	
surprise_22	nature	-0,043	
surprise_220	desert	-0,074	0,01
surprise_220	geology	-0,074	
surprise_220	rock	0,187	
surprise_221	lake	-0,230	-0,19
surprise_221	mountain	-0,196	
surprise_221	nature	-0,145	
surprise_222	mountain	0,000	-0,03
surprise_222	nature	-0,121	
surprise_222	river	-0,154	
surprise_222	tree	-0,053	
surprise_222	valley	0,181	
surprise_225	beach	-0,200	-0,11
surprise_225	horizon	0,080	
surprise_225	sky	-0,212	
surprise_238	landscape	-0,077	-0,05
surprise_238	mountain	-0,031	
surprise_245	blooming	-0,042	0,03
surprise_245	floral	0,096	
surprise_245	flower	-0,081	
surprise_245	petal	0,165	
surprise_25	light	0,009	0,10
surprise_25	room	0,183	
surprise_250	mountain	-0,012	-0,12
surprise_250	nature	-0,138	
surprise_250	sky	-0,210	
surprise_273	outdoors	-0,134	-0,02
surprise_273	rock	-0,169	
surprise_273	tree	0,240	
surprise_29	botanical	0,102	0,07
surprise_29	flora	-0,090	
surprise_29	nature	0,183	
surprise_303	landscape	-0,143	-0,09
surprise_303	nature	-0,095	
surprise_303	sky	-0,172	
surprise_303	sunset	0,039	
surprise_305	nature	-0,206	-0,11

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
surprise_305	rainforest	-0,143	
surprise_305	tree	0,111	
surprise_305	waterfall	-0,203	
surprise_309	nature	-0,162	0,05
surprise_309	road	0,107	
surprise_309	trail	0,190	
surprise_313	beautiful	0,053	-0,01
surprise_313	floral	0,181	
surprise_313	nature	-0,212	
surprise_313	petal	-0,051	
surprise_32	motion	-0,084	-0,01
surprise_32	wildlife	0,064	
surprise_323	fireworks	0,340	0,21
surprise_323	party	0,081	
surprise_324	berry	-0,125	-0,03
surprise_324	fruit	-0,125	
surprise_324	raspberry	0,148	
surprise_35	ice	0,179	-0,03
surprise_35	nature	-0,151	
surprise_35	pine	-0,138	
surprise_35	snow	-0,133	
surprise_35	tree	0,109	
surprise_38	car	0,077	0,15
surprise_38	city	0,077	
surprise_38	street	0,385	
surprise_38	vehicle	0,075	
surprise_48	flamingo	-0,097	-0,02
surprise_48	garden	0,064	
surprise_53	child	-0,013	0,11
surprise_53	water	0,237	
surprise_6	fall	-0,207	-0,11
surprise_6	nature	-0,145	
surprise_6	tree	0,018	
surprise_65	deer	-0,200	-0,16
surprise_65	nature	-0,125	
surprise_66	bird	0,318	0,20
surprise_66	nature	0,081	
surprise_66	tropical	0,211	
surprise_69	bakery	-0,216	-0,11
surprise_69	cake	0,218	
surprise_69	food	-0,207	
surprise_69	pastry	-0,197	

Continued on next page

Table A.2 – continued from previous page

Image	tag	Value of K	Average Value of K for image
surprise_69	sweet	-0,148	
surprise_83	mountain	-0,203	0,09
surprise_83	nature	0,214	
surprise_83	outdoors	0,273	
surprise_85	bird	0,062	-0,14
surprise_85	field	-0,182	
surprise_85	nature	-0,215	
surprise_85	outdoors	-0,242	
surprise_93	garden	-0,053	-0,10
surprise_93	nature	-0,139	
surprise_96	animal	0,021	-0,10
surprise_96	deer	-0,141	
surprise_96	nature	-0,176	

Table A.3 – Top rated areas for images with more than one associated emotion

Emotion	Zone with most votes	Number of images
anger/disgust/sadness	zone_E4	1
disgust/fear	zone_E2	1
disgust/fear	zone_E7	1
disgust/happiness/neutral	zone_E5	1
disgust/neutral	zone_E5	1
disgust/sadness	zone_E6	1
disgust/sadness/surprise	zone_E1,zone_E2,zone_E3,zone_E5, zone_E6, zone_E8,zone_E9	1
disgust/surprise	zone_E5	1
disgust/surprise	zone_E6	1
disgust/surprise	zone_E7, zone_E8, zone_E9	1
fear/happiness	zone_E4	1
fear/happiness	zone_E5	1
fear/happiness	zone_E5, zone_E6	1
fear/happiness	zone_E5, zone_E9	1
fear/happiness	zone_E9	1
fear/happiness/neutral	zone_E6, zone_E8	1
fear/happiness/neutral	zone_E7	1
fear/neutral	zone_E4	1
fear/neutral	zone_E5	1
fear/neutral/sadness	zone_E5	2
fear/neutral/sadness	zone_E2	1
fear/neutral/sadness	zone_E2, zone_E5	1
fear/neutral/surprise	zone_E6	1
fear/sadness	zone_E6	2
fear/sadness	zone_E8	2
fear/sadness	zone_E2, zone_E4	1
fear/sadness	zone_E5	1
fear/surprise	zone_E5, zone_E6	1
happiness/neutral	zone_E2	2
happiness/neutral	zone_E3	1
happiness/neutral	zone_E4	1
happiness/neutral	zone_E5	1
happiness/neutral	zone_E6	1
happiness/neutral	zone_E8	1
happiness/neutral	zone_E9	1
happiness/neutral/sadness/surprise	zone_E5	1
happiness/neutral/surprise	zone_E5	2
happiness/surprise	zone_E2	2
happiness/surprise	zone_E4	1
happiness/surprise	zone_E5	1
happiness/surprise	zone_E7, zone_E8, zone_E9	1
happiness/surprise	zone_E8	1
neutral/sadness	zone_E5	4
neutral/sadness	zone_E3	1
neutral/sadness	zone_E6	1
neutral/sadness	zone_E8, zone_E9	1
neutral/surprise	zone_E1,zone_E2,zone_E4,zone_E7, zone_E8	1
neutral/surprise	zone_E3	1
neutral/surprise	zone_E4	1
neutral/surprise	zone_E5	1
neutral/surprise	zone_E8	1
sadness/surprise	zone_E5	2
sadness/surprise	zone_E4	1

Table A.4 – Least voted areas for images with more than one associated emotion

Emotion	Zone with most votes	Number of images
anger/disgust/sadness	zone_E1,zone_E9	1
disgust/fear	zone_E1	1
disgust/fear	zone_E5,zone_E6	1
disgust/happiness/neutral	zone_E7,zone_E9	1
disgust/neutral	zone_E2,zone_E3,zone_E6,zone_E9	1
disgust/sadness	zone_E3,zone_E8,zone_E9	1
disgust/sadness/surprise	zone_E4,zone_E7	1
disgust/surprise	zone_E1,zone_E2,zone_E3,zone_E6	1
disgust/surprise	zone_E7	1
disgust/surprise	zone_E8	1
fear/happiness	zone_E1	1
fear/happiness	zone_E1,zone_E2,zone_E3	1
fear/happiness	zone_E2	1
fear/happiness	zone_E2,zone_E3,zone_E8,zone_E9	1
fear/happiness	zone_E3,zone_E6,zone_E9	1
fear/happiness/neutral	zone_E1,zone_E3	2
fear/neutral	zone_E5	1
fear/neutral	zone_E1	1
fear/neutral/sadness	zone_E4,zone_E7,zone_E8,zone_E9	1
fear/neutral/sadness	zone_E3	1
fear/neutral/sadness	zone_E1,zone_E2,zone_E3	1
fear/neutral/sadness	zone_E1,zone_E3	1
fear/neutral/surprise	zone_E1	1
fear/sadness	zone_E3	2
fear/sadness	zone_E7,zone_E8	1
fear/sadness	zone_E6	1
fear/sadness	zone_E1,zone_E2	1
fear/sadness	zone_E3,zone_E7	1
fear/surprise	zone_E1	1
happiness/neutral	zone_E3	2
happiness/neutral	zone_E4,zone_E5	1
happiness/neutral	zone_E1,zone_E2,zone_E3	1
happiness/neutral	zone_E1,zone_E2,zone_E3,zone_E8,zone_E9	1
happiness/neutral	zone_E8,zone_E9	1
happiness/neutral	zone_E2,zone_E3,zone_E7,zone_E9	1
happiness/neutral	zone_E2,zone_E9	1
happiness/neutral/sadness/surprise	zone_E1	1
happiness/neutral/surprise	zone_E3,zone_E6,zone_E9	1
happiness/neutral/surprise	zone_E1	1
happiness/surprise	zone_E1,zone_E2,zone_E3	2
happiness/surprise	zone_E7,zone_E8,zone_E9	1
happiness/surprise	zone_E9	1
happiness/surprise	zone_E6	1
happiness/surprise	zone_E7,zone_E8	1
neutral/sadness	zone_E1,zone_E2,zone_E3,zone_E7,zone_E8,zone_E9	1
neutral/sadness	zone_E7,zone_E9	1
neutral/sadness	zone_E3,zone_E9	1
neutral/sadness	zone_E1,zone_E4,zone_E7	1
neutral/sadness	zone_E9	1
neutral/sadness	zone_E2,zone_E8	1
neutral/sadness	zone_E3	1
neutral/surprise	zone_E9	2
neutral/surprise	zone_E8,zone_E9	1
neutral/surprise	zone_E3,zone_E6,zone_E9	1
neutral/surprise	zone_E6	1
sadness/surprise	zone_E2,zone_E3,zone_E6,zone_E7,zone_E9	1
sadness/surprise	zone_E9	1
sadness/surprise	zone_E6,zone_E7,zone_E9	1